Universiteit Leiden
The Netherlands

Spectral imaging and tomographic reconstruction methods for industrial applications

Zeegers, M.T.

Citation

Zeegers, M. T. (2023, May 31). *Spectral imaging and tomographic reconstruction methods for industrial applications*. Retrieved from https://hdl.handle.net/1887/3619550

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3619550

Note: To cite this publication please use the final published version (if applicable).

Spectral imaging and tomographic reconstruction methods for industrial applications

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op 31 mei 2023
klokke 15:00 uur

door

Mathé Thomas Zeegers

geboren te Alkmaar, Nederland
in 1992

Promotor:
Prof. dr. K. J. Batenburg

Co-promotores:
Dr. D. M. Pelt
Dr. T. van Leeuwen (Centrum Wiskunde & Informatica, Universiteit Utrecht)

Promotiecommissie:
Prof. dr. A. Plaat
Prof. dr. T. H. W. Bäck
Prof. dr. ir. F. J. Verbeek
Prof. dr. R. van Liere (Centrum Wiskunde & Informatica, Technische Universiteit Eindhoven)
Dr. M. A. J. M. van Eijnatten (Technische Universiteit Eindhoven)

The research presented in this dissertation was carried out at the Centrum Wiskunde & Informatica (CWI) in Amsterdam.

Financial support was provided by the Netherlands Organisation for Scientific Research (NWO), programme 639.073.506.

© 2023 Mathé T. Zeegers

Contents

1 Introduction **1**
1.1 Product inspection and foreign objects . . . 3
1.2 Radiography with X-rays . . . 6
1.3 Tomography . . . 12
1.4 Spectral X-ray imaging . . . 18
1.5 Spectral tomography . . . 23
1.6 Machine learning . . . 29
1.7 Research questions and dissertation outline . . . 36

2 Tomographic workflows for deep learning training data generation **39**
2.1 Introduction . . . 40
2.2 Preliminaries . . . 42
2.3 Proposed method for training data acquisition . . . 44
2.4 Experiments and results . . . 45
2.5 Discussion . . . 59
2.6 Conclusions . . . 62

3 Hyperspectral data reduction through deep learning **65**
3.1 Introduction . . . 66
3.2 Related work . . . 67
3.3 Materials and methods . . . 69
3.4 Experiments and results . . . 74
3.5 Discussion . . . 92
3.6 Conclusions . . . 93

4 Multi-channel discrete tomography **95**
4.1 Introduction . . . 96
4.2 Problem formulation . . . 98
4.3 Algorithms . . . 100
4.4 Experiments and results . . . 106
4.5 Discussion . . . 110
4.6 Conclusions . . . 111

5 Dictionary-based spectral tomography **113**
5.1 Introduction . . . 114

5.2 Related work . . . 118
5.3 Spectral forward model . . . 119
5.4 Spectral inverse problem . . . 120
5.5 Proposed method - ADJUST . . . 123
5.6 Experiments and results . . . 128
5.7 Conclusions and discussion . . . 145

6 Conclusions and outlook **149**
6.1 Conclusions . . . 149
6.2 Contributions . . . 150
6.3 Future work and outlook . . . 152

Bibliography **157**

A Appendices to Chapter 2 **177**
A.1 Intensity value histograms . . . 178
A.2 Reconstruction and ground truth similarities . . . 179
A.3 Additional quality measure . . . 180

B Appendices to Chapter 3 **181**
B.1 Standard data reduction methods . . . 182
B.2 X-ray projection data computation . . . 183
B.3 Time comparison . . . 185
B.4 Robustness . . . 186

C Appendices to Chapter 5 **189**
C.1 Proof of Theorem 1 . . . 190
C.2 Bi-convexity of ADJUST and partial optimality . . . 191
C.3 Derivation of AAPM . . . 194
C.4 Gradient computations . . . 196
C.5 Dictionary matrix . . . 197
C.6 Performance measures . . . 198
C.7 Numerical studies: Comparison of methods . . . 200
C.8 Numerical studies: Limited measurement patterns . . . 201
C.9 Numerical studies: Mixed material phantom . . . 204
C.10 Numerical studies: 3D phantom . . . 205

List of publications **209**

Summary in English **211**

Summary in Dutch **215**

Curriculum vitae **219**

Acknowledgements **221**

1 Introduction

For well over a century, X-ray imaging has been successfully employed as a non-destructive method to gain insight into the internal structures of various objects. Some well-known applications include baggage scanning at airports and medical diagnosing in hospitals. One particularly vital application area is the food processing chain, where quality preservation and safety checks are the order of the day. Even though most food inspection is surface-based, X-ray imaging has proven to be an essential complement to finding significant anomalies within a food product. Still, a more detailed understanding of the interior of an object can be achieved using Computed Tomography (CT), in which X-ray projections are collected around the imaged object and used to compute a 3D reconstruction of it. These CT acquisitions and algorithms generally take more time than conventional X-ray imaging, which can pose a problem at industrial high-throughput production lines. These conflicting goals of accuracy and processing speed require a trade-off to be made, and new methods and algorithms are needed to optimize this trade-off.

Towards achieving the optimized trade-off, spectral imaging and machine learning are important directions to respectively obtain rich data and achieve faster and better results. Recently, the field of spectral imaging, where the energies within the X-ray beams are taken into account, has been getting more attention. A major driving force is the development of spectral photon-counting detectors, which are able to register the energy properties of each individual photon in the incoming beam independently and can therefore disentangle energetic information into multiple images at different wavelengths. This is a significant information gain compared to conventional X-ray imaging, and can therefore be used to improve X-ray image analysis and to create more accurate CT reconstructions with the same acquisition time. These detectors and reconstruction algorithms are continuously in development. However, the lack of suitable spectral image processing methods and reconstruction algorithms is hampering the application of spectral X-ray imaging in industry in general and food processing in particular. Additionally, recent

developments in machine learning open up possibilities for rapid and accurate X-ray image analysis. In particular, deep learning enables rapid recognition of complicated patterns in data and can therefore aid in image processing tasks. The potential of deep learning has yet to be explored for the examination of X-ray images of natural products in industrial settings.

In this dissertation, several research contributions are put forward for spectral X-ray CT and the incorporation of deep learning in (spectral) image processing routines to improve reconstruction accuracy and performance of image processing tasks. The main contribution of this work is the proposal and detailed analysis of a new deep learning workflow for imaging tasks in which the scientific contributions in all chapters are incorporated. The method can be readily implemented in industrial setups and has a particular focus on food processing. This introductory chapter will first provide the background of product inspection in industry, which will be the common thread for all subsequent motivations. After this, X-ray imaging and its application to product inspection are covered. Next, the basic concepts of computed tomography are introduced. Thereafter, an outline of the current state of the art of spectral X-ray imaging and spectral CT is given. Along with this, several machine learning concepts are introduced, which are used in later sections. This chapter is concluded with an overview of all sections in this dissertation, supplemented by the relevant research questions for each non-preliminary chapter.

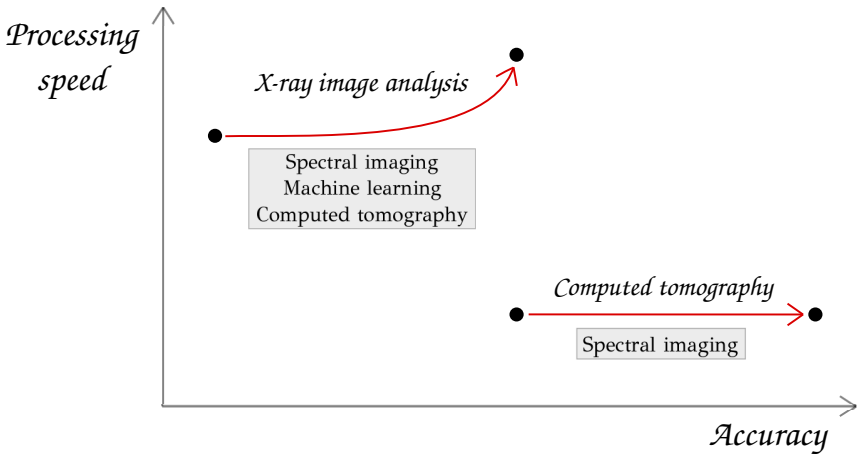


Figure 1.1: Schematic view on the goals in this dissertation. The pursued contributions by the research are indicated by the red arrows. Spectral imaging, machine learning and tomography methods are used to increase the accuracy and speed (i.e. image throughput) of algorithms for tasks on X-ray images. In addition, spectral imaging is used to increase the reconstruction accuracy in computed tomography.

1.1 Product inspection and foreign objects

With the increasing reliance of society on automatic food processing in factories, accurate quality control of food is absolutely essential [305, 327]. Regularly, food products are recalled from supermarkets or other food distributors because of *contamination*: unintended and possibly undetected alterations to the products, which makes them unsuitable for consumption. The contamination and the subsequent recalls waste time, money, effort and energy. On top of that, consumers and product sellers may lose trust in the product manufacturer. It goes without saying that all sorts of contamination should therefore be avoided as much as possible. In general, there are three different distinguishable types of product contamination [89, 287]. Chemical contaminations can appear as pesticides, wrong food additives or inappropriate cleaning substances. Biological contamination encompasses bacterial, fungal, virus and parasite contamination. Both can deteriorate food quality but can also lead to severe food poisoning. In this dissertation, we focus mainly on the third category of *physical contamination*, where food contains items that are not supposed to be there. The existence of this third category is the primary motivator for the realization of the methods presented in this dissertation.

Physical contamination is characterized by the presence of *foreign objects*, also known as foreign bodies. These foreign objects are defined as objects that are not supposed to be present in a product and may harm the end user upon consuming or using it if its presence remains unknown. Foreign objects come in many shapes and conditions. Usually, a distinction is made between intrinsic foreign objects, which are already present in the object but not properly removed (e.g. bones, shells, fruit stalks or pits), and extrinsic foreign objects, which are objects that end up in the product during the food processing stages (e.g. stones, plastic and insects) [76]. Foreign objects of the latter category can be introduced during transport, improperly sanitized factory processing environments, sanitizing processes, automatic or manual deformation processes, and packaging stages. Typical example objects are stones, wood, insects, plastics, hair, pieces of (fallen) gloves, rubber, glass and metal particles. In addition to the harm that can be done after processing, hard foreign objects can also damage the processing machines. Standard control measures such as sieves, filters and magnets can be used to mitigate the problem, but these have technical and operational limitations to minimizing the risks [68]. Furthermore, these methods are automatic and do not necessarily enable the notion of foreign objects being present, as for instance sieves only remove larger foreign objects.

To address the problem of foreign objects, *food inspection methods* are continuously being developed and employed in the food processing industry. Such product inspection methods are in place as well where non-food items are checked for product defects, but also in places such as airports for luggage inspections. Ideally, these methods can eventually be used for assessing products on more advanced quality aspects such as tenderness, ripeness, bacteria presence and other

biological properties as well. Nevertheless, all applications have two common goals: the methods should be fast and accurate. These two conflicting goals ask for the continuous development of new inspection methods.

Until the end of the previous millennium, optical sensing technologies were the only well-established methods [106] for foreign object detection. These methods work well when the product is coarse and loose or when the defects are located on the surface, and when foreign objects can be detected based on differences in shapes or colors. For foreign objects that are hidden in a product, the only techniques that are not surface-based and have widespread commercial availability at low cost were metal detection methods. Because these methods can only detect metals, other *noninvasive inspection* methods (also known as nondestructive methods), in which an object can be inspected without damaging it, were highly desired.

Nowadays, many approaches to nondestructive internal product inspection have been developed. The development and availability are due to scientific progress, decreasing machinery costs and the development of sophisticated signal and image processing techniques over the years. Most of the current methods are based on the interaction between the electromagnetic radiation that penetrates the product and the foreign object [106]. As Figure 1.2 indicates, some state-of-the-art approaches include [117, 194, 200, 305] magnetic resonance imaging (in the radiofrequency range), microwave imaging [281], terahertz imaging [324], thermal infrared imaging, optical hyperspectral imaging [85, 89], X-ray imaging and gamma-ray inspection.

The optimal approach and the chosen type of radiation from the electromagnetic spectrum are usually highly problem-specific and are chosen according to the material properties and inspection goal. Important is the *penetration depth* of the active electromagnetic radiation and its interaction that enables distinction of the materials of interest. Radiowaves have high penetration depth but come with low image acquisition speed. Additionally, they need powerful and costly equipment to generate a strong magnetic field, and are prone to motion artefacts. Microwaves only penetrate by a few centimeters. For other electromagnetic radiation types, the penetration depth is typically very limited. However, it increases with the energy of the wave, with X-rays and gamma rays providing sufficient power to fully penetrate through objects. Both gamma rays and X-rays are ionizing and therefore harmful, but - despite common belief - the effect of X-rays on food products is extremely limited, and radiated food is therefore safe to consume. The high penetration power of gamma rays is mostly used to inspect metallic and welded structures. For softer materials it is less suitable, as differences between materials cannot be observed as well as with X-rays. Another method based on acoustic waves is ultrasound and is, together with X-rays, the most widely used inspection method. However, ultrasound is only useful for imaging soft materials, such as soft tissue. It is less suitable when air or other materials with significant differences in acoustic impedance are involved in the object. On the other hand, X-rays can penetrate all common packaging materials [68]. Compared to widespread simple metal detectors,

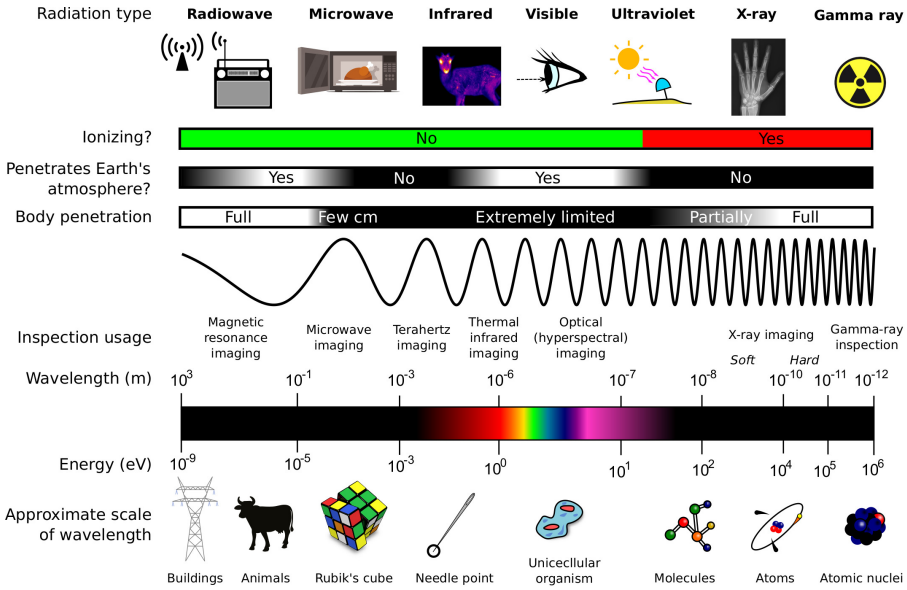


Figure 1.2: Overview of inspection methods based on interaction with the electromagnetic spectrum with radiation types, wavelengths, generic applications, and penetration depth [55, 291].

Method	Speed	Costs	Complexity	Data	Material identification
Standard X-ray imaging	Fast	Low	Low	Small	Low discrimination
Standard X-ray tomography	Slow	Moderately high	Moderate	Moderate	Moderately low discrimination
Spectral X-ray imaging	Fast	High	High	Moderate	Moderately high discrimination
Spectral X-ray tomography	Slow	High	High	Very large	High discrimination

Table 1.1: Different X-ray imaging methods and their properties.

X-ray machinery can expose a wider range of materials including stones, glass, bones, and plastics.

For this reason, this dissertation will focus mostly on X-ray imaging for industrial imaging. State-of-the-art X-ray imaging comes in many modalities. In Section 1.2, *standard X-ray imaging* is explained, which is fast and inexpensive, but does not necessarily yield 2D images with sufficient material differentiation for further analysis. Section 1.3 discusses *X-ray computed tomography*, which by repeatedly capturing radiographs under different angles and reconstructing a 3D volume enables better analysis of an object. Section 1.4 introduces *spectral X-ray imaging*, which generates radiographs at different X-ray energies for better material discrimination than with standard X-ray imaging. Section 1.5 encompasses *spectral X-ray computed tomography*, which combines spectral X-ray imaging and computed tomography for more advanced material discrimination in reconstructed 3D volumes. Table 1.1 gives an overview of all techniques and their relative characteristics.

1.2 Radiography with X-rays

Just as famous as the discovery of X-ray radiation by Wilhelm Conrad Röntgen in 1895 are the subsequently made X-ray images of a hand with a ring, of which the first one Röntgen made revealing the bone structure of his wife's hand. To this day, these images show the enormous potential to visualize the inner structure of objects noninvasively. Therefore, inspection of food is just one of the many applications of X-ray imaging which is, together with gamma-ray imaging, also known as *radiography*. One year after Röntgen's discovery, the first X-ray devices were already in use for clinical observations, and the medical applications of radiography to inspect conditions of various parts of the body (such as possibly broken bones and dental conditions) have been indispensable ever since. With further industrial advances during the 20th century, radiography for baggage screening and cargo inspection, quality inspection of castings and welds and quality control of parts of products such as cars [192] have become crucial as well. Mechanized implementations of these enable *automated X-ray inspection* (AXI). Despite the possibly harmful ionizing radiation of X-rays, radiography has become widely used because it is relatively cheap, fast, easily accessible and has a relatively high penetration depth compared to other inspection methods for such applications [200].

For the electromagnetic spectrum, a single quantum is a *photon*. Depending on the application, photons are described using either properties of particles or waves. Each photon contains a certain amount of *energy* E, measured in electronvolt (eV). The energy of a photon depends on its *frequency* f (in s^{-1}), through

$$E = hf.$$

where h is Planck's constant ($h \approx 4.136 \cdot 10^{-15}$ eV$\cdot s$). Photons and their associated energy are usually expressed either in frequency f or in *wavelength* λ, which is

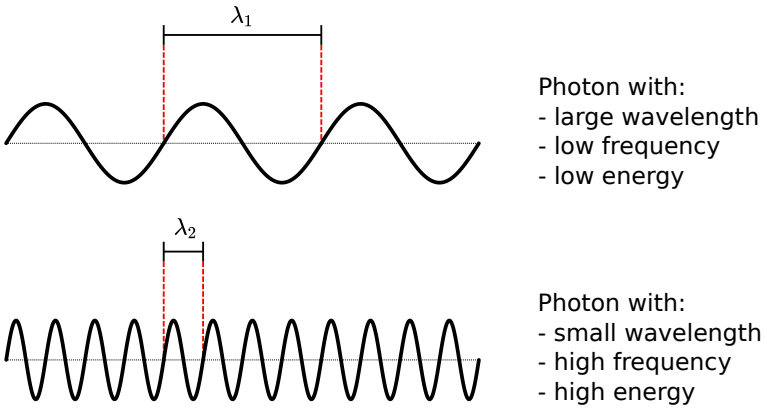


Figure 1.3: Photon model and the comparison of the (equivalent) characteristics between two photons with large wavelength λ_1 and small wavelength λ_2.

inversely proportional to the frequency f through the speed of light c in a vacuum (in $m \cdot s^{-1}$):

$$\lambda = \frac{c}{f}.$$

Electromagnetic radiation is a stream of photons, each carrying its own energy. As indicated in Figure 1.2, the wavelength (or, equivalently, energy or frequency, see Figure 1.3) of the photons determines the type of electromagnetic radiation involved. X-ray beams are made up of photons with wavelengths between 0.01 nm and 10 nm (approximately equivalent to between 100 eV and 100 keV). Photons with high energy - and therefore with small wavelengths - are able to penetrate further into matter than those with low energy. Therefore, a distinction is often made between *soft* and *hard* X-rays. The former consists of photons with maximum energies below 10 keV (above 0.1 nm wavelength) typically used for medical applications, and the latter of photons with maximum energies above 10 keV, which are more prevalent in industrial imaging.

To produce X-rays, an X-ray source consisting of a vacuum tube with a cathode and an anode is typically used (see Figure 1.4). Electrons are emitted from the cathode using the thermionic effect by heating the filament of the cathode to overcome their binding energy, and are accelerated towards an anode due to a voltage applied between the two [51, 55, 192, 242]. The acceleration increases with the *peak voltage* applied (usually indicated in kVp), while the number of released electrons from the filament increases with the applied *current*. The target material of the anode (usually tungsten) is bombarded with electrons, whose kinetic energy is mostly converted to heat, but a small fraction is converted into X-radiation via various processes when the electrons interact with the atoms of the anode material.

In the first of those interaction types, the electron is decelerated due to deflection around the atomic nucleus. The deceleration then converts the kinetic energy of the electron into X-ray radiation, known as braking radiation or *bremsstrahlung*. The energy of the radiation is directly proportional to the amount of deceleration. It is also possible that an incident electron hits an electron in the shell of an atom and kicks it out. When this happens, an electron from an outer shell will take its place, and the atom emits X-rays in the process. The involved shells of the atom characterize the energy of the emitted radiation, and this *characteristic radiation* is therefore highly specific to the material type of the anode. This emission process is known as *X-ray fluorescence* (XRF).

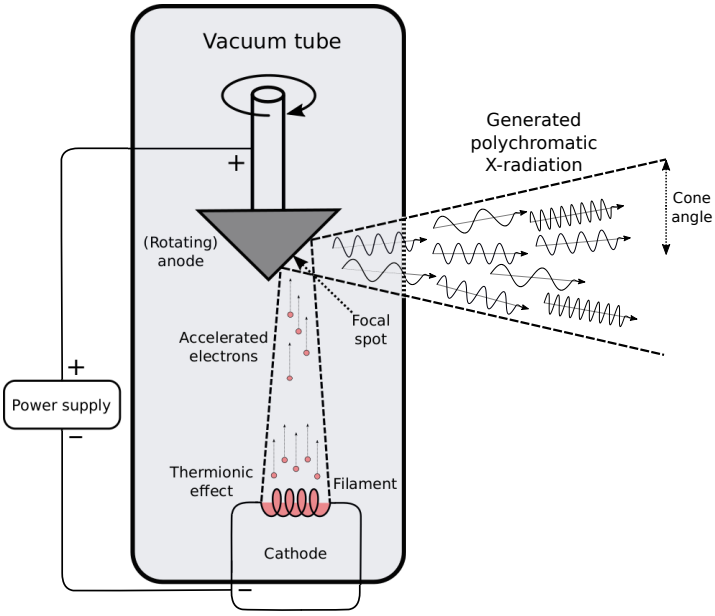


Figure 1.4: Schematic representation of an X-ray source. Electrons are emitted from the filament in the cathode by applying power to heat it up, after which by interaction with the anode a fraction is converted to X-radiation.

The location from which the radiation is emitted in the tube is the *focal spot* of the resulting X-ray beam. Ideally, the focal spot should be a point. In reality, this is an area of a specific size. A small focal spot is better than a large focal spot, as this will reduce the number of blurry edges when forming an X-ray image. On the other hand, a small focal spot will focus the heat in a smaller area and could cause melting of the anode material. For this reason, a rotating anode can be used to distribute the heat more evenly [192]. The generated X-ray beam is *cone*-shaped, and is defined by its *cone angle*. The intensity of the beam is usually referred to as the *flux* of the beam.

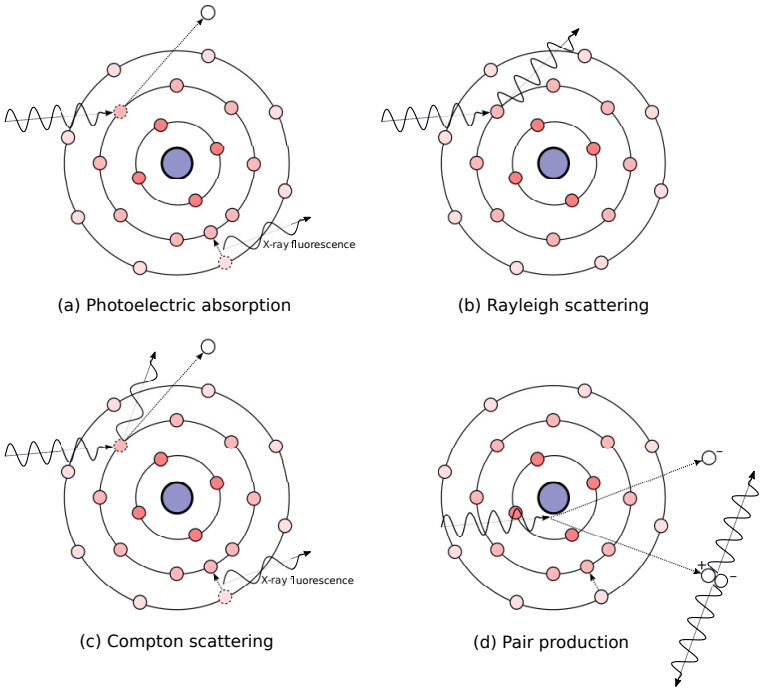


Figure 1.5: Schematic view of X-ray interaction with atoms: (**a**) photoelectric absorption with X-ray fluorescence, (**b**) Rayleigh scattering, (**c**) Compton scattering with X-ray fluorescence and (**d**) pair production.

Alternatively, for much higher-quality beams there are *synchrotrons*. In these ring-shaped facilities, (injected) electrons move at high speeds in a closed loop. Powerful magnets are used to bend the directions of the electrons, at each of which intense radiation is emitted. Typically, synchrotrons are mostly used for scientific research. Researchers acquire beam-time at one of the suitable beamlines (with specific beam characteristics), at which obtaining beam-time is a challenging and involved process. The beams that result from this generation process are typically high-intensity *parallel* beams.

Different types of interactions can occur when a photon in an X-ray beam travels through an object (see Figure 1.5). If a photon has low energy, the dominating interactions are *photoelectric absorption* and *Rayleigh scattering*. In photoelectric absorption, a photon collides with an electron in the shell of an atom, and transfers all of its energy to this electron. If this energy is higher than the binding energy of the electron, the photon is completely attenuated and the electron is ejected from the atomic shell. Due to this unstable ionized state, an electron from a higher shell will take its place and the energy surplus will be emitted as characteristic radiation.

This radiation usually has low energy, but is unique to the material's *atomic number* (usually indicated by Z). The second effect of Rayleigh scattering occurs when the photon energy is not sufficient to eject an electron from its shell. In this case, the direction of the photon is changed without changing its energy. At higher photon energies, *Compton scattering* is the most contributing factor. In this process, the photon energies are much higher than the binding energies of the electrons in the shells. Therefore, an electron is emitted from the shell and characteristic radiation will occur (as with photoelectric absorption). In this process, the photon only loses part of its energy, and will continue its path through the material, possibly starting another interaction process. At the highest photon energies, a fourth interaction called *pair production* will take effect (and dominates for photons with gamma ray energies). In this case, the high-energy photon will create an electron-positron pair in the atom, each travelling further in different directions. When the positron meets another electron, two photons with high energy (gamma rays) are formed in opposite directions. This effect may occur more often in industrial imaging, where high-energy radiation is involved for materials with high atomic numbers.

Given a material, the combined effect of its absorption by the aforementioned mechanisms can be collapsed into an *attenuation rate* μ (and here also includes the material mass density). The absorption of X-rays by a homogeneous material with an attenuation rate μ is given by the *Lambert-Beer law*:

$$I = I_0 \exp\left(-\mu \Delta x\right).$$

Here, I_0 is the *intensity* (proportional to the number of photons) of the incident X-ray beam and Δx is the thickness of the object. An illustration of the Lambert-Beer law is given in Figure 1.6a. The Lambert-Beer law states that the intensity of the beam decays exponentially with the thickness of the material (and the attenuation coefficient). Therefore, a higher incident flux cannot easily compensate for an increase in thickness. If the object being imaged is not homogeneous, the attenuation coefficient becomes dependent on the positions over which the ray passes through the object. Therefore, in this case the Lambert-Beer law reads

$$I(s) = I(0) \exp\left(-\int_0^s \mu(x) dx\right),$$

where $\mu(x)$ is the attenuation coefficient at location x, and s is the length of the ray through the object.

The Lambert-Beer law theoretically quantifies the intensity of the beam after it has traversed the object. To measure this quantity, an X-ray detector is needed. Usually, these detectors do not measure the X-rays directly, but rather contain material that converts the X-rays to other detectable radiation. Nowadays, *flat-panel scintillator* detectors are the most commonly used detector type, where the scintillator converts the X-rays to visible light, which in turn can be detected by a charge-coupled device (CCD) that converts the visible light through photodiodes

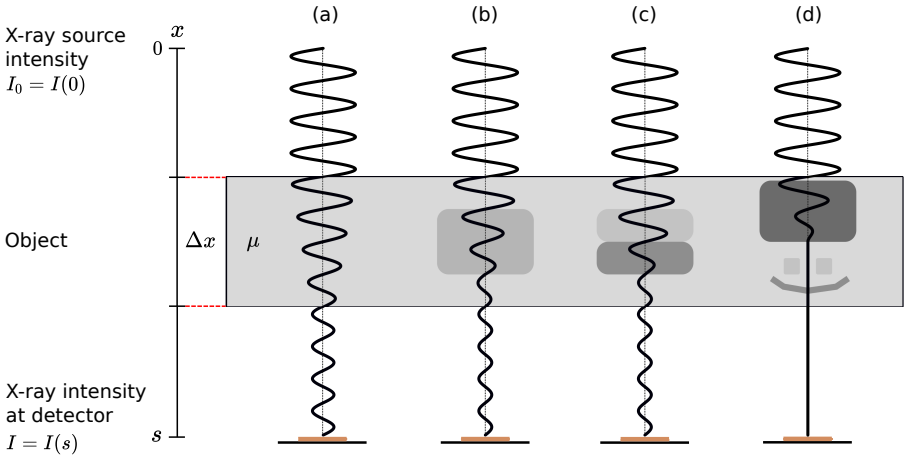


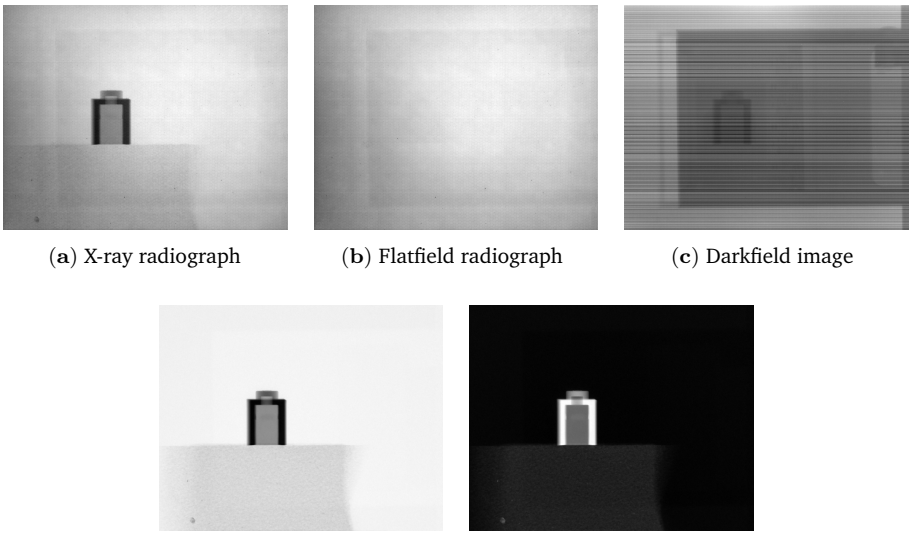
Figure 1.6: Illustration of the Lambert-Beer law: (**a**) ray passing through a homogeneous object with thickness Δx and attenuation μ, (**b**) ray passing through an inhomogeneous object with additional absorption, (**c**) ray passing through a different inhomogeneous object resulting in the same intensity at the detector, (**d**) ray passing through a highly absorbing object, leading to photon starvation and removing any interesting information about the structure of the object.

into electronic signals. The *detector efficiency*, the rate of incoming photons that are detected, depends on the thickness and the material of the scintillator.

According to the Lambert-Beer law, the detector yields an image that approximately gives the intensity of the incident X-rays. For further analysis, the image is first normalized by applying a *flatfield correction* (with an X-ray image without the object) and a *darkfield correction* (with an image with photon count offsets when the X-ray is not in use) to obtain the transmission image. After this, the image is linearized by applying a negative logarithm to obtain the absorption image. The correction process is visualized in Figure 1.7. The result for an X-ray traversing through the object is approximately:

$$-\ln \frac{I(s)}{I(0)} = \int_0^s \mu(x)dx. \tag{1.1}$$

The images obtained by X-ray radiography are 2D projections, which can be analyzed and processed, such as decision-making on the presence of a foreign object in a 3D food product [191]. However, in the resulting *radiograph* (or *X-ray image*), information about the third dimension (perpendicular to the detector plane) is lost. There are situations where prior information can be exploited well enough to detect desired features, such as the presence of bones. In other situations, absorption of a certain material and thickness cannot be distinguished from absorption from another combination of material and thickness. This is the *superposition problem* (Fig. 1.6b and c). On top of this, when highly absorbing objects are involved in the radiograph acquisition, there is the issue of *photon starvation* (Fig. 1.6d). The



(**a**) X-ray radiograph (**b**) Flatfield radiograph (**c**) Darkfield image

(**d**) Corrected radiograph (**e**) Log-corrected radiograph

Figure 1.7: Demonstration of flatfield correction on radiographs of a small brick. (**a**) The radiograph obtained with the object in the field of view. (**b**) Flatfield radiograph without the object in the field of view. (**c**) Darkfield image without the X-ray tube. Note that the recorded values are in reality much lower compared to the values in radiographs (a) and (b). (**d**) Corrected radiograph using the flatfield and darkfield images. (**e**) Log-corrected radiograph resulting in the typical radiograph representation. Note that the acquired images contain some sort of afterglow resulting from X-ray exposure during a previous acquisition, see the panel afterglow in (a-c). The flatfield correction in (d) removes the panel.

number of emitted and detected photons each follow a *Poisson distribution*. The average number of emitted photons is typically sufficiently high to approximate it by a Gaussian distribution and yields a smooth flatfield image. However, this is not the case when a highly-absorbing object prevents most photons from reaching the detector. In this case, the intensities on the detector pixels may vary heavily, and are likely to lead to high *noise* in the radiograph such that desired features are difficult or impossible to extract. Altering the X-ray source properties, such as the current, the peak voltage, or the *exposure time* may alleviate this problem. However, in many cases, these settings cannot be changed because of time, energy and dosage constraints, or the absorption power of the object is unlikely to be overcome.

1.3 Tomography

To gain better understanding of the inner structure of an object, the procedure of *tomography* noninvasively generates images containing the inner structure of an object. There are many different types of tomography with different material penetration and interaction mechanisms, such as magnetic resonance imaging

(MRI) and ultrasound tomography. In this dissertation, whenever tomography is discussed, we will focus on X-ray transmission *computed tomography* (CT). Since the proof of concept of CT scanning in 1969 by Hounsfield and the subsequent first clinical examination with CT in 1972, the methodology has been hugely important in getting good insight into the three-dimensional condition of (parts of) the human body. Additionally, it finds many applications in scientific research, geological studies [97], cultural heritage [44, 151, 201, 255] and (airport) security [193, 204]. In industry, important applications are materials characterization, nondestructive testing and metrology [55], and tomography also finds more and more applications in food processing (such as quality control of pears [174]).

In X-ray CT, an object is exposed to X-rays from a source, while a detector captures the unabsorbed photons in a resulting radiograph. This process is repeated under different angles, by rotating the object with respect to the source and detector (Fig. 1.8). In some instances, such as with CT scanners in hospitals where patients are imaged, it is more practical to rotate the source and detector rather than the object itself (Fig. 1.9). The precise locations of the source, detector and the object in the direction of the beam are recorded during the scan and are part of the complete *scanning geometry*, which can be circular but also helical or spiral when the object is large.

In a more formal context, the object of interest is modelled by the function $f : \mathbb{R}^n \to \mathbb{R}$ which maps the location to a scalar value that represents the attenuation in that location. In case of a two-dimensional object, we have $n = 2$. The *Radon transform* $\mathcal{R}$ maps the object function f to its projection function $P : [0, \pi] \times \mathbb{R} \to \mathbb{R}$

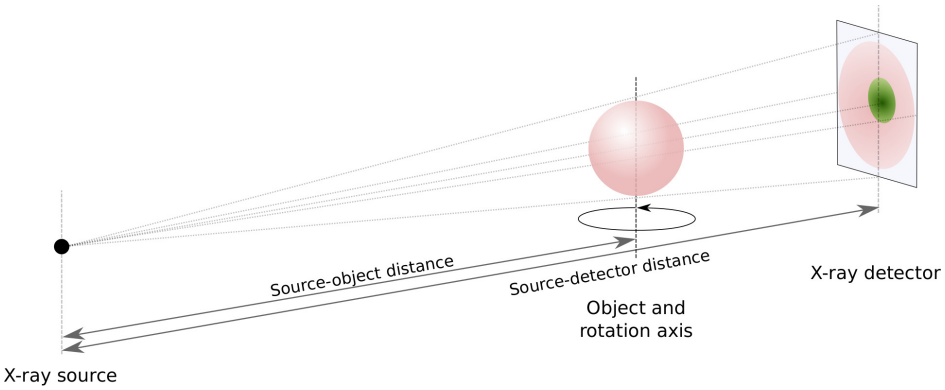


Figure 1.8: The process of obtaining multiple X-ray projections by rotating the source and the detector.

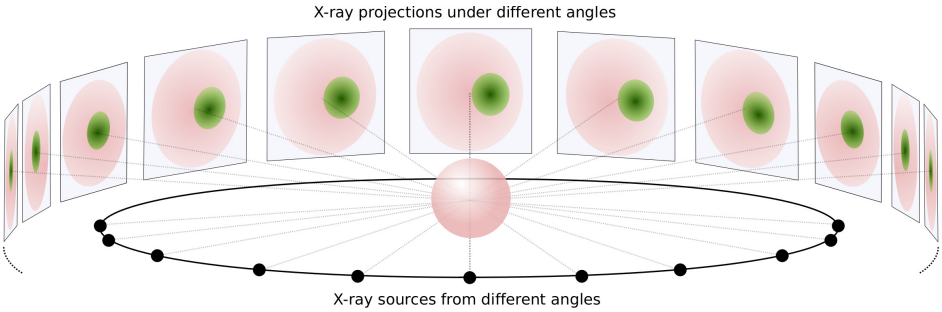


Figure 1.9: The process of obtaining multiple X-ray projections by rotating the object, with the source on the left and the detector on the right.

[84, 209, 232], and is given by the line-integral

$$\begin{aligned}\mathcal{R}[f](\theta, t) &= \int_{L_{\theta,t}} f ds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\delta(x_1\cos(\theta) + x_2\sin(\theta) - t)\ dx_1\ dx_2.\end{aligned}$$

where $L_{\theta,t}$ is the line possibly passing through the object, given by $L_{\theta,t} = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1\cos\theta + x_2\sin\theta = t\}$. Note that for a given line, so for fixed θ and t, the Radon transform is equivalent to the Lambert-Beer law, with the left-hand side being the absorption image (see Fig. 1.10a). The Radon transform of an object f is also referred to as the *sinogram* (Fig. 1.10c). For a given object f, the acquired projections can be viewed as a set of line integrals in which detector position t and angle θ are varied:

$$P(\theta, t) = \mathcal{R}[f](\theta, t).$$

The goal of *tomographic reconstruction* is to retrieve the function f from its Radon transform, which is an *inverse problem*. The series of collected radiographs of an object are used as input for a *reconstruction algorithm* [35, 112]. There are various classes of reconstruction algorithms. The first class is the group of *analytical* algorithms in which explicit analytic expressions are derived for the material attenuations in terms of projections. To retrieve the object from its sinogram, intuitively one can smear out each projection back over the object space. The simplest *backprojection* (Fig. 1.10b) is given by the following:

$$f_{\mathrm{BP}}(x_1, x_2) = \int_0^{\pi} P(\theta, x_1\cos\theta + x_2\sin\theta)d\theta.$$

This backprojection operation assumes that all pixels contributed equally to the projections [286]. Therefore, the resulting reconstructed image is typically blurry (Fig. 1.10d). To prevent this, the projection data are usually filtered in the

Fourier domain by a filtering function $h : \mathbb{R} \to \mathbb{R}$, such that the lower frequencies (accounting for smooth edges) have less contribution than the high frequencies (accounting for finer details) to the backprojection. This leads to the *filtered backprojection* (FBP) method (Fig. 1.10e), where the ramp filter in the Fourier space is used, defined by $\overline{h}(\omega) = |\omega|$, and the complete backprojection step including filtering is given by

$$f_{\mathrm{FBP}}(x_1, x_2) = \int_0^{\pi} \int_{-\infty}^{\infty} P(\theta, \omega) e^{2\pi i \omega (x_1 \cos\theta + x_2 \sin\theta)} h(\omega) d\omega d\theta.$$

To remove noise from the reconstruction, the high frequencies are removed by employing the Ram-Lak filter defined by $h(\omega) = |\omega|$ if $|\omega| \leq c$ and $h(\omega) = 0$ if $|\omega| > c$. Similarly, other filters can be employed as well. In the resulting FBP algorithm, the projection $P(\theta, t)$ is repeatedly *backprojected* for each angle θ. Since the detector positions and the scanning angles are discrete in practice, the integrals in the FBP expression above are substituted by the corresponding summations:

$$f_{\mathrm{FBP}}(x_1, x_2) \approx f_{\mathrm{FBP}}^{\mathrm{disc}}(x_1, x_2) = \sum_{\theta \in \Theta} \sum_{t \in T} P(\theta, \omega) e^{2\pi i \omega (x_1 \cos\theta + x_2 \sin\theta)} h(\omega) d\omega d\theta.$$

The FBP method is designed for parallel-beam geometries, while most CT setups have a fan-beam or cone-beam geometry. However, for fan-beam geometries the method can be employed by applying coordinate transformations, while generalization to three-dimensional cone-beam geometries results in what is known as the Feldkamp-David-Kress (FDK) algorithm [87].

FBP-derived methods are fast, easy to implement, and perform well on data that contain low noise and are near-complete (no missing angles). Therefore, these have been the method of choice until recently. FDK is the most practical analytic algorithm [286] for circular cone-beam geometries.

Another class of algorithms is that of *algebraic methods*, in which the Radon-transform formulation is changed into a system of equations, and the object function domain is discretized as a collection of *voxels* in a grid. The voxels can be modelled as a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, where n is the number of voxels. By putting all measurements in a vector $\boldsymbol{p} = (p_1, \ldots, p_l)$, where l is the number of detector pixels times the number of angles, the problem of tomographic reconstruction can be formulated as finding an $\boldsymbol{x}$ such that the following equality holds

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}.$$

Here, $\boldsymbol{W}$ is the *projection matrix* in which entry w_{ij} indicates how much a ray corresponding to measurement i contributes to the absorption in voxel j (which in the modelling depends on voxel size, ray direction and type of ray-voxel intersection computation). Recovering an accurate image of the scanned object from data p is usually difficult for several reasons. First, the number of measurements l may be too

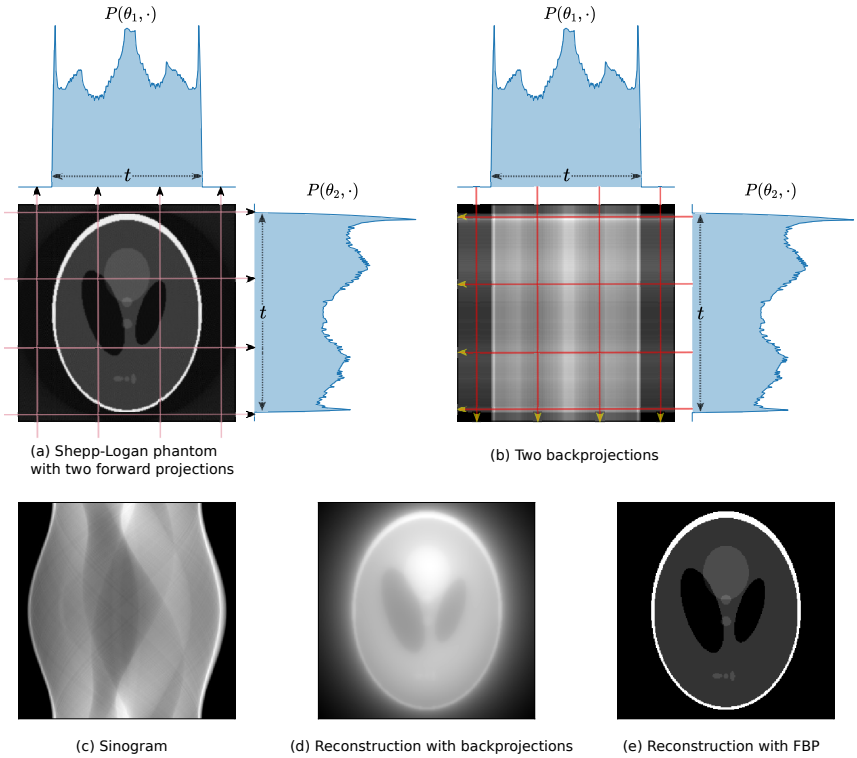


Figure 1.10: Illustration of tomographic reconstruction algorithms on a 512×512 Shepp-Logan phantom. (**a**) Forward projections under two angles on the Shepp-Logan phantom. White areas indicate regions with high attenuation. These are reflected in the large peaks in the graphs of the projection functions $P(\theta_1, \cdot)$ and $P(\theta_2, \cdot)$. (**b**) Backprojections of the two projections onto the image domain. (**c**) The function values of all 256 projections under equidistant angles combined into a sinogram. (**d**) Reconstruction using backprojections resulting in a blurry figure. (**e**) Filtered backprojection (FBP) reconstruction with significantly less blurry edges.

low compared to the number of voxels. This leads to an underdetermined system, which may have infinitely many solutions. Likewise, the number of measurements may be too high, leading to an inconsistent system with no solution that satisfies all measurement equations. Furthermore, the measurements may be corrupted by noise (for instance, electronic noise in the detector). In this case, $\boldsymbol{p}$ could be written as $\boldsymbol{p} = \overline{\boldsymbol{p}} + \boldsymbol{e}$, where both the uncorrupted data $\overline{\boldsymbol{p}}$ and the noise contribution $\boldsymbol{e}$ are unknown. Working with the known data $\boldsymbol{p}$ instead of $\overline{\boldsymbol{p}}$ is likely to render a potentially solvable system unsolvable.

Instead of solving the above equation, the reconstruction problem is often posed as an optimization problem by finding a minimum-norm least-squares solution $\boldsymbol{x}^*$

that minimizes the *data discrepancy* or *data fidelity* term:

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ ||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||.$$

Algebraic iterative methods repeatedly improve the reconstructions with respect to the data discrepancy above, by alternating between forward and backward operations through matrix $\boldsymbol{W}$. One of the most commonly used methods is the Simultaneous Iterative Reconstruction Technique (SIRT). In this algorithm, the current reconstruction is forward projected, after which the difference between this forward projection with the measured data is backprojected using the transpose of the projection matrix. The update step for SIRT is given by the following:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{C}\boldsymbol{W}^T\boldsymbol{R}(\boldsymbol{W}\boldsymbol{x}^{(k)} - \boldsymbol{p}).$$

The diagonal matrices $\boldsymbol{C}$ and $\boldsymbol{R}$ contain normalizations with respect to the sums of columns of $\boldsymbol{W}$ and rows respectively. In the ART method, the update step is carried out by taking only one ray each time. For the SART algorithm [15], this is extended to a full update for every projection angle. However, SIRT has a more stable convergence than ART, but an adequate solution takes longer to reach.

Iterative algebraic methods can deal better with inconsistent, incomplete, and noisy projection data than analytical methods. Furthermore, these methods are also flexible with respect to the acquisition geometries, as these can be modelled in the projection matrix $\boldsymbol{W}$. However, the iterative methods require a proper stopping criterion, as running for too long can lead to overfitting to noise and therefore noisy reconstructions may appear. *Variational methods* combine the above optimization problem with a *regularization* term in which prior knowledge about the object can be incorporated:

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||_2 + \lambda\mathcal{F}(\boldsymbol{x}).$$

In this formulation, $\mathcal{F}$ is the regularization operator and $\lambda \in \mathbb{R}_{\geq 0}$ is a parameter that controls the importance of the regularization with respect to the data-fidelity term. Two often-used regularization methods are Tikhonov and Total Variation minimization (TVmin) regularization. The former promotes solutions with small norms and reduces noise by setting the regularization function to $\mathcal{F}(\boldsymbol{x}) = ||\boldsymbol{x}||_2^2$, while the latter promotes sparsity of the gradients and smooth regions (yielding more piecewise smooth solutions), by setting $\mathcal{F}(\boldsymbol{x}) = ||\nabla\boldsymbol{x}||_1$. Regularization can steer towards more robust solutions containing less noise, but needs to be adjusted to each specific problem and requires the regularization parameter to be properly chosen.

Prior knowledge can also be incorporated by restricting the solution space. In *discrete tomography*, the resulting image has discrete values, meaning that all values x_i are contained in a fixed discrete set R. This type of tomography is a vast

subfield on its own [121, 122], as the prior knowledge can be exploited for more efficient and robust algorithms (in terms of accuracy, noise handling and required number of projections [30]) when the object of interest is discrete, or even binary [29, 176].

While iterative algebraic methods have advantages over analytic methods, their major disadvantage is that the repeated computation of the forward projections and backprojections makes them *computationally expensive*. However, with the increase in computation power of modern computers, most notably the Graphical Processing Units (GPU), this problem is alleviated. This is reflected in the fact that algebraic methods have been the reconstruction method of choice since 2009 commercially [292], while before that time these were FBP-derived methods. On top of that, advances in parallel computing and efficient splitting of the tomographic reconstruction problem make the computation of algebraic methods more tractable [50].

Tomography overcomes the superposition problem found in 2D radiography, and a precise reconstruction of a product can be made and inspected with this technique. A number of parameters influence the quality of the reconstruction of a given object. During the acquisition, the quality of the resulting radiographs is generally improved by a higher current, peak voltage and exposure time. The detector resolution and the rate at which the detector can detect photons also influence the radiograph quality. For the reconstruction, the given computation time, stopping criterion and the type of algorithm (along with its parameter settings) all influence the result. The reconstruction quality is constrained by the allowed acquisition and computation time, the X-ray dosage limit and the energy budget. As opposed to medical CT, there is less concern about the dosage constraints in industrial applications, allowing for more precision and higher resolution in the reconstruction [55]. Nevertheless, in high-throughput applications, the acquisition times and computational times are severe limiting factors. These constraints can be met by for example reducing the exposure time and increasing the photon flux, and by adjusting the reconstruction algorithms, but this may lead to poor reconstructions. Therefore, even though CT gives more object information than 2D radiography, it is considered to be less suitable for online inspection [186].

1.4 Spectral X-ray imaging

X-rays consist of photons of different wavelengths. Since photons with different energies can exhibit different interactions with matter, this can possibly contribute to a deeper analysis of the imaged object. When the notion of X-ray energy is accounted for in the radiography acquisition, this is referred to as *spectral X-ray imaging*, and encompasses a wide range of approaches.

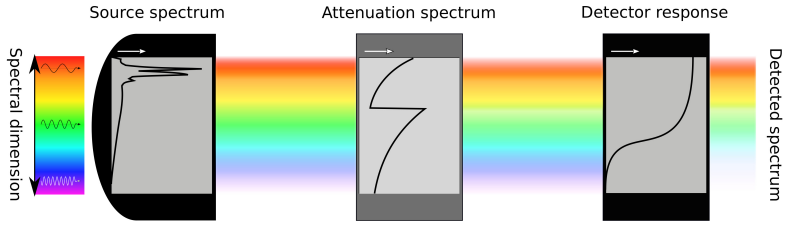


Figure 1.11: Schematic view of the different effects on a polychromatic beam.

To comprehend the methods, it is first important to understand where and how photon energies come into play (see Figure 1.11). First of all, the X-ray source emits a *polychromatic* X-ray beam. The distribution of energies of the photons emitted during the X-ray generation, also called the *source spectrum*, depends on the anode material and the applied peak voltage. An increase in peak voltage increases both the range of photon energies and the overall photon counts per energy. At the same time, the anode material determines the characteristic peaks that emerge from characteristic radiation due to the inner electron shell transitions in the atoms of the anode material when bombarded with fast electrons.

When X-ray photons interact with materials on their path, the resulting absorption is dependent on processes such as photoelectric absorption, pair production and Rayleigh and Compton scattering. The combined contribution of these processes to the beam attenuation depends both on energy and material type. See Figure 1.12 for a number of example spectra. The total attenuation μ_α of the X-ray by a material α for different energies is given by its *attenuation curve*, and is unique to each elementary material. Of particular note in these curves are the *absorption edges* caused by the photoelectric absorption. Due to the ejection of electrons in the shell and the subsequent electron transition from a higher shell, a characteristic radiation with fixed energy is emitted. As an electron will only be ejected when the incoming photon energy exceeds its binding energy, a sharp edge in the attenuation curve will appear. The electron binding energy is proportional to the atomic number, and inversely proportional to the distance to the nucleus of the atom. The electrons reside in shells with a fixed distance to the nucleus, with the innermost shell (K-shell) being the shell with the lowest distance. The binding energy is highest for the K-shell, and therefore requires the highest incident photon energy to be ejected. However, when the energy of the incident photon does not meet the binding threshold, no absorption will occur in this shell. While absorption edges also appear for other electron shells (L_1-, L_2-, M_1- edges, among others), the corresponding *K-edge* is located at the highest energies and provides the highest difference in attenuation, which makes it useful for applications dependent on this contrast (*K-edge imaging*) [94, 249]. Note that the attenuation curve is unique for each element, and so are its K-edge and other absorption edge locations. For elements with a high atomic number, such as metals, the K-edges are located at higher energies. These elements are therefore much more distinguishable than those

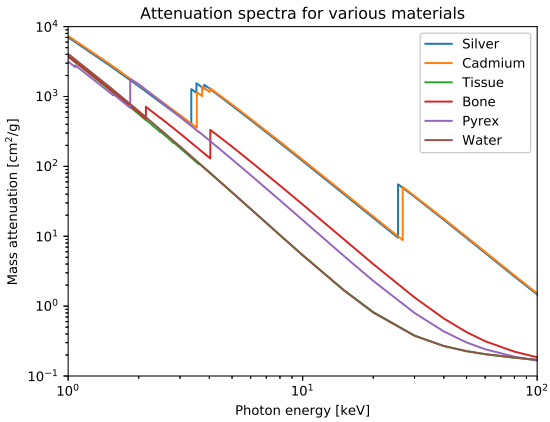


Figure 1.12: Attenuation spectra for various materials. The spectra are unique to each material, but some materials are easier to separate than others: silver and cadmium mostly differ through the slightly different locations of their absorption edges, and the difference in the spectra of water and tissue are almost indiscernible.

of elements with low atomic numbers, whose K-edges may not even be located within the X-ray wavelength range.

As can be observed from the attenuation curves, X-ray photons with low energies tend to be attenuated more strongly than those with high energy. For this reason, when a beam traverses through matter, the mean photon energy shifts to a higher value. This phenomenon is known as *beam hardening*. In standard CT, the X-ray beam is assumed to be *monochromatic*, consisting of photons with the same energy. However, because it is polychromatic in practice, the assumption of monochromaticity leads to *beam hardening artefacts* in the subsequent reconstruction, showing as dark and less attenuating regions in the CT reconstruction. To avoid this, ideally, the beam should be monochromatic, but a sufficiently high flux of such a beam can only be achieved at synchrotron facilities. To mitigate beam hardening artefacts resulting from a wide energy range, the low-energy photons in the incident beam are usually removed by *filtering* the beam with a (metallic) material with known spectral properties. With this approach, beam hardening is actually used to an advantage. Nevertheless, filtering does not completely eliminate beam hardening artefacts but reduces them, and longer scanning times are needed for similar signal-to-noise ratios.

The interactions of different X-ray photon energies with matter can be exploited in various ways. The most elementary way is to scan an object with beams of different energies. In the case of two energies, this is referred to as *dual-energy X-ray absorptiometry* (DEXA). This can be achieved by applying different values for the X-ray tube voltage (and possibly some additional beam filtering). For each

detector pixel, this methodology results in two intensity values corresponding to two different voltages. As these define a point in a two-dimensional attenuation space, the projected materials are likely to be more separable from intensity value tuples from other material compositions than in radiographs from single peak voltages.

Instead of modifying the energy distribution in the polychromatic beam from the source, in *detector-based approaches* energetic information of photons reaching the detector can be directly obtained. *Dual-layer* scintillator detectors consist of two layers on top of each other. These spectral detectors have different photoelectric absorptions depending on their sensor materials, and the upper layer typically, on average, detects the low-energy photons while the bottom layer detects the high-energy photons [99, 118]. The advantage of such detectors is that these simultaneously capture images of the same object, avoiding any spatial differences due to motion between subsequent captures.

Similarly, as opposed to regular charge-integrating detectors (also known as energy-integrating detectors), (hybrid) *photon-counting detectors* register single incoming photons and are able to measure the energy of each photon individually [94] instead of integrating the energies of all incident photons [249]. Compared to the two-step process in charge-integrating detectors, a photon-counting detector consists of a semiconductor layer, in which incident X-ray photons generate electron-hole pairs. A bias voltage between a cathode on top of the detector and an anode on the other side separates the charge carriers to the contacts for electronic readout [263]. The process directly converts photons into electric charge, allowing for faster detection speed than with charge-integrating detectors. On top of this, the electronics in these detectors allow for spectral separation of photons into spectral *bins*, in which the number of photons between a preset energy range is counted. Compared to the first dual-energy methods, which were available in the '70s, photon-counting technology is relatively new. Nevertheless, the ability to simultaneously collect radiographs at different energies holds a huge potential for the future [94]. Therefore, several research groups worldwide are developing readout electronics for hybrid photon-counting detectors [91], such as the PILATUS chips at PSI, Medipix [26] and Timepix chips at CERN, PIXIE chips [36] at INFN and the PXD18k at AGH University.

An alternative class of spectral detectors operate by measuring the deposited energy in every pixel during a given exposure time and repeating this for many frames at high rates, yielding a binned energy distribution for each pixel [78]. For this reason, these *hyperspectral* X-ray detectors can attain a high spectral resolution. State-of-the-art detectors, such as the HEXITEC family of detectors [282] and SLcam [213], provide data containing up to hundreds of spectral energy bins and a spectral resolution of less than 1 keV, providing a rich spectral footprint of the measured object. Nevertheless, this relatively new technology requires comparatively longer acquisition times, as well as powerful data processing frameworks, which are currently in development [18].

For photon-counting detectors, different semiconductor materials like silicon (Si), gallium arsenide (GaAs), cadmium telloride (CdTe), or cadmium zinc telluride (CdZnTe) can be used. While Si-based detectors are generally cheaper, it suffers from low absorption, especially at higher energies where the detector layer becomes transparent. In comparison, CdTe and GaAs are more costly but have better absorption at higher photon energies. The extent to which the detector is able to register photons at different energies is given by the *spectral detector efficiency* or *detector response*.

The Lambert-Beer equation (where the line from the source to a the detector pixel is now given by ℓ) can generalized by incorporating the energy dependence :

$$I(E^{\max}) = \int_0^{E^{\max}} I_0(E)D(E)e^{-\int_\ell \mu(x,E)dx}dE.$$

In this formulation, the energy is integrated up to the maximum photon energy $E^{\max}$, which typically is the peak voltage energy or the maximum energy registered by the detector. Now, I_0 models the source spectrum as an incident photon rate as a function of energy, while D is a simple form of modelling the photon detector efficiency. In some formulations, I_0 and D functions are combined into a function known as the effective spectrum. The attenuation μ is now a function, dependent on both location and energy. As opposed to the monochromatic Lambert-Beer law, due to the energy integral, this expression cannot be converted into a linear right-hand side. This means that the data dependency on the energy-dependent attenuation coefficients is nonlinear (as opposed to Equation (1.1)).

While dual-energy absorptiometry has found many applications in food processing [16] and medical imaging, spectral X-ray imaging is especially promising for tasks involving the use of *contrast agents* which enhance the contrast between certain parts of an object. Examples of these include iodine and gadolinium, which have K-edges in the energy range of X-rays used in diagnostic radiography. Photon-counting detectors are valuable when at least two contrast agents need to be detected simultaneously [264]. More generally, spectral X-ray imaging is useful for material identification in an object [96], especially with multiple different materials. Therefore, it opens up new possibilities such as multi-energy X-ray testing [192] and multi-material decomposition [274], for example leading to improvements in tissue discrimination [293].

Even though the photon-counting detector technology is promising for energy-resolved imaging, there are still several challenges to overcome [293]. Firstly, an incident photon may be registered in a location other than where it hits the detector. This can be caused by fluorescence or Compton scattering, or by an effect known as *charge-sharing*, where the incident photon generates charges close to the border of a pixel that may be partly registered by an adjacent detector

pixel. Secondly, when multiple photons arrive with too little intervening time, an effect known as *pulse pile-up* may occur [54, 99]. In this effect, the photons will be registered as one photon with an energy equal to the sum of the individual photons. Even though these detectors are able to detect hundreds of millions of photons per second per square millimeter, pulse pile-up is likely to occur. Both charge-sharing and pulse pile-up can lead to artefacts since the spectral distribution of photon counts is perturbed. Furthermore, a trade-off is required for the spatial resolution of the detector. Increased resolution increases the effects of fluorescence, charge-sharing and noise thresholding, and decreases the effects of pile-up [67]. Also, pixels may yield a different output with the same incident photon energy distribution. Therefore, accurate equalization and calibration need to be developed and carried out.

Apart from improvements with respect to energy differentiation and speed, there are more benefits from photon-counting detectors. The thresholding mechanism of photon-counting detectors prevents contribution of electronic noise [264]. This means there is no contribution of dark current, and darkfield correction is therefore not needed. Additionally, depending on the task carried out on the resulting spectral radiographs, the additional spectral information may reduce the total dose compared to regular radiograph acquisition. On top of this, due to the direct photon conversion, photon-counting detectors generally have much higher spatial resolution than energy-integrating detectors [159].

While hyperspectral detectors generate data with high spectral resolution and therefore circumvent the issue of tuning spectral thresholds to obtain a desired result, the registration of photon counts in every small energy bin yields huge data volumes. Additionally, there is the issue of spectral redundancy [60], indicating that spectrally adjacent X-ray images from the data cube are likely to be very similar. Therefore, given a specific task on the hyperspectral data cube, it is not clear in which spectral bins the important features are located.

1.5 Spectral tomography

X-ray CT has yet to reach its full potential as spectral information of X-rays is not always recorded [274] or used in reconstruction algorithms. Similar to how radiography paved the way for X-ray transmission tomography, spectral radiography opens up possibilities for *spectral CT*, where the input radiographs are energy-resolved.

Exploitation of spectral characteristics in CT scans can be done in various ways, as demonstrated in Figure 1.13. Historically, the first and most simple spectral CT setup is to perform the same scan twice with different tube voltages. The possibility of artefacts resulting from object motion between the scans and the increased scanning time or X-ray dose are obvious downsides of this approach. Therefore, to reduce the disadvantages of sequential scanning, rapid *tube voltage switching*

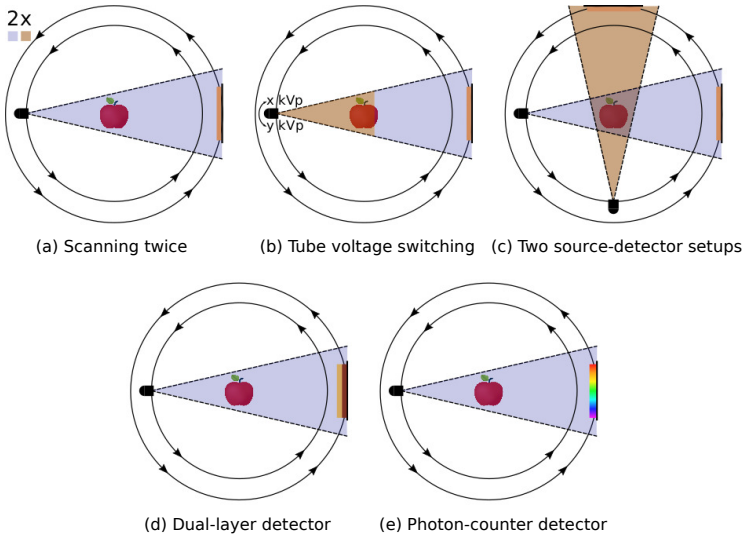


Figure 1.13: Schematic representation of various spectral CT setups.

sources were developed [11, 51, 55, 99]. However, even though the tube voltage switches at high speed at every scanning angle, this still results in considerably more scanning time than with conventional CT scanning, and relatively slow potential switching may still lead to motion artefacts. Another approach is to combine two source-detector combinations into one setup, known as *dual-source CT*, each source with a different peak voltage [11, 51, 55]. This method requires a more expensive setup, requires spatial co-registration of the obtained data, and the scattering resulting from the two different beams may increase the noise in the projections obtained by either detector. However, the setup of each detector-source combination can be configured independently. Both fast-switching potential dual-energy CT and dual-source CT are well-established in clinical practice. More recently, integration of *dual-layer detectors* in clinical settings has also been established [11, 40], with the benefit of simultaneously obtaining perfectly spatially aligned radiographs with different average photon energies [233]. Nevertheless, dual-layer detectors are more costly and designed for specific tasks, which makes these inflexible with respect to various material separation tasks [99]. In medical imaging, dual-energy CT is often used for contrast enhancement, artefact reduction, material decomposition and radiation dose reduction [233]. For industrial CT, dual-energy CT mainly improves reconstructions in terms of beam hardening artefact reduction and feature detection with low contrast [55]. The advent of photon-counting detectors enables *multi-spectral* CT, which generalizes the concept of dual-energy CT to more than two energy bins. As photon-counting detectors record data into narrower energy bins, the polychromatic beam drastically reduces beam-hardening artefacts found in conventional CT. Despite this, industrial CT most often employs energy-integrating detectors [55]. In medical CT, it is expected that dual-energy and spectral CT will

be fully integrated into clinical practice and eventually replace conventional CT [99, 274].

In terms of modelling the spectral reconstruction problem, we take the spectral Lambert-Beer law and integrate over the energy interval $[E_c^{\min}, E_c^{\max}]$ of the c-th measurement with $1 \leq c \leq C$:

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\int_\ell \mu(x,E)dx} dE. \tag{1.2}$$

Here, C is the total number of measurements, each with different energy properties. In *dual CT*, this number is $C = 2$. When prior knowledge of factors contributing to the attenuation is available, a basis decomposition can be incorporated that assumes the dependence of attenuation on location and energy can be written as a (finite) linear combination of functions that depend only on energy or location:

$$\mu(x, E) = \sum_{m=1}^{M} \widetilde{\mu}_m(E)\alpha_m(x).$$

Substituting this assumption into (1.2) yields

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\int_\ell \sum_{m=1}^{M} \widetilde{\mu}_m(E)\alpha_m(x)dx} dE. \tag{1.3}$$

The goal of dual CT is to obtain a decomposition of the reconstructed object in terms of the contributions α_1 and α_2. This can be approached in two ways. First, each spectral channel can be reconstructed separately. This can be done by assuming attenuation at an effective energy E_c^{eff} (such as the mean energy in the bin) to remove the energy integral and therefore the nonlinearity of the expression. Then the object can be reconstructed with conventional CT techniques. For every position x, this yields a system of equations

$$\begin{cases} f(x, E_1^{\text{eff}}) & = \sum_{m=1}^{M} \widetilde{\mu}_m(E_1^{\text{eff}})\alpha_m(x) \\ f(x, E_2^{\text{eff}}) & = \sum_{m=1}^{M} \widetilde{\mu}_m(E_2^{\text{eff}})\alpha_m(x) \end{cases}.$$

in which $f(x, E_c^{\text{eff}})$ is the solution of the tomographic inverse problem in channel c at position x. With this system of equations, an *image-based decomposition* $\{\alpha_m\}_{1\leq m\leq M}$ can be computed. Note that in dual CT this can only yield a unique solution if $M = 2$. While the image-based decomposition is relatively easy to implement, proper estimations for the attenuations and the photon flux are required. When the energy range in each channel is wide, the reconstructed images will suffer from beam hardening artefacts [274, 283].

Alternatively, by changing the order of the summation and the integral, a *projection-based decomposition* can be attempted:

$$I_c = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E)D(E)e^{-\sum_{m=1}^{M} \widetilde{\mu}_m(E) \int_\ell \alpha_m(x)dx} dE.$$

In this approach, the integrals $\int_\ell \alpha_m dx$ are computed first by solving the system of nonlinear equations, for instance by polynomial approximations [14, 63, 179, 266]. After the projection-based decompositions are obtained, each of these components are reconstructed separately, usually referred to as Virtual Monochromatic Images. Image-based decomposition is typically found to be faster, but projection-based decomposition yields superior decomposition results [179, 208].

The choice of base decomposition is mostly empirical. From the start of dual-energy CT, the decomposition into photoelectric effect (dominant at low energies) and Compton scattering contributions (dominant at high energies) [14] has been well-investigated. The former can be approximated by a function cubically inversely proportional to the photon energy, and the latter by the Klein-Nishina function. For materials that occur in the body (with an atomic number lower than 25), this decomposition is sufficient for the expression of the attenuation coefficients [99]. The energy-dependent attenuation coefficients can also be decomposed by attenuations of sufficiently different materials. For medical CT, the decomposition of water and bone is helpful, since most materials in the body have attenuation curves similar to either of those [118]. Alternatively, the K-edge in the diagnostic range of iodine makes this material suitable for material decomposition combined with tissue or water, since the attenuation of bone is similar to that of iodine-enhanced blood.

In advanced photon-counting CT, the number of preset spectral bins C is usually larger than two. In this type of spectral CT, the reconstruction techniques can be classified into two categories [143]. The first category is concerned with *multi-channel* reconstruction [64], which opts for the reconstruction of each spectral channel. In general, a channel in multi-channel CT does not necessarily refer to a spectral channel, but to any set of measurements of the same object with different acquisition settings. However, in this dissertation, we will mostly refer to a multi-channel image as a stack of images with each image resulting from different spectral properties.

The second category of spectral CT techniques is *material decomposition* from multi-channel data with more than two spectral bins. Many industrial and medical applications require more than $M = 2$ materials to be separated [172]. However, material decomposition with more than two materials poses challenges for standard binary reconstruction techniques [266], as found in dual CT imaging.

The difficulty of the multi-spectral CT problems is that these consist of many nonlinear equations, resulting in a nonlinear forward operator and an ill-posed problem. Approaches to these problems can be divided into two classes. The first class of methods are *two-step methods*. As a natural extension of a dual-energy CT approach, one subclass of two-step methods consists of methods that first reconstruct each spectral channel separately, after which an *image-based decomposition* (also known as post-reconstruction processing [27]) is carried out by

solving for the material distribution with more than two channels (i.e. $C > 2$):

$$f(x, E_c^{\text{eff}}) = \sum_{m=1}^{M} \widetilde{\mu}_m(E_c^{\text{eff}})\alpha_m(x). \tag{1.4}$$

The second subclass performs a projection-based material decomposition (also known as pre-reconstruction processing), followed by the reconstruction of each material separately. The problem here is that the material decomposition in the projection domain is a nonlinear inverse ill-posed problem [6]. A common drawback of all two-step methods is that the separation of the material decomposition and the image reconstruction steps imply loss of information during the first step, for which the other second step cannot compensate [292]. The drawback of image-based decomposition is that it may suffer from beam hardening artefacts in the first step if the energy bins are too wide. In contrast, projection-based material decomposition suffers from sensitivity to noise [251] and the nonlinearity of the problem, of which the introduced errors are computationally costly to alleviate [274].

The class of *one-step methods* is designed to avoid the information loss and related problems inherent to two-step methods. One-step methods attempt to find a solution to the system of equations resulting from Equation 1.3 directly to obtain material maps, but the complexity of the problem (for which finding an analytic solution is challenging, if possible at all) necessitates the use of iterative algorithms. Such algorithms are even more computationally heavy than standard iterative reconstruction methods for non-energy-resolved CT problems. Moreover, many state-of-the-art approaches are limited to only a few materials.

For spectral detectors with a higher energy resolution, the approximation by energy discretization resulting in Equation 1.4 is more accurate than with detectors with lower resolution. Additionally, this modality is useful for K-edge imaging as sharp edges are more clearly visible in the multi-channel reconstruction. Since the bins are narrow, beam hardening problems are also avoided. However, unless extremely high acquisition times are applied, the low signal-to-noise ratios in the channels will have a high impact on the reconstructions by the aforementioned methods.

To mitigate the impact of high noise levels in bins because of low photon counts [244, 320], prior knowledge about the solution can be incorporated into the reconstruction [60]. Regularization is required to obtain satisfactory solutions for the ill-posed spectral CT problems [274]. Similar to the standard iterative techniques, variational methods can be employed with spectral CT, especially in the reconstruction routines of two-step methods. Intensity-gradient sparsity constraints [61], such as Total Variation Minimization (TV) and Non-Local Means (NLM) are among the most commonly used methods in spectral CT [244]. Other variational regularization methods exploit the structural redundancy between the channels [61, 64], with hyperspectral images providing an even stronger structural

correlation between channels [288]. Numerous regularization methods have been proposed [274, 302], many of which are multi-channel enhancements of TV or NLM regularization. A few examples include spectral-means (based on NLM) [320], and Total Nuclear Variation (TNV) [64, 126] (which is an extension of TV, leading to low-rank solutions and therefore encouraging common edges in multiple images) [238, 320, 325], as well as parallel level sets methods such as directional Total Variation (dTV) [33, 64, 143]. Another approach is Total Generalised Variation (TGV) (designed for multi-modal and multi-channel imaging) [131, 147], which generalizes TV to higher orders of differentiation, and uses it for spectral CT to combine the channels. Both the sparse gradient and structural redundancy regularization methods can be combined into so-called spatiospectral regularization, for instance by combining spatial TV and spectral TGV [288]. All regularization methods require at least one parameter to be properly set and tuned, while the regularization method of choice is always dependent on the specific application. Regularization is generally easier to incorporate in one-step methods [27]. However, only a few methods incorporate spectral prior information into the one-step method that allows for the reconstruction of more material volumes.

Since photon-counting detectors are a new technology relative to energy-integrating detectors [284], multi-spectral CT has yet to find widespread use in medical and industrial fields. Nevertheless, the possibility to combine multi-spectral imaging with CT has enabled contrast agent imaging by exploiting energy-dependent K-edges of different materials [249]. Additionally, photon-counting detectors are also expected to improve spatial resolution [261]. Therefore, spectral CT and hyperspectral CT are beginning to find applications in medical imaging, such as enhancing tissue contrast [99, 296], and spectral CT is expected to be fully integrated into health care in the coming years [274]. This will further increase the interest in the underlying machinery and methods. However, it also requires further algorithmic developments for spectral image processing and CT reconstruction. At the same time, spectral and hyperspectral CT have great potential for security and industrial imaging [296]. Still, iterative reconstruction algorithms for photon-counting detectors are currently too computationally intensive for fast reconstruction [292], although the continued growth of interest [99] and algorithmic developments in methods with applications in medical CT are expected to improve this. Moreover, the difficulty of large-scale production of affordable high-quality photon-counting detectors still limits their commercial usage [159], and hyperspectral cameras are currently even more expensive. However, the first industrial X-ray cameras able to perform spectral imaging for food processing are scheduled to be launched in the coming years. For this reason, we will mainly focus on spectral X-ray CT in this dissertation. Furthermore, possible downsides of (hyper)spectral imaging, such as long acquisition times and computation times, may be alleviated by machine learning (covered in the next section) which is likewise getting much more attention in recent years in inverse problems and CT reconstruction.

1.6 Machine learning

Many routines in industrial imaging, for example X-ray based foreign object detection and crack detection, can in principle be carried out by humans. As the previous sections show, X-ray imaging and subsequent innovations enable doing these tasks in a noninvasive manner, but the work can be tedious, laborious or even harmful to humans. Moreover, the results are subjective, often irreproducible and possibly severely inaccurate. These aspects are some of the many motivating examples for the development of artificial intelligence, particularly with respect to food processing, with the aim of intelligent agents taking over the decision-making in various stages of the processing chain.

With machine learning, a model has the ability to learn (i.e. increasingly improve performance on a specific task) from data to carry out a task without being explicitly programmed to do so. Machine learning algorithms are concerned with predicting outputs y from inputs x using a function f. In supervised machine learning, this function can be estimated from labelled training data $\{(x_i, y_i)\}_{i=1}^m$ with input features $x_i \in \mathcal{X}$ and corresponding labels $y_i \in \mathcal{Y}$ [205], with $\mathcal{X}$ and $\mathcal{Y}$ being the input and output spaces, respectively. Machine learning aims to minimize the errors on unseen examples, as opposed to minimizing the errors on a training set only such as in optimization. Therefore, the training data should represent the unseen data well and the function f should both predict the training input features correctly as well as generalize to unseen data. As opposed to unsupervised machine learning, the labels y_i are known in *supervised* machine learning, which we will address in this section (for more details we refer to more technical and in-depth work [38]).

More formally, the labelled data space can be written as a product of the input and the output space $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$. The training data are assumed to be realizations of independent and identically distributed random variables $\boldsymbol{Z} = (Z_1, \ldots, Z_m)$, with $Z_i = (X_i, Y_i)$ from an unknown (joint) probability distribution denoted by $\mathbb{P}_Z$. Let $\mathcal{M}(\mathcal{X}, \mathcal{Y})$ be the set of all (measurable) functions from $\mathcal{X}$ to $\mathcal{Y}$. The *loss* function $\mathcal{L} : \mathcal{M}(\mathcal{X}, \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}_{\geq 0}$ measures the performance, i.e. how much the prediction $f(x)$ by a given function $f \in \mathcal{M}(\mathcal{X}, \mathcal{Y})$ differs from the target y with respect to a chosen metric. Given the data space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, the goal is to find a function $f : \mathcal{X} \to \mathcal{Y}$ that minimizes the *risk* $\mathcal{R}(f)$, which is the expected loss of the function f:

$$\mathcal{R}(f) = \mathbb{E}_{\mathbb{P}_Z}[\mathcal{L}(f, X, Y)].$$

Since the distribution according to $\mathbb{P}_Z$ is unknown, minimizing the risk over all measurable functions (resulting in the Bayes risk $\mathcal{R}^*$) is not possible. Therefore, the idea is to choose a *hypothesis set* or *hypothesis class* (a set of candidate functions) $\mathcal{F} \subset \mathcal{M}(\mathcal{X}, \mathcal{Y})$ and construct a learning algorithm to find a function $\hat{f} \in \mathcal{F}$ for

training data $\boldsymbol{Z}$ that minimizes the *empirical risk*, which is given by:

$$\hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z}) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{f}, Z_i).$$

The empirical risk $\hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z})$ measures the average loss on the training data $\boldsymbol{Z}$ and converges by the law of large numbers almost surely to the true risk $\mathcal{R}(\hat{f})$ as the number of training samples m goes to infinity. Given training data $\boldsymbol{Z}$, an *empirical risk minimizer* is a function $\hat{f}_{\boldsymbol{Z}}$ that minimizes the empirical risk over all functions in the hypothesis set $\mathcal{F}$:

$$\hat{f}_{\boldsymbol{Z}} \in \arg\min_{\hat{f} \in \mathcal{F}} \hat{\mathcal{R}}(\hat{f}, \boldsymbol{Z}).$$

To assess the performance of an empirical risk minimizer $\hat{f}_{\boldsymbol{Z}}$, the difference between the expected risk $\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})]$ (where expectation is with respect to $\boldsymbol{Z}$ and the randomness in the learning algorithm to choose $\hat{f}_{\boldsymbol{Z}}$ from $\mathcal{F}$) and the Bayes risk can be examined. This excess risk can be rewritten in the following way:

$$\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})] - \mathcal{R}^* = \underbrace{\mathbb{E}[\mathcal{R}(\hat{f}_{\boldsymbol{Z}})] - \inf_{f \in \mathcal{F}} \mathcal{R}(f)}_{\text{estimation error}} + \underbrace{\inf_{f \in \mathcal{F}} \mathcal{R}(f) - \mathcal{R}^*}_{\text{approximation error}} .$$

The first two terms form the estimation error and give the difference between the expected risk of the empirical risk minimizer and the risk minimizer over the hypothesis set $\mathcal{F}$. The last two terms together form the approximation error and give the difference between the risk minimizer over the hypothesis set $\mathcal{F}$ and the Bayes risk (over all measurable functions). In other words, the estimation error reflects how good the data are for finding the best prediction function in $\mathcal{F}$ for the entire data space (with respect to the data distribution). In contrast, the approximation error reflects how restrictive the hypothesis set $\mathcal{F}$ is for finding the optimal prediction function. As illustrated in Figure 1.14, the hypothesis set $\mathcal{F}$ is an important parameter. In case $\mathcal{F}$ is very small, the estimation error can become small, but the approximation error may become large, leading to *underfitting* of the dataset and, therefore, poor results on the training dataset. Conversely, if the hypothesis set is large, the approximation error is low, but the empirical error is high, leading to *overfitting* and bad generalization. Hence, given training data, a well-chosen hypothesis set is necessary for proper learning, of which the complexity provides a good trade-off for function estimation from the data and approximation of the other elements in the distribution. The analysis of this problem is complex and we will refer to literature for more details [38].

While there is no universal learning algorithm for every data distribution, there are many approaches to tackle this problem that provide suitable learning methods in practice. First, given a dataset and a function f, it is not clear whether f overfits

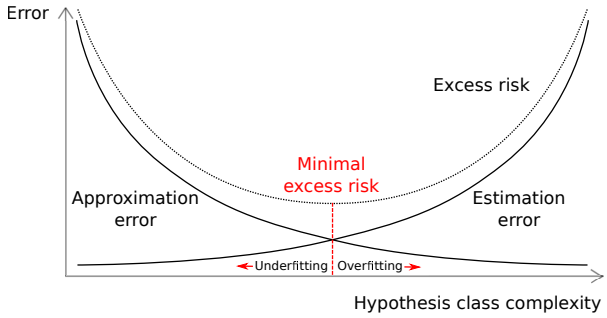


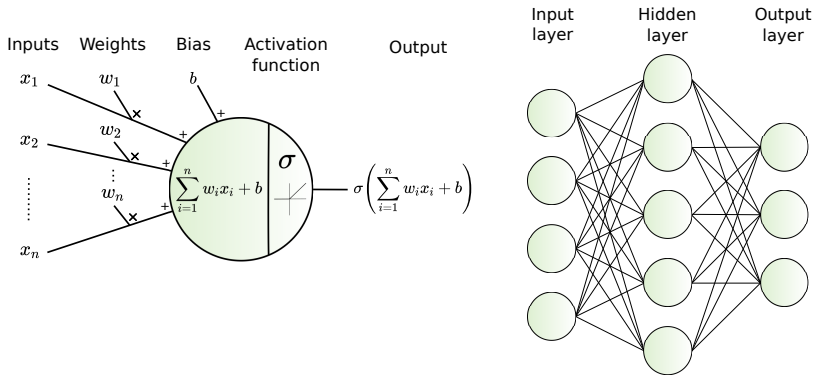
Figure 1.14: Schematic view of the estimation error and approximation error as a function of the hypothesis set complexity. The minimal excess risk is attained at a hypothesis class complexity for which the class contains a function that does not underfit the data, but no function that overfits the data.

or underfits the data. A common approach is to split the dataset into a training set and a validation set. Using a learning algorithm, candidate functions can be computed by minimizing the empirical risk on the training dataset. After that, the performance of a candidate function can be evaluated on the validation set. The idea is that a function overfitting on the training set can be observed by a weak performance on the validation set. Instead, the functions that perform best on the validation set can be used as candidates for the trade-off between estimation and approximation.

In traditional machine learning, there are many approaches for the hypothesis set. These include classes of functions described by decision trees, k-nearest neighbors and support vector machines, to name a few. Due to the increasing data size and complexity of their analysis, traditional machine learning methods have mostly become less suitable for many problems, including those related to food inspection [327]. However, one class of hypothesis sets that deal particularly well with these challenges is that of *neural networks*. Neural networks are inspired by the functionality of biological neurons in the brain. The basic blocks are the artificial *neurons* (Fig. 1.15a), that consist of a number of *input weights* $\boldsymbol{w} = (w_1, \ldots, w_n)$, a *bias* value b and an *activation function* σ, producing an *output value* $\phi(\boldsymbol{x})$ from its input values $\boldsymbol{x} = (x_1, \ldots, x_n)$ in the following way:

$$\phi(\boldsymbol{x}) = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right).$$

For the activation function σ, the Heaviside (threshold) function is an intuitive choice: only when the sum of the bias and the value of the inputs multiplied by the weights exceeds a certain threshold, the neuron will output a nonzero value. Alternative activation functions include the logistic functions and - more recently used - the Rectified Linear Unit (ReLU) activation function [102, 178], which is the identity for positive numbers and zero otherwise. By connecting the neurons,



(**a**) Artificial neuron structure

(**b**) Example network with fully connected layers

Figure 1.15: Building blocks of a neural network: (**a**) artificial neurons that weigh each input value and passes on the input value modified by a given activation function σ (as an example, the ReLU function is shown) and (**b**) the composition of multiple neurons into (fully connected) layers, and the composition of layers into a network of depth $d = 2$.

a *neural network* is formed (Fig. 1.15b). Different neurons can act on the same input, and can be organized in a *layer* with each neuron producing its own output $\phi_j(\boldsymbol{x})$ on the same input $\boldsymbol{x}$. By placing the layers ϕ_j^i in sequential order, with $i = 0, \ldots, d$ being the index of the layer and d being the depth of the network, a *multilayer perceptron* is created which takes input and propagates it through all the layers. The architecture of the network is given as $A = (C, \sigma)$, where σ is the activation function, and $C = (c_0, c_1, \ldots, c_{d-1})$ specifies the number of neurons per layer. The architecture A gives rise to the set Θ_A, which contains all value combinations of parameters of the network, i.e. the weights and biases. A realization of the network architecture A is a network $F_{A_\theta} : \mathbb{R}^{c_0} \to \mathbb{R}^{c_{d-1}}$ with values $\theta \in \Theta_A$ assigned to the parameters. The hypothesis set of a network architecture is then $\mathcal{F} = \{F_{A_\theta} : \theta \in \Theta_A\}$. The aim is to find a parameter configuration θ^* that minimizes the loss of the predictions of the realization of the network on the training set [188]:

$$\theta^* = \underset{\theta \in \Theta_A}{\arg\min} \sum_{i=1}^{m} L(F_{A_\theta}(x_i), y_i).$$

By feeding the neural network examples from a training set, the loss of the model's predictions and the *ground truth* can be computed. Since the functions ϕ_j^i in the neurons in the neural network are differentiable (almost everywhere), the gradient of the loss can be computed with respect to the network parameters using the *backpropagation* algorithm. Using a stochastic gradient algorithm, the error can be repeatedly used to re-adjust the parameters and optimize the network performance on the training dataset. Often used derivations of stochastic gradient descent for

this purpose are AdaGrad [74] and ADAM [146]. The training procedure is carried out and evaluated repeatedly on the validation set. To assess the result of the training procedure, the obtained network is eventually evaluated on a separate test set.

Deep neural networks have recently gained popularity, where *deep* refers to the networks having many layers [205]. Such networks allow for more complex functions and trainable parameters, and deep learning methods can in some cases therefore perform better data analysis than traditional machine learning methods [327]. Deep learning applications include speech recognition, natural language processing, classification, partial differential equation solvers, image generation and image processing [12, 327]. Consequently, these approaches are nowadays used for health care, recommender systems, fraud detection and self-driving cars, to name a few applications.

For imaging tasks, deep Convolutional Neural Networks (CNNs) have found substantial usage for tasks such as denoising, object detection, edge detection, classification and *segmentation* (assigning a label to each pixel in an image) [296]. The key feature of a CNN is that the neurons use *convolution* operators (which generalize the weights of standard neural networks), where the output of a neuron is expressed in the following way:

$$\phi(\boldsymbol{x}) = \sigma\left(\sum_{i=1}^{n} C_{\boldsymbol{h}_i}(x_i) + b\right). \tag{1.5}$$

The convolution $C_{\boldsymbol{h}_i}$ of the neuron from input i is characterized by a *filter* $\boldsymbol{h}_i$. For 2D imaging applications, the convolution operator of h_i with the function $g_{\boldsymbol{x}}$, defined by $g_{\boldsymbol{x}}(k,l) = x_{k,l}$, is given by:

$$\begin{aligned} C_{\boldsymbol{h}_i}(\boldsymbol{x}) &= (\boldsymbol{h}_i * g_{\boldsymbol{x}}) \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \boldsymbol{h}_i(m,n) x_{k-m,l-n}. \end{aligned} \tag{1.6}$$

In practice, the summations are taken over a range where the chosen filter is mostly nonzero (which is often indicated by the size of the filter). For instance, a filter can consist of a 3×3 matrix *kernel* containing weights that indicate the contribution of pixel $x_{k,l}$ and its neighbouring pixels to the corresponding pixel $C_{\boldsymbol{h}_i}(\boldsymbol{x})_{k,l}$ in the resulting convolved image (see Fig. 1.16 for an example). The expression in Equation 1.5 is an image of one *channel*. When images from multiple channels are combined, these are referred to as *feature maps* (for instance, an RGB image is a feature map with three channels). Each *convolutional layer* creates a new feature map from input feature maps. The input maps could be those of the last previous layer or of all the previous layers combined. By repeatedly passing on an input image through multiple convolutional layers, complicated features can be extracted from images. Deep convolutional neural networks apply many of such

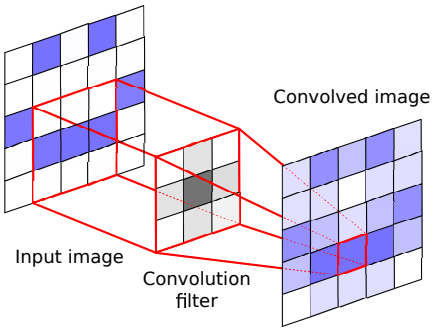


Figure 1.16: Visualization of a convolution. For every target pixel, the input pixel values are multiplied by the corresponding weights in the convolution filter, yielding the convolved image. The colors indicate the values of the pixels. In this case, the resulting image is a blurry version of the input image.

convolutional layers. Along with the biases in the neurons, the weights in the convolution operators are parameters that need to be optimized during the learning process, giving rise to a significantly higher number of trainable parameters than in conventional neural networks with similar architectures.

Nevertheless, despite seemingly overparametrization and the possibility of overfitting of CNNs, these have proven to yield outstanding results for many imaging problems. However, the reason for the success of such deep architectures is not understood well [38]. On top of this, many different strategies can be applied to such networks, giving rise to many classes of CNNs. For instance, the setup of the kernels can be tweaked, such as varying the *stride* (the step size in the summations of Equation 1.6), the *dilation* (the spacing between the kernel points), and *padding* (adding values at the edge of the image), which can all be considered as *hyperparameters* (parameters set beforehand that control the learning process). Additionally, there is a high degree of freedom in the architecture, such as the dependence of convolutional layers on the previous layers. Other examples are the addition of other operations between the convolutional layers, such as up- or downsampling operators or pooling layers (which also reduce the spatial resolution of the feature maps). Chapters 2 and 3 and will go into more detail about relevant CNN architectures and notation, respectively.

The many possibilities in the design of deep CNN architectures results in a large number of CNN classes without a uniform understanding of how each class works. However, this has not prevented its widespread use for imaging problems or limited its suitability for detection and segmentation problems in X-ray imaging and computed tomography. The main assets of CNNs for these fields are the recognition of complicated patterns and the ability to deal well with noisy images, which is for instance useful with (hyper)spectral images that are noisier than normal radiographs. Similar to the feasibility of iterative reconstruction techniques

for tomography, learning with convolutional neural networks has become more prevalent due to improved hardware, particularly the GPUs, allowing to train networks with up to millions of parameters. Toolboxes for advanced deep learning, such as PyTorch [222, 223] and TensorFlow [4], provide even easier access to the application of these methods to imaging problems than before.

Deep learning is applied to many problems in 2D X-ray imaging. An important field of application is X-ray security imaging, such as cargo and baggage scanning at airports and other public areas, to detect and classify potentially harmful objects from radiographs [8]. In medical imaging, deep learning has many applications in radiography [195], including the analysis of chest X-ray images [53], as well as bone fracture classification [277] and bone age estimation [155]. For industry, deep learning methods with X-ray imaging have been developed for tasks such as defect segmentation of castings from radiographs [72, 90], weld inspection [90, 185, 192], and food processing [327]. However, deep learning combined with radiography is underused in those application areas, particularly in food processing [192], although it has been getting more attention in recent years [81]. The usage of deep learning is more common in X-ray CT, where it can be applied for 3D reconstruction and segmentation tasks with limited data [120, 187, 226, 326], such as limited angle and limited view measurements [160, 300], noisy data acquisitions [124, 157, 160, 304], as well as reduction of cone-beam artefacts [197] and reduction beam hardening [319, 329] and metal artefacts [32, 157]. On top of that, (parts of) the tomographic inversion operator can also be learned with deep learning [269, 285, 300]. Such approaches can also be combined with spectral imaging for improved image artefact reduction [49]. Another important field of application of deep learning is image compression [177], with network structures that encode data to a lower dimensional space and decode it to either the original data (autoencoder) or data corresponding to a specific task [198, 229]. In hyperspectral imaging, image compression by means of deep learning is carried out to reduce the size of hyperspectral data cubes of satellites and planes that are transmitted to remote locations for further analysis [73]. For the new hyperspectral X-ray imaging methods, such approaches may turn out useful when applied to high-throughput industrial tasks such as food inspection.

Despite the high success of deep learning in recent years, it also comes with many challenges. Apart from the incomplete mathematical understanding [38] of deep learning and their high computational costs [13], another practical problem is the reliance on data to successfully employ deep learning. Deep learning is very data-demanding [183], with the performance on vision tasks increasing logarithmically with the data volume [271]. On top of this, in supervised learning, the data often need to be annotated by experts, which is a tedious and time-consuming process. For food processing, there is a lack of large datasets [8] to employ deep learning for real-time applications. Moreover, very few open datasets are available for X-ray testing [192], most likely due to the limited number of experts for data annotation. Because of this, even though there is much potential, multi-channel acquisitions still need to be fully used in deep learning based X-ray imaging applications [8].

1.7 Research questions and dissertation outline

This dissertation presents a number of tools and techniques as building blocks that can be integrated into automated X-ray inspection systems for industry. As previously pointed out in this chapter, recent advances in X-ray imaging and machine learning enable the use of spectral X-ray imaging and deep learning. This allows for the construction of learning-driven automated spectral X-ray inspection systems. The main driving question throughout this dissertation is therefore the following:

Main question: *How can we use the possibilities of spectral imaging and deep learning in industrial imaging and industrial tomography?*

To provide an answer to this question, we investigate various aspects of spectral imaging, deep learning, or a combination of these. Each of the four forthcoming chapters will address a separate research question.

Chapter 2: *How can tomography assist industrial 2D radiography, in particular foreign object detection?*

In Section 1.1, we have seen that detection of unwanted ('foreign') objects within products is a standard procedure in many branches of industry for maintaining product quality. In Section 1.2, we have seen that X-ray imaging is a fast, non-invasive and widely applicable method for foreign object detection. As noted in Section 1.6, deep learning has recently emerged as a powerful approach for recognizing patterns in radiographs, enabling automated X-ray based foreign object detection at high rates. However, these methods require a large number of training examples and manual annotation of these examples is a subjective and laborious task.
In **Chapter 2**, we will propose a new workflow that solves this manual annotation problem. The workflow combines CT scanning with segmentation to efficiently and objectively produce annotated training data. With this data, a deep convolutional neural network can be trained to learn the foreign object detection task at hand, after which it can be applied to similar tasks. The method relies on high-quality CT scans and subsequent processing. As opposed to medical CT, in industry, there are generally no dosage constraints on the CT scan and therefore higher-quality reconstructions can be achieved by utilizing long scanning times and high penetration power by powerful X-ray sources. The workflow is demonstrated on a real X-ray CT dataset, with results suggesting that the method can be applied to industrial food inspection.

Chapter 3: *How can we use deep learning for task-driven (hyper)spectral data compression?*

In Section 1.4, we have seen that the arrival of (hyper)spectral X-ray imaging may provide additional insights into objects exposed to foreign object detection tasks. However, an important challenge in hyperspectral imaging tasks is to cope with the large number of spectral bins in the generated data. Common spectral data reduction methods do not consider prior knowledge about the task. Consequently, sparsely occurring features that may be essential for the imaging task may not be preserved in the data reduction step. As noted in Section 1.6, convolutional neural network approaches are capable of learning the specific features relevant to the particular imaging task, but applying them directly to the spectral input data is constrained by the computational efficiency.
In **Chapter 3**, we therefore introduce a novel supervised deep learning approach for image analysis. The method combines data reduction and image analysis in an end-to-end network architecture. The proposed data reduction network architecture, which we name DRCNN, consists of a data reduction block paired with a CNN. The network component that performs the reduction is trained jointly with the CNN such that image features most relevant to the task are preserved in the reduction step. The approach is demonstrated on a number of artificial datasets, and suggests the possibility for higher compression and accuracy in various applications, including hyperspectral X-ray imaging.

Chapter 4: *Can we use multi-channel imaging to improve reconstructions in discrete tomography?*

As noted in Section 1.3, discrete tomography is concerned with objects that consist of a small number of materials, which makes it possible to compute accurate reconstructions from severely undersampled projection data. For cases where the allowed intensity values in the reconstruction are known a priori, the discrete algebraic reconstruction technique (DART) has been shown to yield accurate reconstructions from few projections. However, a key limitation is that the benefit of DART diminishes as the number of different materials increases. Since new tomographic imaging techniques can simultaneously record tomographic data at multiple channels and provide multi-channel data, a new reconstruction algorithm can exploit this additional information.
In **Chapter 4**, we present Multi-Channel DART (MC-DART). This method is a generalization of DART to multi-channel data and combines the information for each separate channel-reconstruction in a multi-channel segmentation step. By iteratively applying these steps, a final accurate segmented reconstruction can be produced from multi-channel data. Through a number of simulation experiments, it is shown that MC-DART is capable of producing more accurate reconstructions compared to (single-channel) DART.

Chapter 5: *How can we use prior information on spectral material signatures in tomographic reconstruction and material decomposition algorithms?*

The advances noted in Section 1.5 in multi-spectral detectors are causing a paradigm shift in X-ray CT. Spectral information acquired from these detectors can be used to extract volumetric material composition maps of the object of interest. The image reconstruction step is relatively straightforward if the materials and their spectral responses are known a priori. If these are not known, however, the maps as well as the responses need to be estimated jointly. The most conventional workflows in spectral CT involve performing volume reconstruction followed by material decomposition, or vice versa. However, these methods inherently suffer from the ill-posedness of the joint reconstruction problem.
In **Chapter 5**, we present a new one-step method for tomographic reconstruction in spectral CT. In this method, which we name ADJUST, prior information about spectral attenuation curves of materials that may appear in the object of interest is taken into account by formulating the problem in a specific way. In essence, the to-be reconstructed volume is written as a multiplication of a material map matrix, a material-indicator matrix, and a spectral dictionary matrix. In the proposed optimization method, the first two matrices are jointly iteratively estimated.

In **Chapter 6**, we will return to the main research question and discuss how the methodologies developed in each chapter can support industrial X-ray imaging and tomography. In addition, the relevance of each method is discussed with respect to the fields of spectral and hyperspectral tomography, discrete and multi-channel tomography, hyperspectral imaging, machine learning and dimensionality reduction, along with all their application areas. Furthermore, we outline the contributions, implications and future research directions.

2

Tomographic workflows for deep learning training data generation

In food inspection, identifying foreign objects is essential. X-ray imaging and machine learning allow for doing this in automated manner. Nevertheless, for reliable performance a substantial body of training data is needed, which is difficult to come by.

In this chapter, we present a Computed Tomography (CT) based method for producing training data for supervised learning of foreign object detection, with minimal labor requirements. In this approach, a few representative objects are CT scanned and reconstructed in 3D. The radiographs that are acquired as part of the CT-scan data serve as input for the machine learning method. High-quality ground truth locations of the foreign objects are obtained through accurate 3D reconstructions and segmentations. Using these segmented volumes, corresponding 2D segmentations are obtained by creating virtual projections. We outline the benefits of objectively and reproducibly generating training data in this way. In addition, we show how the detection accuracy depends on the number of objects used for the CT reconstructions. We show that in this workflow generally only a relatively small number of representative objects (i.e. fewer than 10) are needed to achieve adequate detection performance in an industrial setting.

This chapter is based on:

M. T. Zeegers, T. van Leeuwen, D. M. Pelt, S. B. Coban, R. van Liere, and K. J. Batenburg. "A tomographic workflow to enable deep learning for X-ray based foreign object detection". *Expert Systems with Applications* 206 (2022), p. 117768.

2.1 Introduction

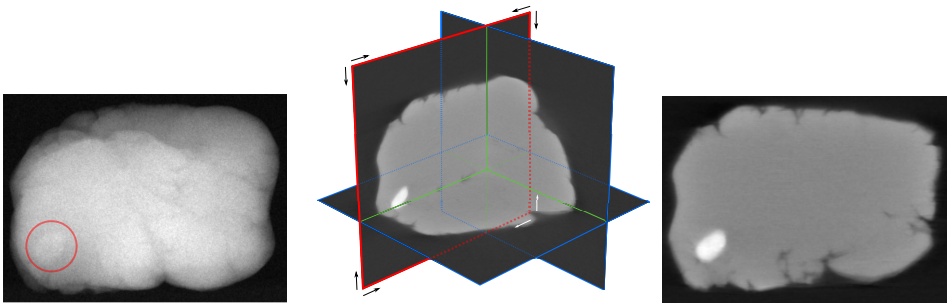
Foreign object detection in an industrial high-throughput setting is essential for guaranteeing quality and safety of objects processed in factory lines. Foreign objects may, for example, appear in products such as meat, fish or vegetables as small pieces of glass, bones, plastic, wood or stone that could harm consumers [16, 294, 327]. Conventional nondestructive methods for detecting foreign objects include ultrasound imaging, X-ray imaging, magnetic resonance imaging, fluorescence imaging, (hyperspectral) spectroscopic imaging and thermal imaging [116, 162, 200, 207, 210, 303]. X-ray imaging provides the unique opportunity to visualize the interior structure of an object in a fast, low-cost, and non-invasive manner. This enables *X-ray based foreign object detection*, in which the goal is to detect unwanted smaller objects inside base objects based on their distinct attenuation or attenuation patterns, as observed in generated *radiographs* (i.e. standard 2D X-ray images). The possibility to reveal hidden foreign objects in radiographs has lead to its extensive use in various industrial applications [80, 111, 153, 186, 191, 207, 323], for which low-cost, adaptive and efficient image processing methods are essential [186, 303]. One way to achieve better discrimination of foreign objects in radiographs is to use multispectral X-ray imaging detectors, simultaneously capturing radiographs at two or more energy levels [199, 274]. As the attenuation properties of each material have their own characteristic dependence on the X-ray energy, these multispectral images can be analyzed to extract material composition information.

However, superposition of materials gives rise to similar levels of intensities for different objects in 2D radiographs. This problem limits the application of commonly used segmentation methods, such as threshold-based, clustering-based, and boundary-based or edge-based segmentation [256, 260], to extract different components of the object. Additionally, high-throughput acquisition may lead to high noise levels in radiographs, and this increases the difficulty of successful foreign object detection even further [186, 303]. Commonly used segmentation methods can be unsuitable in case of poor image qualities caused by conditions such as noise, low contrast and homogeneity in regions close to foreign objects [260]. Most conventional unsupervised methods can therefore not achieve high accuracies [260] without extensive manual parameter tuning to use a method for a specific problem [239, 245].

Machine learning is a powerful tool for recognizing patterns in images [322] and can potentially detect foreign objects in radiographs [327]. Recent machine learning methods address a wide variety of segmentation problems [98, 260], and provide a remarkable improvement over more classical segmentation methods in many practical applications [108]. A key obstacle in the application of machine learning is the need for large datasets [57, 69, 297], which is particularly prominent in machine learning for foreign object detection as each new combination of sample, foreign object, and imaging settings requires additional data. On top of that, supervised

learning uses labelled datasets for training. However, manual annotation (as in e.g. [245, 260]) requires tremendous efforts [8], is time consuming and tedious [276], is subjective and can be prone to errors.

The key contribution of this chapter is the proposal of a workflow based on 3D Computed Tomography (CT) for efficiently creating large training datasets, overcoming the aforementioned obstacle. CT scans of a relatively small number of objects are carried out with low exposure time – as in a high-throughput setting – yielding a large number of radiographs that are used as input for the supervised machine learning method. The same set of radiographs is also used offline for generating multiple high-quality tomographic 3D reconstructions, from which foreign objects can easily be segmented in 3D and projected back onto a virtual 2D detector to give the corresponding ground truth locations of the foreign objects in the radiographs. Without the effort of extensive manual labelling, this results in a large dataset with which deep learning can be carried out to detect foreign objects from fast-acquisition radiographs at a high rate. The example in Figure 2.1 illustrates the difference in ease of segmentation for a CT reconstructed 3D volume versus a 2D radiograph. Whereas segmenting the foreign object in a radiograph is a challenging task, simple global thresholding can be applied to the CT volume to separate the foreign object from the base object. Additionally, more sophisticated and accurate segmentation and denoising rules can be imposed on 3D volumes [98, 173, 217] than on 2D radiographs.



(**a**) 2D radiograph (**b**) 2D slices in the reconstructed 3D volume space (**c**) Red-bordered 2D slice of the 3D reconstruction

Figure 2.1: Different views of an imaged product (Play-Doh) with a foreign object inserted (a piece of gravel). A 2D radiograph with the location of the foreign object (red circle) is shown (**a**), as well as multiple slices through the 3D volume of the reconstructed object (**b**), of which the slice with the red border is highlighted (**c**). The images show the difference in contrast: the foreign object is much easier to distinguish based on intensity values in the reconstructed 3D volume (**b** and **c**) than in the 2D radiograph (**a**).

The structure of the chapter is as follows. Section 2.2 provides the background of applying machine learning for foreign object detection, and Section 2.3 explains the proposed method of data generation to apply machine learning. In Section 2.4, the workflow is demonstrated in a laboratory experiment, and shows how the number of imaged objects affects the detection accuracies. Additionally, the robustness of the workflow is analyzed. Section 2.5 discusses various aspects of the results and the flexibility and modularity of the workflow. Section 2.6 presents the conclusions from this work.

2.2 Preliminaries

In this section, we introduce the X-ray foreign object detection problem and the machine learning concepts used in this chapter.

2.2.1 Foreign object detection with X-ray imaging

We consider the problem of foreign object detection in an industrial high-throughput conveyor belt setting. The problem and the usage of X-ray imaging to solve this are schematically shown in Figure 2.2. In foreign object detection, the aim is to correctly determine for each object whether a foreign object is contained in it or not, for instance a piece of bone within a meat sample.

For this problem, we focus on finding an accurate *segmentation* for each radiograph. A segmentation partitions an image into sets of pixels with the same label. In our case, the formed segmented image is binary and indicates on which detector pixels a foreign object is projected. The segmentation depends on the type of objects that are considered to be foreign (by for instance a manufacturer). Any further classification (based on the minimum size of a foreign object for example) can be carried out after the segmented image is produced.

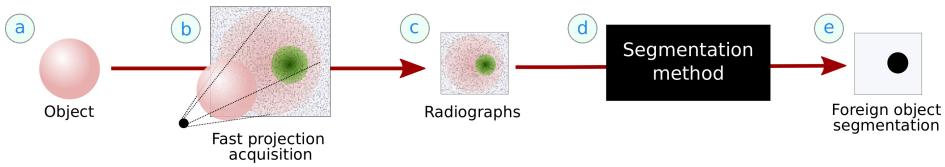


Figure 2.2: A schematic overview of the foreign object detection problem and the segmentation-based approach to solving it. Each object (**a**) is assumed to have a correct segmentation (**e**). By using X-ray imaging (**b**), a radiograph of the object (**c**) can be acquired. Using a segmentation method (**d**), a segmented image (**e**) can be produced. The main challenge is to find a suitable segmentation method such that this approach to foreign object detection produces the correct results.

Throughout this chapter, we use the term *radiograph* for radiographs corrected using flatfield radiographs (without an object) and darkfield images (without the X-ray beam) that serve as input to the segmentation method. The quality of a radiograph depends on a number of properties of the scan, including exposure time, tube intensity, photon energy windows and the geometric setup [242]. In a high-throughput setting, the steps in Figure 2.2 should be fast to carry out, typically resulting in high noise levels and a challenging segmentation task.

2.2.2 Supervised learning

Machine learning is a widely used approach for difficult imaging tasks, as it can extract complicated patterns from complex images. In the foreign object detection problem, supervised machine learning can be used to learn the segmentation task such that it generalizes well for all possible fast-acquisition radiographs of similar objects with similar acquisition settings. To do so, a set of examples $\{(x_i, y_i)\}_{i=1}^N$ is used, where $\{x_i\}_{i=1}^N$ are acquired radiographs and $\{y_i\}_{i=1}^N$ are their corresponding foreign objects segmentations. The aim is to find the unknown segmentation function F that maps each radiograph x_i to its segmentation y_i. To find an approximate solution that generalizes well, the set of images is partitioned into a training set, a validation set and a test set. The training set is used to learn the function F_{train} that minimizes the loss L on the training set, which is the sum of errors between the segmented images $F_{\text{train}}(x_i)$ produced by the segmentation function and the true segmented images y_i. To find a suitable segmentation function, a (convolutional) neural network is often used as a model and parametrized using weights and biases that are optimized during the training process. While carrying out the training with a chosen loss function and optimization algorithm, the performance of the model is evaluated on the validation set. Several stopping criteria can be used for this, for example stopping the training when the error on the validation set increases, or training for a fixed time (and recording the network that gives the best results on the validation set). To avoid any bias towards the training and validation data, the accuracy of the trained model is finally assessed using the test set.

Since the introduction of Fully Convolutional Networks [171], in which successive contracting convolutional layers are utilized for pixel-wise semantic segmentation, many convolutional neural network (CNN) architectures have been proposed that can be used for the object segmentation task. U-Net changes the FCN architecture by - along with downsampling operators and skip connections - introducing upsampling operators instead of pooling operators, giving it an U-shaped appearance [240]. Similarly, Deconvnet [212] also introduces an auto-encoder structure with deconvolution and unpooling operations (without skip connections). The success of these methods on medical image segmentation and object detection spawned other commonly used CNN architectures for segmentation such as SegNet [22], RefineNet [167], PSPNet [321], and Mask R-CNN [115] for instance segmentation. Although

some of the listed architectures need relatively few training examples for successful segmentation, the annotation of these examples still requires considerable efforts.

2.3 Proposed method for training data acquisition

When attempting to perform machine learning for X-ray based foreign object detection, the major obstacle is to acquire (manually) annotated training data. In this section, we explain the methodology of our proposed CT-based workflow for efficiently creating this annotated training data.

Our proposed workflow for using CT to obtain annotated training images is schematically displayed in Figure 2.3. First, we select a set of representative objects as training objects (Fig. 2.3a). For each object, a set of fast-acquisition radiographs is collected from a set of predefined angles (Fig. 2.3b). These fast-acquisition radiographs will form the input set of the intended training dataset (Fig. 2.3c). The total number of examples in the resulting dataset is the number of training objects multiplied by the number of selected angles.

The same set of radiographs is used to carry out a tomographic reconstruction of the object and acquire high-quality CT volumetric data (Fig. 2.3d and e). The next step is to segment the reconstructed volume such that a possible foreign object is separated from the base object (Fig. 2.3f). This segmentation step can be automated and many methods are available to implement this [158]. Here, we consider volumetric segmentation methods that consist of a global thresholding step.

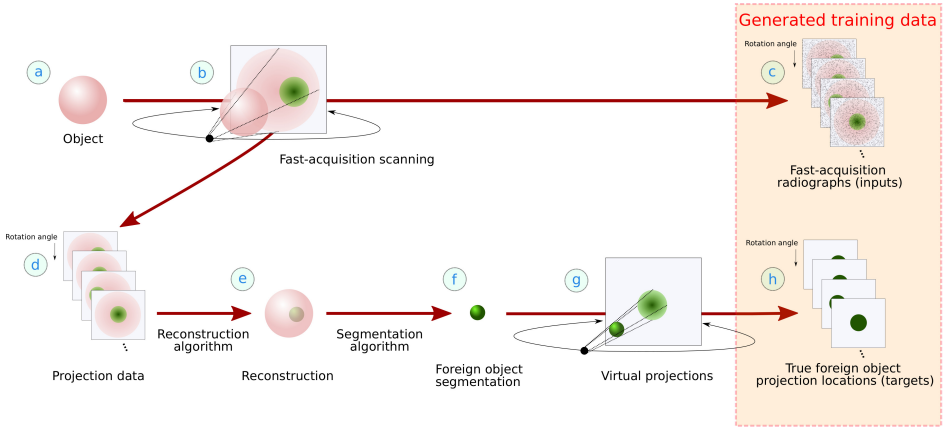


Figure 2.3: The complete workflow of data acquisition (**a**,**b**) and the generation of training data (**c**,**h**) for deep learning driven foreign object detection, through 3D reconstruction from the CT scan (**d**, **e**), segmentation (**f**), and virtual projections (**g**). The reconstruction reveals the hidden foreign objects inside the main object. Note that the projection data (**d**) are usually just the set of fast-acquisition radiographs (**c**).

Binary segmentation by global thresholding is defined by the following function $S : \mathbb{R} \to \{0, 1\}$ that acts on every voxel z_{ijk} in reconstruction volume z:

$$S(z_{ijk}) = \begin{cases} 1 & z_{ijk} \geq \theta, \\ 0 & z_{ijk} < \theta. \end{cases}$$

Here, θ is the segmentation threshold. The more angles and other high-quality settings are used to obtain projection data, the easier it is to accurately segment the foreign object. Easier segmentation can also be accomplished by carrying out a separate high-quality scan of the same object and making a reconstruction with these high-quality radiographs. Additionally, for segmentation, prior information about the objects can be used, such as bounding boxes on the foreign object location [144]. Also, 3D denoising [70, 119] can be used to remove non-foreign object pixels captured by the thresholding operation.

From the constructed foreign object segmentation, virtual ground truth projections are generated by simulating projections of the foreign objects onto a virtual detector (Fig. 2.3g). This results in the set of ground truth images, which will serve as target images in the machine learning procedure (Fig. 2.3h). These virtual projections need to be taken under the same angles as in the fast-acquisition scan (Fig. 2.3b). When this procedure is repeated for all objects, this results in a large dataset with annotated training examples with which supervised machine learning can be carried out (Fig. 2.3c and f). The trained model can then be applied to similar new objects scanned in the same fast-acquisition setting, without the need for acquisition of high-quality radiographs or CT scans.

2.4 Experiments and results

In this section, we demonstrate the proposed workflow using the in-house FleX-ray CT system at CWI [62] (Fig. 2.4), and investigate the relation between machine learning performance and the number of training objects used.

2.4.1 Base objects and foreign objects

As test objects, we use base objects that are created from a fixed amount of modelling clay (Play-Doh, Hasbro, RI, USA). Play-Doh is primarily made of a mixture of water, salt and flour and we therefore consider it to be a representative example of products in the food industry, where foreign objects may be pieces of stone, plastic, or metal. A basic shape is deformed and remolded for every object instance (Fig. 2.5a) in such a way that they are similar from object to object, but still exhibit some natural variation. For the foreign objects we choose to use gravel (Fig. 2.5b), with the stones having an average diameter of ca. 7mm (ranging from 3mm to 11mm). These stones have slight variations in shape and material. We

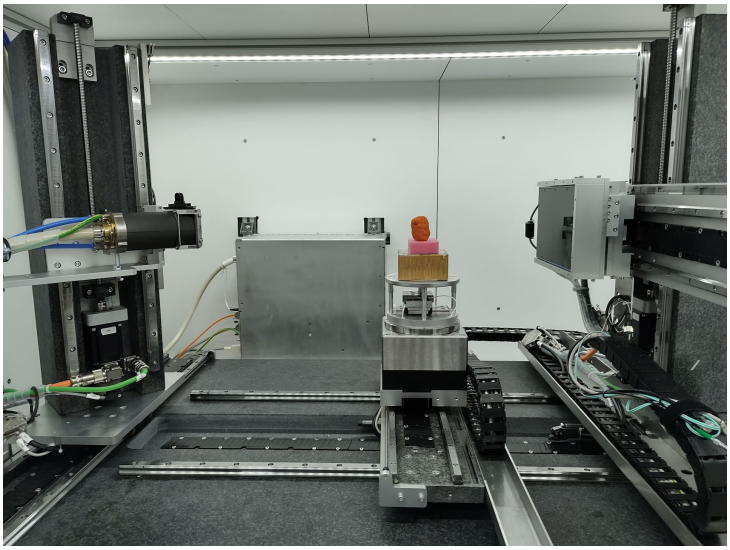


Figure 2.4: The scanning setup in the FleX-ray laboratory with the X-ray source on the left and the detector on the right.

(**a**) Play-Doh (**b**) Stones (**c**) Play-Doh with stone inserted

Figure 2.5: An example of a base object (**a**) and examples of foreign objects (**b**) used in the laboratory experiments, as well as an example of a combined object (**c**).

create 3 objects with three inserted stones, 35 with two stones, 62 with one stone (Fig. 2.5c) and 11 without a stone.

2.4.2 CT scanning and data preparation

A fast CT scan is made for each of the objects, which yields both a series of radiographs (i.e. the X-ray projections) and a reconstructed 3D volume of the object. The objects are scanned in the FleX-ray laboratory [62] (Fig. 2.4). The FleX-ray CT scanner has a cone-beam microfocus X-ray point source with a focal spot size of 17 μm, and a Dexela1512NDT detector. The source, object and detector positions can be configured flexibly, and are arranged such that the distance between the source and detector is 69.80 cm, and the distance between the source and the object 44.14 cm. For the radiographs a voltage of 90 kV with a power of 20 W is used, while the exposure time is kept low at 20 ms, with the intention to emulate the imaging conditions of in-line industrial systems and produce sufficiently noisy radiographs. To achieve high-quality reconstructions, 1800 projections of each object are obtained at equidistant angles over a full 360° rotation. All projection angles are precisely recorded during the scan for the later stages of the workflow. Before and after each scan, 10 darkfield images and 10 flatfield projections are obtained. Each object is positioned in a random manner, and the cylinders may therefore be standing upright or be laying down on the long edge. Example radiographs are shown in Figure 2.6. Separating the projected foreign objects from the base object in these radiographs is not a trivial task, illustrating the problem of obtaining annotated training data for automated segmentation using machine learning directly from these images.

The Simultaneous Iterative Reconstruction Technique (SIRT) [140, 265] algorithm (100 iterations) as implemented in the ASTRA toolbox [1, 2] is used to compute the reconstructed 3D CT volume of the object. A visualization of the reconstruction from the third object in Figure 2.6 and its foreign object is shown in Figure 2.7. The CT reconstruction allows to slice the object along different axes. As the CT voxel intensity is directly related to the attenuation coefficient of the material in a voxel, the segmentation task for the 3D CT volume is, in this

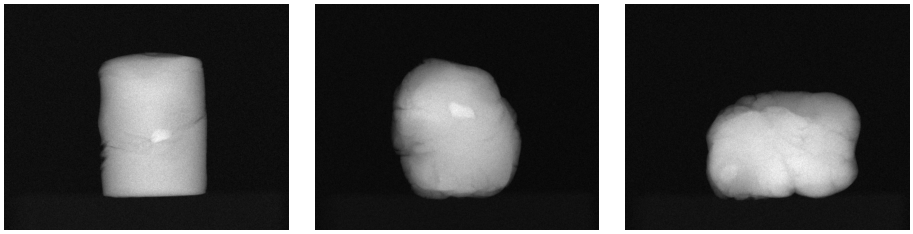
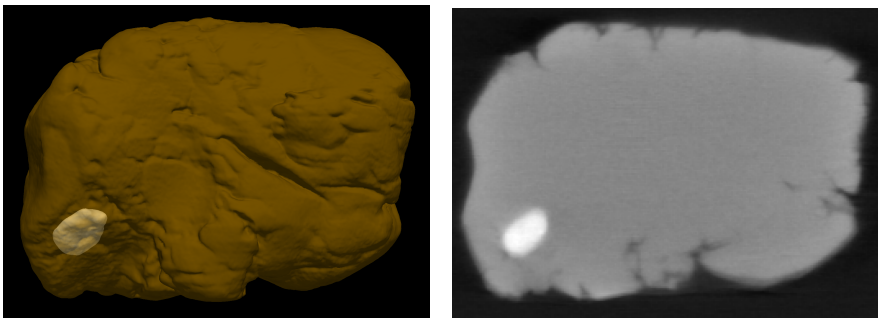


Figure 2.6: Example of radiographs (size 965 × 760 pixels) of three objects that are scanned. In the first and second radiographs the foreign object is clearly visible, but in the third it is more difficult to distinguish it from the base object, even though it is visible on the bottom left.



(**a**) Reconstructed 3D volume (**b**) Slice of the 3D reconstruction

Figure 2.7: Visualization of reconstructions by 3D rendering (**a**) and slicing (**b**) of the third object in Figure 2.6.

case, much more straightforward and can be carried out by global thresholding (see Appendix A.1 for additional details on intensity value distributions). Therefore, a simple global threshold based on Otsu's method [214] is sufficient to segment the foreign objects.

From the 3D segmented objects obtained from the CT scans, 2D segmentations for the individual radiographs are computed. This is done by computing the projections of the segmented parts with the ASTRA toolbox using the same geometric properties and recorded angles as in the radiograph acquisition of the actual CT scan, to ensure geometric consistency in the training examples. Every nonzero pixel on the detector is marked as a projected foreign object location. The result is a dataset containing 1800 radiographs and corresponding segmented images for each object.

2.4.3 Machine learning

We use the Mixed-Scale Dense (MSD) [227] and the common U-Net convolutional neural network architectures [240] to train the task of image segmentation. For our experiments with U-Net, we have slightly changed the architecture, as we observed this improved performance in the experiments compared to the standard version. We downsample twice, with a stride of 2. The initial number of feature maps is set to 128, and the number of feature maps doubles for each downsampling layer. For upsampling, bilinear interpolation is used. A spatial 3×3 convolution operation with zero padding and a ReLU activation function are carried out before and after all downsampling and upsampling operations. The biases and convolution weights are initialized by sampling from $\mathcal{U}(-\sqrt{k}, \sqrt{k})$, with $k = {}^{1}/_{c_{\text{in}} \cdot a^2}$ being the range, c_{in} the number of input channels and a the kernel size. ADAM optimization on the average of the binary cross entropy loss and the dice loss [135, 270] between the data and the predictions is used for training. The network is implemented with PyTorch [222, 223]. For comparison between architectures, we also use the

MSD network for training. MSD is a compact network architecture that has been demonstrated to be suitable for real-time segmentation of X-ray and CT images using relatively few training examples compared to larger networks [227], including the U-Net architecture. We use a depth of 100 intermediate layers and width of 1 channel per intermediate layer and increase the dilation parameter repeatedly from 1 to 10 dilations in each layer, which are common settings for the MSD network [154, 226, 227]. Xavier initialization is used for the convolution weights. ADAM optimization [146] is used during training on the cross-entropy loss between the ground truth and the segmented images, and the batch size of training examples is set to 10. We use the GPU implementations in Python that are available [225, 227]. For both architectures, the learning rate is set to 0.001 and all networks are trained on a GeForce GTX TITAN X GPU with CUDA version 10.1.243. All hyperparameters of both network architectures are kept the same during all experiments. Data augmentation is applied by rotation and flipping of the input examples. All networks are trained for 9 hours, and the network with parameters resulting in the lowest error on the validation set is used for testing.

With these networks, we carry out an image-to-image training from radiographs (Fig. 2.3c) to their corresponding foreign objects segmentations (Fig. 2.3f). For training, 60 randomly chosen base objects containing a foreign object are used. The remaining 51 objects are used for testing. All images are resized using cubic interpolation to 128×128 to speed up the training process (global thresholding with parameter $\theta = 0.5$ is applied to the resized ground truth images to make these binary again). We test the performance of the trained networks for different numbers of objects included in the training scheme. To compare the workflow with labor-intensive 2D data annotation, we compare the following training strategies:

- **Workflow approach**: For each network, we fix the total number of training examples to 1800. A random but fixed order of the 60 training objects is created and the first i objects among these are used for the training set. The training examples are selected from the set of radiographs and ground truths created by the workflow from these i training objects in equal amounts. Every 10th example is used for validation during training.
- **Classical approach**: For each network, only one randomly chosen training radiograph with the corresponding ground truth (generated using the workflow) is selected for each of the first i included training objects. The resulting set of training examples is separated such that $9/10$ part is used for training (rounded down to the nearest integer) and $1/10$ part is used for validation (rounded up).

2.4.4 Quality measures

To evaluate the accuracy of the trained networks on the test set, in which the target images are generated using the workflow on the test objects. Three different measures on the segmented images and the corresponding target images are computed. The collection of these measures both assess the image segmentation accuracy and the object detection accuracy. An image segmentation accuracy is based on the classification of each pixel in the segmented image, and there are standardized ways to measure this that do not depend on any parameters [105]. An object detection accuracy compares connected components (groups of pixels connected by their edges) in the segmented image with the ground truth images. Although these accuracy measures require additional parameters to define the notion of detection, they are more relevant to the foreign object detection application.

The first measure is an *image-based average class accuracy* (also called *balanced accuracy* [105]) to assess the accuracy of a produced segmentation. The average class accuracy of a segmented image relative to the target image is given by the sum of the true positives divided by the true positives and false negatives (the recall) of each class, averaged over the number of classes. In the binary case this becomes

$$\frac{1}{2}\left(\frac{\mathrm{TP_{FO}}}{\mathrm{TP_{FO}}+\mathrm{FN_{FO}}}+\frac{\mathrm{TP_{BG}}}{\mathrm{TP_{BG}}+\mathrm{FN_{BG}}}\right). \tag{2.1}$$

Here, $\mathrm{TP_{FO}}, \mathrm{FN_{FO}}, \mathrm{TP_{BG}}$ and $\mathrm{FN_{BG}}$ are the true positives and false negatives of the foreign object and the combined base object and background pixel classifications respectively over the entire segmented image relative to the target image. The average class accuracy as given in (2.1) is averaged over all target images.

The second measure is an *object based detection rate*. A *connected component* is a maximal set of nonzero-valued pixels such that each pixel is reachable from another pixel in the set via a sequence of neighboring pixels in the set. Each connected component in the target image with a minimum size of 8 pixels (0.05% of the image size) is considered as an object that should be detected. We define such an object as *detected* if its pixel-wise recall relative to the segmented image is higher than a certain threshold η:

$$\frac{\mathrm{TP}^{\mathrm{tar}}_{\mathrm{obj}}}{\mathrm{TP}^{\mathrm{tar}}_{\mathrm{obj}}+\mathrm{FN}^{\mathrm{tar}}_{\mathrm{obj}}} > \eta. \tag{2.2}$$

Here, $\mathrm{TP}^{\mathrm{tar}}_{\mathrm{obj}}$ and $\mathrm{FN}^{\mathrm{tar}}_{\mathrm{obj}}$ are the true positive and false negative pixels in the target object relative to the segmented image. The threshold indicates the percentage of pixels of a projected foreign object in the target image that should be indicated as foreign object pixels in the segmented image produced by the network to be marked as a detected object. In our experiments, we set $\eta = 0.3$. We define the *detection rate* as the percentage of components in all target images for which condition (2.2) holds.

The third measure is an *object based false positive detection rate*. Each connected component in the segmented image with a minimum size of 8 pixels is considered as a potentially detected object. We define such a potentially detected object as a *false positive* if its pixel-wise recall relative to the target image is lower than a certain threshold δ:

$$\frac{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}} + \mathrm{FN}_{\mathrm{obj}}^{\mathrm{seg}}} < \delta. \tag{2.3}$$

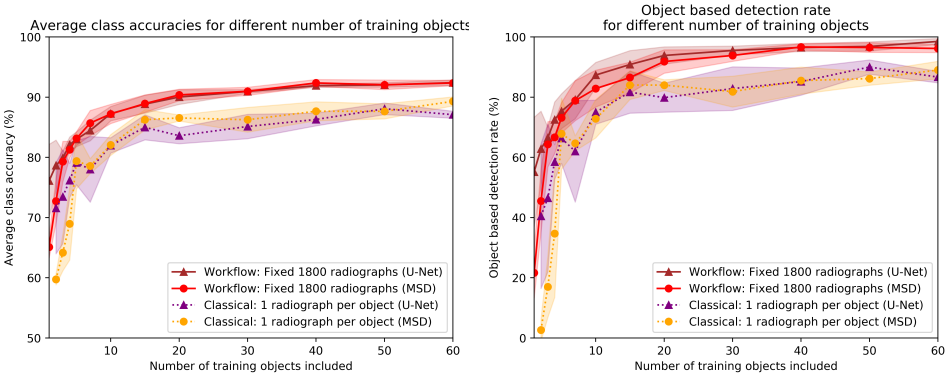
Here, $\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}$ and $\mathrm{FN}_{\mathrm{obj}}^{\mathrm{seg}}$ are the true positive and false negative pixels in the segmented object relative to the target image. The threshold indicates the percentage of pixels of a foreign object in the segmented image produced by the network that are correctly labelled as foreign objects compared to the foreign object in the target image. In our experiments, we set $\delta = 0.3$. We define the *false positive detection rate* as the percentage of potential objects in all segmented images for which condition (2.3) holds.

2.4.5 Results

For the test set, we select a random angle and an orthogonal one for each test object, making the total number of testing radiographs 102. We measure the average class accuracy, the object based detection rate and the object based false positive detection rate of segmentations created by the network on the projections from the test set. The results are given in Figure 2.8.

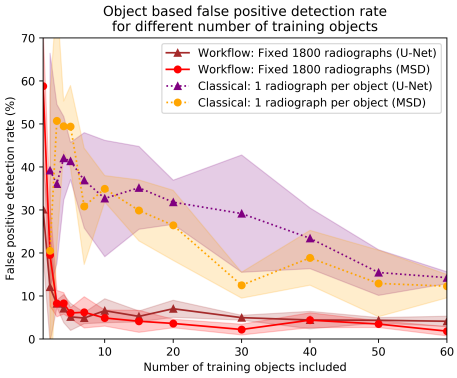
For all measures, the quality of the foreign object segmentations in the radiographs using networks trained with the workflow data is low for a few training objects. This initially improves with the addition of relatively few training objects, but this improvement stagnates beyond 20 objects. However, the detection accuracy still shows slight improvements beyond this point, but almost completely stabilizes from 40 objects onwards. Based on a decided accuracy goal, a certain number of objects need to be scanned and used for training to achieve that accuracy. The false positive rate decreases strongly and maintains a low level value from including 3 objects in the training onwards. Note that the results between the U-Net and MSD architectures agree well with each other.

When we compare the usage of a fixed number of training radiographs among all training objects with the classical approach of using only one radiograph per object, we see that this leads to inferior results in all aspects. The average class accuracies and the object based detection rates are lower for all numbers of included training objects, while the false positive rates are higher. The difference between architectures only shows for the false positive detection rate, which is generally higher with the U-Net architecture.



(**a**) Average class accuracy

(**b**) Object based detection rate



(**c**) Object based false positive detection rate

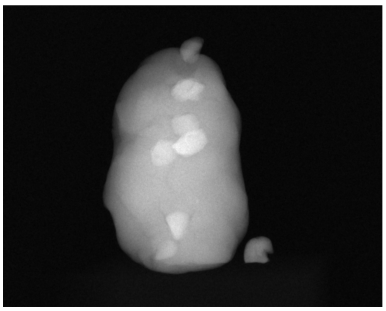
Figure 2.8: The average class accuracy (**a**), the object based detection rate (**b**), and the object based false positive rate (**c**) of segmentations with trained U-Net and MSD networks on laboratory data for different number of training objects. The results are shown for the fixed number of training radiograph approach (workflow) and the one training radiograph per object approach (classical). The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the respective standard deviations.

2.4.6 Laboratory experiments with many foreign objects

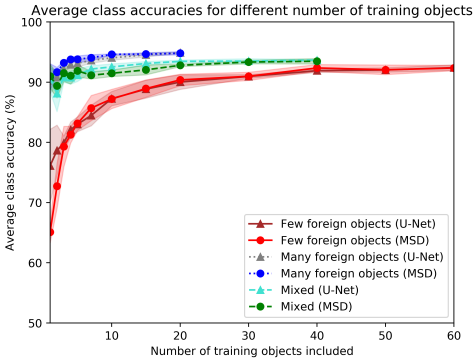
A natural way to reduce the number of objects used for training that need to be scanned for obtaining accurate segmentations may be to include more foreign object in the imaged objects. To test this, we repeat the experiments of the previous section, but we insert 5 to 8 foreign objects instead of 0 to 2. The foreign objects are placed within the base object such that overlapping of foreign objects in the radiographs is minimized. We have scanned an additional set of 20 objects with these characteristics. An example of a radiograph of an object with many foreign objects is shown in Figure 2.9a. We compare the following training strategies in

which the workflow data come from the following sets of training objects:

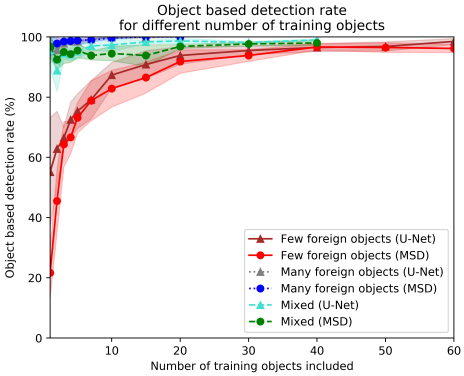
- **Few foreign objects**: Base objects with 0 to 2 foreign objects
- **Many foreign objects**: Base objects with 5 to 8 foreign objects
- **Mixed**: 50% – 50% mix of base objects with 0 to 2 foreign objects and base objects with 5 to 8 foreign objects.



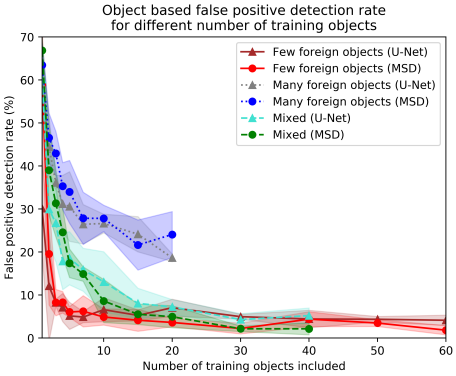
(**a**) Base object with many foreign objects



(**b**) Average class accuracy



(**c**) Object based detection rate



(**d**) Object based false positive detection rate

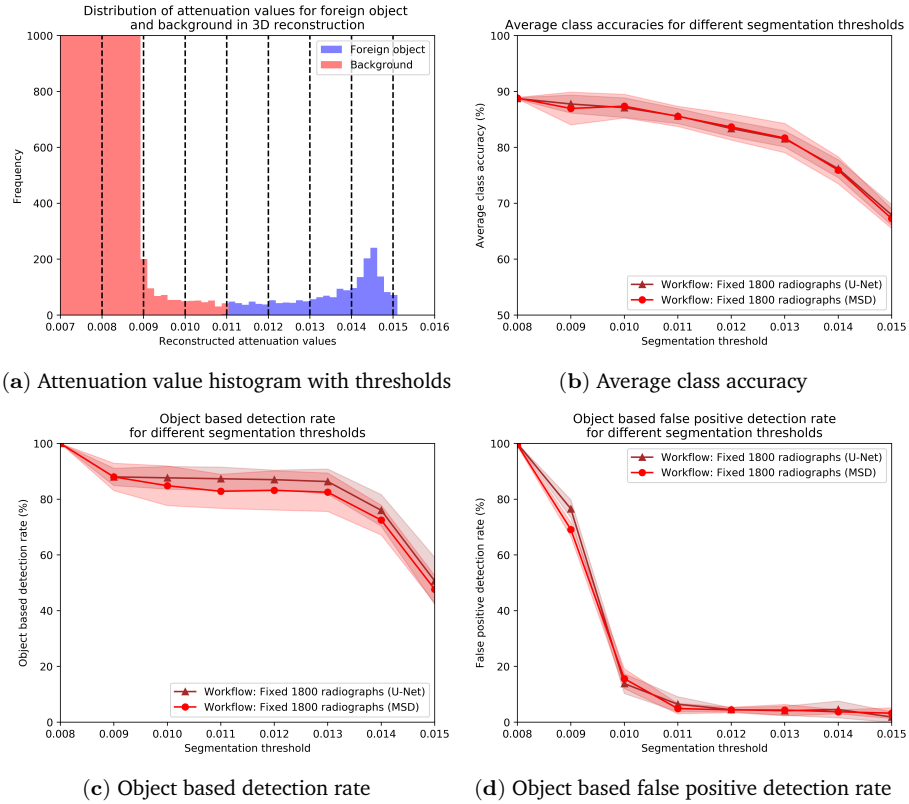
Figure 2.9: Example of a radiograph of a base object with many foreign objects (**a**). The eight foreign objects are vertically placed in the object, although there still may be some overlap. The average class accuracy (**b**), the object based detection rate (**c**), and the object based false positive detection rate (**d**) of segmentations with trained U-Net and MSD networks for different number of training objects and training strategies are shown. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

All networks are evaluated on the testing set from the previous section (with test objects containing few foreign objects). The average class accuracies, detection accuracies and the false positive rates of the trained neural networks with these schemes on the test set are shown in Figure 2.9. From the graphs in Figures 2.9b and 2.9c we see that the average class accuracies and detection accuracies are higher for the many foreign object training scheme, but Figure 2.9d indicates that false positive rate is also roughly 5 times higher. The mixed approach appears to find middle ground between the two other approaches for all measures. We see that from 20 objects onwards the mixed approach is as good as the approach with few foreign objects in terms of the false positive rate, while being superior in terms of average class accuracy and detection accuracy for up to 40 training objects. This shows that including many foreign objects in the training set for detecting few to no foreign objects in the test set has limited additional value, but mixing these with examples with objects containing a few foreign objects may result in higher detection quality while maintaining a similar false positive detection rate.

2.4.7 Robustness of the workflow

In the previous experiments, the trained networks are tested on a set of projections that are generated using the same 3D segmentation threshold parameter in the workflow as in the generation of the data for the training and validation sets. To assess the robustness of the workflow to different segmentation parameters, we generate the training datasets with different values of the segmentation parameter θ (see Figure 2.10a). For each of these values, networks are trained and assessed on the test set from the previous sections. The number of training objects that are included in the workflow is fixed to 10 (which has led to equivalent results in the previous experiments as with 60 objects in the classical approach).

In Figure 2.10, the average class accuracies, detection accuracies and the false positive rates of the trained neural networks are shown for the different thresholds. The results for U-Net and MSD are very similar. As the threshold value increases, the average class accuracy decreases, with significantly lower values for $\theta = 0.014$ and $\theta = 0.015$. The same holds for the detection rate, but it reaches a plateau between $\theta = 0.009$ and $\theta = 0.013$ where this accuracy measure gives similar values. For low values of the threshold parameter, the false positive values are high, and from $\theta = 0.011$ and higher these are low and similar to each other. Taken together, threshold parameters between $\theta = 0.011$ and $\theta = 0.013$ lead to very similar results. We conclude that for the class of objects considered in these experiments, the workflow is robust against moderate variation of the segmentation parameter and that suboptimal segmentation methods can also be used in the workflow.



(**a**) Attenuation value histogram with thresholds

(**b**) Average class accuracy

(**c**) Object based detection rate

(**d**) Object based false positive detection rate

Figure 2.10: The eight considered thresholds for the generation of the training datasets in the workflow, drawn in the histogram of attenuation values of the third object in Figure 2.6, and the average class accuracy (**b**), the object based detection rate (**c**), and the object based false positive detection rate (**d**) of segmentations with trained U-Net and MSD networks on data resulting from these segmentation thresholds. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

2.4.8 Simulation experiments

In this section, we will demonstrate the workflow in a controlled simulated setting. In this way, we can verify the results with larger training and test sets when more objects are available. Furthermore, the test set previously consisted of data generated with the workflow, but in a simulated setting 'absolute' ground truth can be created for the test set by directly projecting the simulated foreign objects (see Figure 2.11). We verify that the proposed workflow (with CT scanning, reconstruction and segmentation) results in segmented foreign objects of which the projections are similar to absolute ground truth projections, which further supports the confidence we can have in the experimental test results.

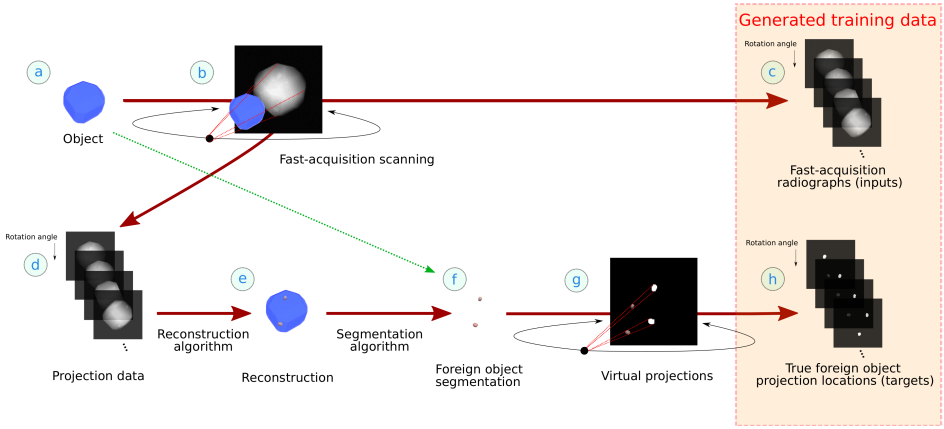


Figure 2.11: The complete workflow of data acquisition (**a**,**b**) and the generation of training data (**c**,**h**) for the simulation experiment by 3D reconstruction from the CT scan (**d**, **e**), segmentation (**f**), and virtual projections (**g**). The 3D reconstruction reveals the hidden foreign objects inside the object. The dotted green arrow (**a** to **f**) indicates that because of the simulated nature of the objects, the reconstruction and segmentation steps are skipped for the generation of ground truth for objects in the test set.

We have generated a set of 500 objects, each in an object space of 128^3 voxels. Each object is a cube of size 64^3 voxels, which is placed in the center of the volume. To create sufficient variety among the objects, the cube is cut off by eight planes. For each corner of the cube, a plane is created by selecting points on each of the three outgoing edges of the corner, randomly between the corner point and the midpoint of that edge. The pixels are cut off whose location is on side of the plane opposite to the center of the cube. See Figure 2.12 for a visualization. Additionally, we rotate the resulting object with random angles around all axes. After that we include a foreign object as an ellipsoid with a radius randomly chosen between 3 and 7 voxels at a random location within or on the edge of the base object. These ellipsoids have a random orientation as well. As a result, the foreign objects vary in shape, size, orientation and location. With 50% probability, we include two of these foreign objects instead of one in the base object.

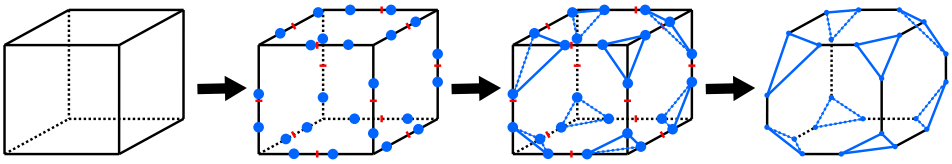


Figure 2.12: The process of cutting off corners with planes from cubes for creating the simulated base objects. The red stripes indicate edge midpoints and the blue dots are the randomly chosen points between those midpoints and the corners.

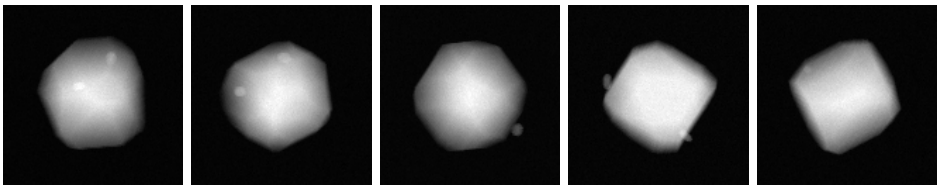


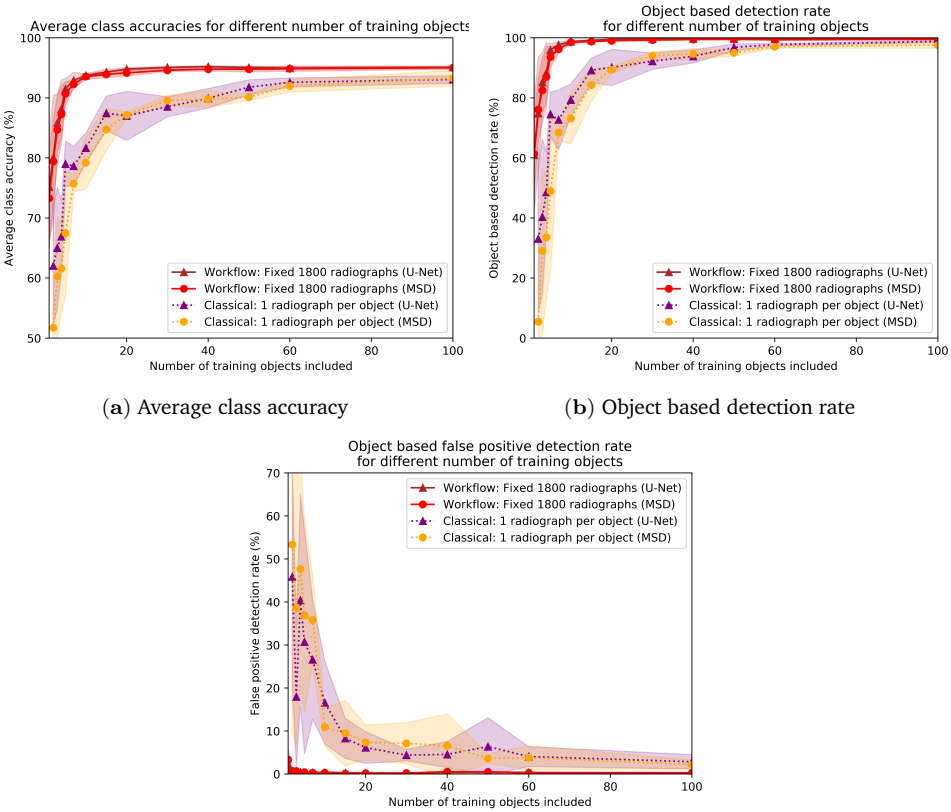
Figure 2.13: Example radiographs of five simulated objects. Foreign objects are located at various positions in or on the border of the base object in the radiographs, and there can be one or two of these present.

Based on the spectral properties of the assigned materials, we create simulated radiographs (Fig. 2.11b). Details of the computation can be found in Appendix B.2. First, we make projections of each material separately by computing cone beam forward projections using the ASTRA toolbox [1, 2]. From this, the simulated radiographs are computed by taking the spectral properties of each material into account (taken from the National Institute for Standards and Technology (NIST) [130]). We model the foreign objects as bone and the base object as tissue for each object. We take the spectral material characteristics between 15 keV and 90 keV into account, and use an exposure time of 0.002 seconds for each radiograph, for which the Poisson noise that is applied is relatively high. These settings are chosen such that there is sufficient contrast in the radiographs, but not as much that it can be very easily identified with simple segmentation methods. The simulated detector size – and therefore the projection image size – is 128×128 pixels. Examples of radiographs from five objects are given in Figure 2.13.

A total of 100 objects are reserved as training objects, while the other 400 objects are reserved for testing. For each training object, the ground truth corresponding to each radiograph is generated with the workflow, with the same strategy and parameters as in Section 2.4.2. Global thresholding with parameter value $\theta = 0.04$ is used for the reconstructions. For each test object, the 'absolute' ground truth corresponding to each radiograph is generated by directly projecting the virtual foreign objects (Fig. 2.11a and f), thereby skipping the reconstruction and segmentation steps. The projections are segmented such that every nonzero pixel on the detector is a projected foreign object location.

To verify that the direct use of the generated 3D volumes results in very similar ground truth projections compared to when to workflow is followed, the resulting ground truth projections are compared for the training set. The Jaccard index between the resulting ground truth pairs, averaged over all projection angles for all 100 training objects, is 0.961 for SIRT with 100 iterations. This result indicates that the resulting ground truth projections resulting from both approaches are very similar, and both are likely to yield the same quality measures.

To further confirm this, the training of networks as described in Section 2.4.3 is repeated with the simulated projections, with the trained networks this time being evaluated on the test set with 'absolute' ground truth. The results for the three measures are given in Figure 2.14, and are in accordance with the experiments with the laboratory data. A notable difference is that the average class accuracy and detection accuracy reach their maximum values for a relatively lower number of training objects (and the same goes for the minimum value of the false positive rate). This is most likely because the simulated objects are less complex, resulting in radiographs with less complicated structures. Nevertheless, the results again show inferior results for the approach where one radiograph per training object is used, since 100 objects are needed to reach similar quality measure values as for the workflow with only 7 objects.



(**a**) Average class accuracy

(**b**) Object based detection rate

(**c**) Object based false positive detection rate

Figure 2.14: The average class accuracy, the object based detection rate, and the object based false positive detection rate of segmentations with trained U-Net and MSD networks on simulated data for different number of training objects and different training strategies. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

2.5 Discussion

Regarding the difficulty of transferring the method to industrial settings, practical issues such as differences in the CT setup can be expected, but we expect that these are straightforward to overcome. In general, we expect that the proposed workflow can be transferred to industrial applications if the following three conditions are met. First of all, the materials should have different absorptions in the energy range of acquired radiographs (determined by the peak voltage of the source and the energetic detection range of the detector) to be distinguishable at all. Secondly, it should be possible to carry out a segmentation in the reconstruction domain, and the foreign objects should therefore not be too small. Lastly, the foreign objects should be detectable in the radiographs.

In our experiments, the Play-Doh is selected to be representative of many example products in the food industry, and the stones for the related foreign objects. This particular case meets the conditions stated above, and therefore stone can be detected in our examples. Moreover, we expect no problems with metal detection because of its higher X-ray attenuation and visibility in the radiographs, and this high attenuation may allow for even smaller metal foreign object sizes as they are more likely to still appear in the reconstruction and radiographs. On the other hand, if the objects contain large metal pieces and other materials need to be detected, it could lead to artifacts in the reconstructions, but there are many artifact reduction methods available that can be used to mitigate this [101]. Regarding other less dense materials such as plastics, successful application of the workflow strongly depends on the visibility of the foreign objects in the radiographs in the first place. If foreign objects are impossible to discern in radiographs, creating training data by manual annotation is also not possible. Of course, many solutions can be proposed for this invisibility problem, but this discussion is independent of the workflow for training data generation. However, even without advanced imaging methods, applying the workflow and generating 2D ground truth could lead to networks retrieving patterns in the radiographs that are difficult to find by human inspection. Additionally, we have shown in the experiments that multiple foreign objects in a sample do not pose a problem. When the material types among these are varied, we do not expect any problem as long as they are distinguishable in the radiographs. Finally, when transferring the workflow to an industrial setup, we may expect practical issues regarding the CT setup, but this is not a fundamental problem of the proposed method.

Overall, the graphs presenting the foreign object detection accuracies of stones in Play-Doh in Section 2.4 indicate an increase of segmentation and detection accuracy with increasing the number of objects from which the training data are created. The accuracies initially increase strongly with the number of training objects but this increase decays when the number of training objects is further increased. The detection rates and false positives rate introduced in Section 2.4.4

depend on the thresholds η and δ, respectively. If a higher threshold η is used in condition (2.2), the detection rate will decrease, because fewer objects in the target images will meet this condition. Conversely, if the threshold decreases, the detection rate will decrease. Similarly, if the δ threshold in condition (2.3) is increased, there will be more false positives (and fewer when the threshold is decreased). Changing the thresholds η and δ will therefore change Figures 2.8, 2.9 and 2.14 as well, but we expect the overall shape of the curves and their relative distances to each other will remain similar.

When the thresholds are fixed, the maximum detection accuracy that can be achieved depends on the nature of the foreign detection problem. For instance, if the X-ray flux is low and the noise is high, foreign objects are more difficult to detect from the radiographs. In the case of the laboratory experiments, foreign objects are difficult to detect when the cylindrical shape is located with the long edge on the ground and oriented orthogonal to the detector. The radiographs should contain sufficient discriminatory information such that foreign object detection with deep learning is possible. Additionally, for the dataset to be suitable for supervised machine learning, the ground truth should also be of sufficient quality, although this seemed to be less of an issue in our experiments as we observed no negative effects from occasional noise in the ground truth on the training and detection accuracy.

With the above considerations in mind, the workflow is designed to be modular. Every stage of the proposed workflow can be designed according to the available data-acquisition equipment, the intended detection accuracy, the type of base objects and foreign objects, and the available computer memory, among other things. We highlight some possible considerations for every stage:

- **Objects** (Fig. 2.3a): The set of objects can be enlarged or diversified when the accuracy of the trained neural network is not satisfactory. Also, more objects can be added to obtain a more diverse representation of objects when a more diverse array of objects or orientations are considered to be subjected to X-rays in the industrial application, such as on a conveyor belt. When a completely new type of objects is considered, these objects should be added to the workflow as well.

- **Scanning routine** (Fig. 2.3b): In our experimental setting we have used data resulting from low exposure times as input for both the neural networks and the reconstruction algorithm. If the foreign objects turn out to be too difficult to separate in the reconstructions, more scanning angles may be considered. Additionally, if the factory settings are allowed to be altered, higher fluxes, different tube voltages or longer exposure times can be used to obtain radiographs of higher quality, as long as the processing times remain acceptable. Also, more discrimination can be achieved by applying

spectral imaging (dual-energy [235] or multi-energy imaging [82, 199, 274]) such that the neural network can distinguish the foreign objects from the base objects. If changing the quality of the radiographs is not possible, a separate high-quality scan of the same object can be made under the same angles, to achieve more contrast of the foreign object in the reconstructions. The scanning routine can be carried out in any lab, as long as it is done under similar conditions as in the intended industrial X-ray imaging setting.

- **Reconstruction algorithm** (Fig. 2.3e): Depending on the type of data, different reconstruction algorithms may be considered [51, 112]. In this work, we have used the SIRT algorithm to account for the noise in the data, but other reconstruction algorithms such as Feldkamp-Davis-Kress (FDK) algorithm [87] or the Conjugate Gradient method for Least Squares (CGLS) [123] can be considered as well. Also, when dealing with spectral or generic multi-channel data, multi-channel reconstruction methods [143, 238, 246, 254, 312] can be used to increase the reconstruction accuracy even further. When dealing with objects that may change in time, dynamic reconstruction methods can be considered [71, 114, 211].

- **Segmentation algorithm** (Fig. 2.3f): In this work we have used a simple global thresholding scheme, but many more segmentation methods are available, as well as approaches to reduce possible noise [70], or bounding boxes when the location of the foreign object is known [144]. In case of multi-channel data, a multi-dimensional thresholding scheme can be used, as well as clustering methods. Discrete reconstructions algorithms that combine reconstruction and segmentation are also available [30, 121].

- **Virtual projection** (Fig. 2.3g): When creating the virtual projection, post-processing on the generated ground truth projections can be applied to increase the training target quality, for instance by denoising the obtained ground truth projections.

- **Supervised learning** (Fig. 2.3c and h): To validate the workflow, we have used the U-Net architecture with ADAM optimization on cross entropy loss and dice loss, as well as the MSD network with ADAM optimization [146] on the cross-entropy loss. Other neural network architectures (see Section 2.2.2) can also be considered, as well as different optimization strategies and loss functions. Note that the foreign object detection problem considered in this work may be ambiguous, since for a base object containing a foreign object another base object can theoretically be constructed (without foreign object) that results in the same radiograph. This constructed base object may have an unnatural shape when compared with other base objects, but if it happens, it may lead to inconsistent training data for the network. However, this possible problem is independent of the workflow and can be resolved by multi-spectral imaging or multi-angle imaging, and training the networks with multiple images from the same object resulting from these imaging

methods. However, creating reconstructions with data from these advanced imaging methods would not be necessary.

In Section 2.4, we have compared the segmentation results with workflow-generated data from many angles for each object to results with workflow-generated data with only one radiograph for each object. According to the results in Figures 2.8 and 2.14, segmentation and detection accuracy can be improved by using multiple annotated radiographs for each training object. For a comparative classical approach, ideally manual annotation of the data should be carried out, taking into account the variation that may occur in different annotations for the same set of radiographs. In addition to the tremendous effort required to manually annotate a large number of radiographs, there are a number of issues that arise in the food industry. To carry out segmentation on 2D radiographs, highly specific knowledge is required for which it is difficult to find experts, as opposed to trained radiologists in medical imaging. This is also reflected in the lack of suitable publicly available datasets with annotated radiographs of food products. Also, the manual annotations may vary depending on the detection goal by a manufacturer (such as object sizes, positions or number of foreign objects). For these reasons, we eventually chose not to compare the proposed data generation with manual annotation. Manual annotation could still be used in conjunction with this workflow to replace the segmentation step. Whether or not this is feasible and will yield better results will ultimately depend on the specific application at hand.

2.6 Conclusions

In this research, a new workflow is proposed for generating training data for supervised deep learning for foreign object detection in an industrial setting. In this workflow, a number of representative objects are scanned using X-ray imaging, reconstructed using computed tomography, segmented and virtually projected in an objective and reproducible manner to obtain the true foreign object locations in a large set of radiographs, after which supervised machine learning can be applied to detect foreign objects with high accuracy depending on the number representative objects included. We demonstrate this workflow on both laboratory and simulated data using neural networks for the deep learning task. Through laboratory experiments, we have verified that the workflow produces adequate target images. The introduced measures assess the quality of foreign object detection with networks trained using datasets generated with this workflow. All experiments show a consistent result in which the accuracy increases significantly with a few number of training objects, and less significantly for every additional training object. In the laboratory experiment, we consistently obtain high accuracies for detecting gravel in modelling clay with low exposure times using this workflow, demonstrating its application potential in an industrial setting.

Code and data availability

The source code to reproduce all results in this chapter are available on `https://github.com/mzeegers/DeepFODDataGenerator` [309]. These Python scripts make use of open-source toolboxes, in particular the ASTRA toolbox [1, 2], MSD-net [227] and PyTorch [222, 223]. The datasets generated for this chapter are available at Zenodo. Separate submissions are made for the processed data resulting in radiographs with ground truth for object detection [308], as well as the unprocessed CT scan data for complete reproduction of the results in this chapter [307].

3

Hyperspectral data reduction through deep learning

The possibility of hyperspectral imaging opens exciting opportunities to see features of objects that are otherwise invisible. The high data volumes that these techniques generate pose challenges for transmission, storage and processing. The need for reduction of those data volumes is not only important in hyperspectral X-ray imaging, but also in remote sensing, where hyperspectral images taken from planes and satellites are typically large and data reduction through compression before transmission is crucial.

In this section, we propose a novel supervised deep learning approach for combining data reduction and image analysis in an end-to-end architecture. In our approach, the neural network component that performs the reduction is trained such that image features most relevant for the task are preserved in the reduction step. Results for two convolutional neural network architectures and two types of generated datasets show that the proposed Data Reduction CNN (DRCNN) approach can produce more accurate results than existing popular data reduction methods, and can be used in a wide range of problem settings. The integration of knowledge about the task allows for more image compression and higher accuracies compared to standard data reduction methods.

This chapter is based on:

M. T. Zeegers, D. M. Pelt, T. van Leeuwen, R. van Liere, and K. J. Batenburg. "Task-driven learned hyperspectral data reduction using end-to-end supervised deep learning". *Journal of Imaging* 6.12 (2020), p. 132.

3.1 Introduction

In hyperspectral imaging, data are collected in a large number of spectral bins from a wavelength range in the electromagnetic spectrum. It is used in various fields [83], including agriculture classification [110, 184], medical imaging [56, 175, 184], luggage and cargo inspection [79, 92, 184] and food quality assessment [272], as well as with energy-dispersive X-ray spectroscopy (EDX) and electron energy loss spectroscopy (EELS) [267]. In addition to the spatial dimensions, hyperspectral data include the spectral dimension which is typically large [23], often in the order of 10^2 to 10^3 spectral bins [259]. The data have rich information for image processing tasks (for instance segmentation and classification) [259]. Hyperspectral imaging can circumvent calibration issues found in (multi)spectral imaging with a low number of energy bins, such as carefully setting spectral measurement ranges.

A key challenge for hyperspectral imaging systems is handling the *size* of the data, which can be prohibitively large for online processing [273]. Efficient data compression is essential to save storage and reduce transmission load [203], for instance in remote sensing with satellites sending spectral images to the Earth or in high-throughput food quality assessment tasks [83, 272]. For industrial applications, training and running algorithms for classification tasks on full hyperspectral data may be very time consuming [128]. Therefore, *data reduction* steps need to be carried out to reduce data redundancy and size. However, it is not known a priori which spectral bins contain important information, and combining information from many bins may be required for the data interpretation. In addition, bins may contain a low signal-to-noise ratio [161, 267], possibly exacerbated by limited acquisition times in some applications. Moreover, bins that are located close to each other are highly correlated which results in redundant information [257].

The goal of this chapter is to propose a new *convolutional neural network* (CNN)-based approach for hyperspectral data reduction that combines high computational efficiency with strong data reduction (down to just 1 or 2 channels), by making effective use of the relation between the spectral signatures in the data and the specific task that needs to be performed. By attaching a data reduction network to a CNN component for segmentation, the combined network simultaneously learns how to effectively reduce the input data to a low number of images, eliminate spectral redundancy and successfully perform a given task, without the need for parameter tuning. The network adapts to different problem settings and learns how to effectively compress the data for the problem while maintaining accurate segmentation with fast processing times. We assess the performance of the method on a simulated dataset consisting of attenuation-based hyperspectral X-ray projection images, as well as on a simulated dataset based on spectral properties found in remote sensing. These multi-image datasets allow for the method to be evaluated without risks of information leakage between training and test sets [206]. We show that the method is applicable to different CNN architectures by applying

it on a Mixed-Scale Dense (MSD) [227] and a U-Net [240] architecture. The results show that our method is robust to noise and to cases where many different materials or classes are involved, for which standard data reduction methods, such as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA), are not sufficient. To summarize, the main contribution of this chapter is providing a flexible learned supervised data reduction approach with convolutional neural networks with spectral data reduction to a very limited number of images, while retaining high segmentation accuracies.

The remainder of this chapter is structured as follows. Section 3.2 gives an overview of methods for spectral data reduction. In Section 3.3, we introduce notation and the general set of functions in convolutional neural networks to optimize. In addition, we give the problem statement for supervised segmentation and a brief explanation of the most widely used hyperspectral data reduction methods. Most importantly, we introduce our end-to-end hyperspectral data reduction method. In Section 3.4, we describe our experimental setup and the CNN architectures and datasets that are used, including a description of the creation of our simulated attenuation-based hyperspectral X-ray image dataset and simulated remote sensing dataset. Then, we outline the experiments and discuss the results of the data reduction approaches. Section 3.5 discusses the introduced method and the results, and additionally gives some further possibilities for future research. Section 3.6 summarizes the chapter and presents the conclusions.

3.2 Related work

As a result of the importance of data reduction in practical applications, a wide variety of approaches have been developed in earlier work. Two approaches for reducing the dimensionality of a hyperspectral image are hyperspectral band selection and feature extraction methods. Hyperspectral band selection methods select a small number of the bands (bins) to be used for the imaging task, based on searching, ranking, clustering or learning methods [273]. Hyperspectral *feature extraction* methods project the data into a new feature space with a lower dimension. While it changes the meaning of the data, more (combined) information can be stored in the lower dimensional images than the selected bins of band selection methods. In feature extraction methods, a wider range of reduced images can be found and used for the specific task to be carried out than in band selection methods. Common approaches for feature extraction include Principal Component Analysis (PCA) [152, 279] and Linear Discriminant Analysis (LDA) [83, 129, 137, 165], which are popular for their low complexity and absence of parameters. Other common data reduction techniques include Nonnegative Matrix Factorization (NMF) [164, 231], Independent Component Analysis (ICA) [141, 152, 231, 234] and many variants of PCA [152]. More details about the aforementioned methods that appear in this chapter are given in Appendix B.1.

Popular feature extraction methods are fast and do not require parameter tweaking. However, since no task-specific information is used in commonly used unsupervised data reduction methods, features that are important for the given task (i.e. segmentation, classification) may not be preserved in the reduction step. Additionally, other problems such as the inability of PCA to deal with noisy channels and LDA generating only at most one reduction image less than the number of classes [137], as well as the linear nature of these transformations make these approaches less suitable for complex data and feature distributions [278].

There is a wide range of other linear and non-linear data reduction approaches that require different prior knowledge on the data for the image processing task [66]. For example, Kernel PCA (KPCA) makes the transformation of PCA non-linear, but requires the selection of a suitable kernel [10, 86] and introduces the need for parameter tuning. In Locally Linear Embedding (LLE) and similar manifold learning methods, one or more parameters have to be chosen, and the optimal values are different for every dataset [150]. In several cases, the classical linear data reduction methods can outperform the non-linear data reduction techniques [180].

Convolutional neural networks (CNNs) are a powerful tool for classification and segmentation tasks [9, 133, 228]. These have the property to generalize well, as they can non-linearly extract distinctive spatial [163] and spectral properties [132] on different scales for segmentation tasks on noisy data. The current convolutional neural networks for hyperspectral imaging can be classified into three categories [152, 163, 219]. *Spectral CNN* methods apply one-dimensional convolutions in the spectral dimension to classify each pixel. These methods do not take into account that essential spectral information for classification may be located in very distant bins. Additionally, these methods also disregard spatial information [218]. *Spatial CNN* methods first reduce the data with a separate method, for example with PCA. The pixels in the remaining feature maps are then classified using 2D convolutions. This can, for example, be used when executing on-ground recovery of image information by a CNN after compression on-board of a satellite [280]. In these approaches, feature extraction and CNN classification are disconnected [168, 184], so the feature extraction is not tailored to the CNN classification task. *Spectral-spatial CNN* methods take both spectral properties and spatial information into account and many possible designs and strategies can be developed for this, making this set of approaches highly flexible [219]. Data reduction can be integrated (both explicitly and implicitly) into the architecture of this type of CNN. Possibilities include performing 1D convolutions in the spectral dimension and in different layers before applying 2D convolutions in the spatial dimensions (1D+2D CNNs) [21, 59] or applying convolutions in all dimensions simultaneously (3D CNNs) [21]. However, with 3D CNNs, it is not possible to retrieve purely spectrally reduced images. Additionally, some of these spectral-spatial CNN approaches require hyperparameters to be properly tuned [132].

For classification methods that use spectral-spatial CNNs, the computation time can increase significantly with high-dimensional data [169, 218]. In addition, the large number of training parameters makes the network difficult to train and easy to overfit [58], especially if only a few training examples are available, referring to the limited amount of labelled data that is available in hyperspectral imaging [88]. Therefore, some CNN-based approaches still need simple reduction methods such as PCA as a preprocessing step to keep computation times tractable [88, 216], but a learned data reduction approach may outperform these standard reduction approaches [7]. Most current methods reduce to a relatively large number of reduction images (i.e. 32). In contrast, in this work we introduce a learned data reduction approach for CNNs to reduce data to a very limited number of channels (i.e. 1 or 2). By adding a data reduction network to the CNN and training the combined network in an end-to-end fashion, the data reduction becomes task-specific and can be applied with high compression, low computation times and without parameter tuning.

3.3 Materials and methods

3.3.1 Notation and concepts

Hyperspectral imaging

We consider the supervised hyperspectral image segmentation problem. A *hyperspectral image* is a three-dimensional image $x \in \mathbb{R}^{N_b \times m \times n}$ with two spatial dimensions of size m and n and one spectral dimension of size N_b. The number of spectral bins N_b is typically large compared to multispectral images, i.e. between about 100 to 1000 [259].

A *segmented image* is an image $y \in \mathcal{C}^{m \times n}$ in which a class is assigned to each pixel from a finite set $\mathcal{C}$ of classes. A segmentation y of an image x divides the image into regions where the pixels have similar characteristics. For instance, it can divide a hyperspectral satellite image up into regions of classes including water, roads, vegetation, etc. We assume the existence of a true segmentation function $F_s : \mathbb{R}^{N_b \times m \times n} \to \mathcal{C}^{m \times n}$ that maps a hyperspectral image x to its segmented image $y = F_s(x)$. The problem, of course, is that this underlying function F_s is generally not known. Therefore, the aim is to find an approximating segmentation function F such that $F \approx F_s$. Note that both spectral and spatial information are needed for a good segmentation (for example, vegetation can have the same shape but different spectral reflectance, and roads can have different shapes but similar spectral properties).

Supervised learning and neural networks

To solve the problem of finding an appropriate function F, *supervised learning* can be used. In this setting, a set of examples $\{(x_i, y_i)\}_{i=1}^{N^{\text{im}}}$ of hyperspectral images with their segmentations is available, with N^{im} being the number of images. The aim is to approximate the function F_s based on segmentation training data with $y_i = F_s(x_i)$ for every i. In other words, the problem can be summarized as:

$$\text{Find a function } F : \mathbb{R}^{N_b \times m \times n} \to \mathcal{C}^{m \times n} \quad \text{such that } F(x_i) \approx y_i \text{ for every } i. \tag{3.1}$$

The set of images and their segmentations can be partitioned into training, validation and test sets. To solve the supervised hyperspectral segmentation problem, the aim is to find a function $F : \mathbb{R}^{N_b \times m \times n} \to \mathcal{C}^{m \times n}$ such that the loss L, the error between the predicted classes by F from the training examples and their respective target images, is minimized:

$$\min_F \sum_{i=1}^{N^{\text{train}}} L(F(x_i^{\text{train}}), y_i^{\text{train}}). \tag{3.2}$$

To prevent overfitting on the training data, it is evaluated on a separate validation set $\{(x_i^{\text{val}}, y_i^{\text{val}})\}_{i=1}^{N^{\text{val}}}$. The error on this set determines whether training should be continued or not by defining a stopping criterion. Subsequently, the function is tested on a separate test set $\{(x_i^{\text{test}}, y_i^{\text{test}})\}_{i=1}^{N^{\text{test}}}$ to assess the overall performance.

A common approach to find a suitable function F to satisfy Equation (3.1) with supervised learning is to parameterize it as a *neural network*. In many popular neural network architectures for imaging, the input is passed on from layer to layer to create *feature maps*, denoted by $z_i \in \mathbb{R}^{c_i \times m_i \times n_i}$, where c_i is the number of *channels* in the feature map of layer i. The structure is schematically shown in Figure 3.1, and we adopt notation from [227]. To finally produce the output feature map, the feature map $z_0 \in \mathbb{R}^{c_0 \times m_0 \times n_0}$ in the input layer is iteratively passed on from layer $i-1$ to i to produce feature maps $f(z_{i-1}) = z_i$ with $f_i : \mathbb{R}^{c_{i-1} \times m_{i-1} \times n_{i-1}} \to \mathbb{R}^{c_i \times m_i \times n_i}$. In networks for segmentation problems, the number of channels c_d in the output layer is equal to the number of classes N_c, and the feature map z_d contains probability maps for every class.

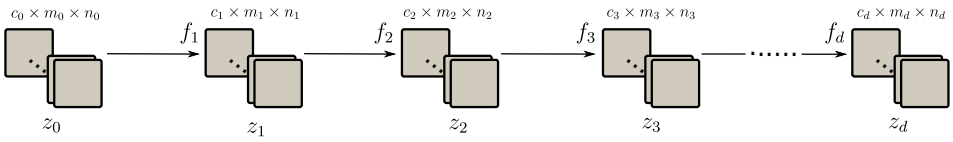


Figure 3.1: Schematic architecture of a neural network for image processing. Squares denote feature map channels and arrows denote function inputs. The depth of the network is given by d.

There also exist more intricate neural network architectures (for instance in [227, 240]) where the feature map z_i in layer i can be written as function depending on the feature maps in all previous layers: $z_i = f_i(z_0, z_1, \ldots, z_{i-1})$, with $f_i : \mathbb{R}^{c_0 \times m_0 \times n_0} \times \ldots \times \mathbb{R}^{c_{i-1} \times m_{i-1} \times n_{i-1}} \to \mathbb{R}^{c_i \times m_i \times n_i}$. This is schematically shown in Figure 3.2.

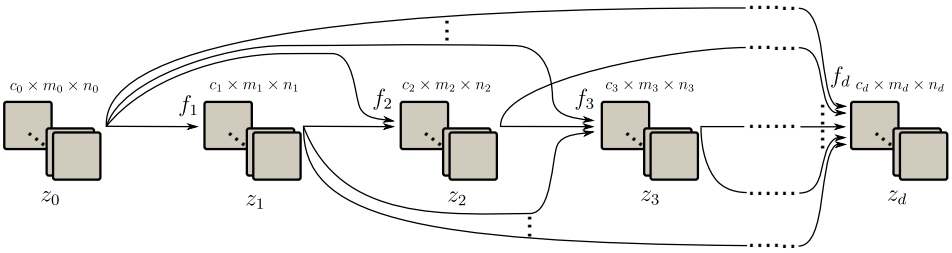


Figure 3.2: Schematic architecture of a neural network with dependencies between all layers. Squares denote feature map channels and arrows denote function inputs. The network depth is given by d.

The network is parameterized with weights and biases which are typically involved in the functions f_i. The entire network can then be written as a function $F_\theta : \mathbb{R}^{c_0 \times m_0 \times n_0} \to \mathbb{R}^{c_d \times m_d \times n_d}$, where $\theta \in \Theta$ contains given values for all weight and bias parameters. Since the final feature map z_d produced by the network contains probability maps, it is usually compared with the *one-hot encoding* of the target image, which marks a pixel in channel i as probability one if the target class label of that pixel is i, and zero otherwise. Denote the one-hot encoding function by P. The aim is now to find a set of parameter values θ such that $P \circ F_s$ is approximated by F_θ.

For imaging, *convolutional neural networks* (CNNs) have proven successful. In these networks the functions f_i are typically operations involving activation functions, bias functions, weighting functions and, by definition, *convolutions*. These functions depend on the previous feature maps z_{i-1} only or on feature maps of all previous layers. The latter case is more general, in which we have the following:

$$f_i(z_0, \ldots, z_{i-1})^j = \sigma\left(\sum_{l=0}^{i-1} \sum_{k=0}^{c_{l-1}} C_{ijkl}(z_l^k) + b_{ij}\right). \tag{3.3}$$

Here $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function, $b_{ij} \in \mathbb{R}$ are the bias parameters and $C_{ijkl} : \mathbb{R}^{m_l \times n_l} \to \mathbb{R}^{m_i \times n_i}$ is the convolution function (including convolution filters) from feature map channel k in layer l to feature map channel j in layer i. During training, the parameters that are being optimized are the biases b_{ij} and the convolutional filters in the convolutions C_{ijkl}. For a CNN $F_{\text{Net}} : \mathbb{R}^{c_0 \times m_0 \times n_0} \to \mathbb{R}^{c_d \times m_d \times n_d}$ with depth d and $c_0 = N_b$ spectral inputs, similar to Equation (3.2), the loss function is minimized over the network parameters $\theta \in \Theta_{\text{Net}}$, to obtain the

network F_{Net_θ}:

$$\min_{\theta \in \Theta_{\mathrm{Net}}} \sum_{i=1}^{N^{\mathrm{train}}} L(F_{\mathrm{Net}_\theta}(x_i^{\mathrm{train}}), P(y_i^{\mathrm{train}})). \tag{3.4}$$

Spectral data reduction

If the uncompressed hyperspectral image data are used directly as input for the CNN, the large data size results in prohibitively long training times [317] and memory requirements [280]. A potential solution to these possible issues can be found by employing a *spectral data reduction* method. This method can be viewed as a function $G : \mathbb{R}^{N_b \times m \times n} \to \mathbb{R}^{N_r \times m \times n}$ that acts on the data x and transforms it to $G(x)$ in a lower-dimensional space with $N_r \ll N_b$. Given a chosen reduction function G, the aim is now to find a CNN $F' : \mathbb{R}^{N_r \times m \times n} \to \mathbb{R}^{N_c \times m \times n}$ that segments the reduced data, such that $F' \circ G$ segments the original data, minimizes (3.4), and therefore approximates $P \circ F_s$:

$$\begin{aligned} &\text{Find a function } F' : \mathbb{R}^{N_r \times m \times n} \to \mathbb{R}^{N_c \times m \times n} \\ &\qquad\qquad \text{such that } (F' \circ G)(x_i) \approx P(y_i) \text{ for every } i. \end{aligned} \tag{3.5}$$

Since the function F' has only N_r inputs, the input data size is strongly reduced, reducing the size of the minimization problem (3.4) and decreasing memory requirements. If $G(x)$ preserves the relevant features in the image x that are required for the segmentation task, the function F' that minimizes (3.4) could, in principle, be more easily found.

In this chapter, a new approach is proposed to reduce the data to a very limited number of input feature map channels, without the need for parameter tuning or prior information about the problem and providing possible advantages such as higher processing speeds.

3.3.2 Learned data reduction method

We will now introduce our proposed task-driven end-to-end *Data Reduction CNN* (DRCNN) approach. The key idea of the method is to include the data reduction in the problem as a neural network to approximate the function $P \circ F_s$:

$$\begin{aligned} \text{Find functions} \quad & G : \mathbb{R}^{N_b \times m \times n} \to \mathbb{R}^{N_r \times m \times n} \\ & F' : \mathbb{R}^{N_r \times m \times n} \to \mathbb{R}^{N_c \times m \times n} \\ & \qquad \text{such that } (F' \circ G)(x_i) \approx P(y_i) \text{ for every } i. \end{aligned} \tag{3.6}$$

Therefore, the method includes a supervised data reduction tailored to the segmentation task, which can be separated from the CNN after being trained

together with the CNN. The resulting new network is a combination of a subnetwork that spectrally reduces data to a given number N_r of feature map channels and a CNN that segments the image from $c_0 = N_r$ input feature map channels. A high-level overview of this approach is given in Figure 3.3. Given a CNN F_{Net}, a compatible data reduction network is given by $G_D : \mathbb{R}^{N_b \times m_0 \times n_0} \to \mathbb{R}^{N_r \times m_0 \times n_0}$ with functions that consist of linear combinations of spectral feature map channels, as opposed to containing spatial convolutional operators as in the CNN. The data reduction layer is characterized by the data reduction list $D = [r_0, r_1, \ldots, r_{d_D}]$, where d_D is the depth of the subnetwork. The number of feature map channels r_i in each layer dictates the data reduction, with $r_0 = N_b$ and $r_{d_D} = N_r = c_0$. The feature maps $\overline{z}_i$ in each layer with $i > 0$ are only dependent on those in the previous layer $\overline{z}_{i-1}$. Therefore, the architecture is of the form shown in Figure 3.1 (where the number of channels c_i in layer i equals r_i and the other spatial dimensions remain unchanged). As a result that the functions in this network are linear combinations (or equivalently, spectral pixel-wise 1×1 convolutions), similar to Equation (3.3), the functions $g_i : \mathbb{R}^{r_{i-1} \times m \times n} \to \mathbb{R}^{r_i \times m \times n}$ mapping the images from layer $i-1$ to i in the data reduction network have the form

$$g_i(\overline{z}_{i-1})^j = \overline{\sigma}\left(\sum_{k=1}^{r_{k-1}} \overline{w}_{ijk} \cdot \overline{z}_{i-1}^k + \overline{b}_{ij}\right).$$

Here, $\overline{\sigma} : \mathbb{R} \to \mathbb{R}$ is the activation function in this subnetwork, $\overline{w}_{ijk} \in \mathbb{R}$ are the linear weights between the k-th image in layer $i-1$ and the j-th image in layer i, $\overline{b}_{ij} \in \mathbb{R}$ is the bias of image j-th image in layer i. The weights $\overline{w}_{ijk}$ and biases $\overline{b}_{ij}$ in these functions determine the set of parameters ϕ that has to be optimized in this data reduction subnetwork. For neural networks, the Rectified Linear Unit (ReLU) function is a commonly used activation function, but it can lead to dying nodes that become inactive and whose activation functions only output zeros once the nodes produce negative output values [178]. Since the final layers in the data reduction subnetwork contain a low number of feature map channels (at times only one), it is more likely that dying nodes will affect the performance of the network more negatively than in other network architectures. Therefore, for activation function $\overline{\sigma}$, we propose to use Leaky ReLU functions in the data reduction subnetwork with leakage parameter $a = 0.01$ in order to avoid the dying ReLU output problem [178].

For this subnetwork, linear layers are used instead of convolutional layers, because we want to compress exclusively in the spectral direction and convolve exclusively in the spatial directions. Since adjacent spectral bins are highly correlated, applying local spectral convolutional operations is not expected to result in informative feature maps. Instead, bins that are more distant from each other should be combined to achieve this. A network with linear combinations is therefore more suitable than a convolutional neural network for learning this transformation as it can learn complex non-linear functions to combine the information from all spectral bins.

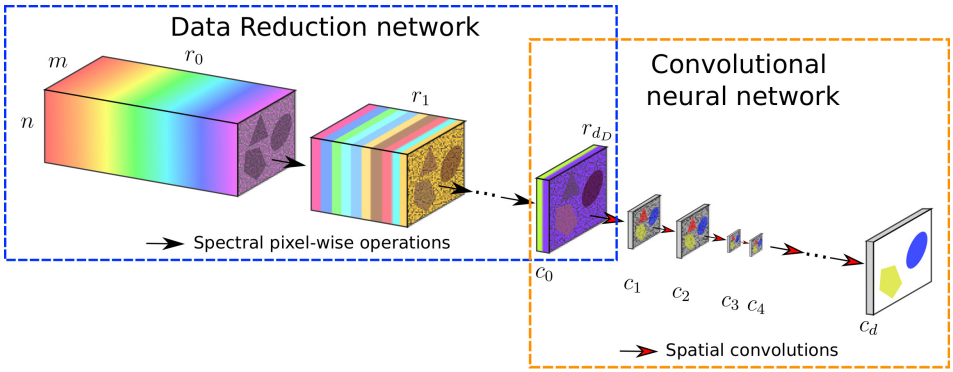


Figure 3.3: Schematic structure of a Data Reduction CNN example. The entire setup consists of a CNN of choice, for example with down- and upscaling layers as shown here, and a data reduction subnetwork in front. This subnetwork repeatedly decreases the number of images from r_0 down to r_{d_D} by taking linear combinations of the input images. After that, the CNN carries out the segmentation task.

Let Θ_{G_D} denote the parameter space in this data reduction subnetwork G_D, compatible with the given CNN F_{Net}. The optimization problem for the joint data reduction and image processing becomes

$$\min_{\substack{\theta \in \Theta_{F_{\text{Net}}} \\ \phi \in \Theta_{g_D}}} \sum_{i=1}^{N^{\text{train}}} L(F_{\text{Net},\theta}(G_{D,\phi}(x_i^{\text{train}})), P(y_i^{\text{train}})). \tag{3.7}$$

The number of parameters to be trained in the data reduction layer is equal to $|\phi| = \sum_{i=0}^{d_D} r_i \cdot r_{i+1} + r_i$. The number of weights in the first layer of the data reduction network provides the leading order of the number of trainable parameter in this subnetwork. As the number of parameters in a CNN can be in the order of millions, the number of parameters in the data reduction layers is relatively small. Moreover, since the CNN has fewer input images, depending on the architecture, the data reduction network may also reduce the number of parameters in the CNN, possibly making a pass through the network faster.

3.4 Experiments and results

3.4.1 Data reduction network architectures

The proposed data reduction approach is designed to be compatible with any existing CNN. In this work, we present results for both the popular U-Net CNN architecture [240] and the recent Mixed-Scale Dense (MSD) architecture [227]. Their architectures and the data reduction integration are explained first, after which the datasets, experiments and results on the datasets are outlined.

Data reduction multi-scale dense net

In the MSD network, all features maps are fully-connected and the operations are dilated convolutions (also called atrous convolutions) to capture image features at different scales. As in the paper that introduced the MSD network structure [227], a network width of $w = 1$ works well in our experiments (or equivalently, setting all values c_i to 1 for $i > 0$). Figure 3.4 gives an example of a Data Reduction MSD (DRMSD) net layout with a reduction to $N_r = 2$ feature map channels, where the depth of the data reduction net is equal to $d_D = 2$, the reduction scheme is $D = [8, 4, 2]$ and the depth of the MSD net is $d = 5$. In this work, we use a common depth of $d = 100$ for this MSD network.

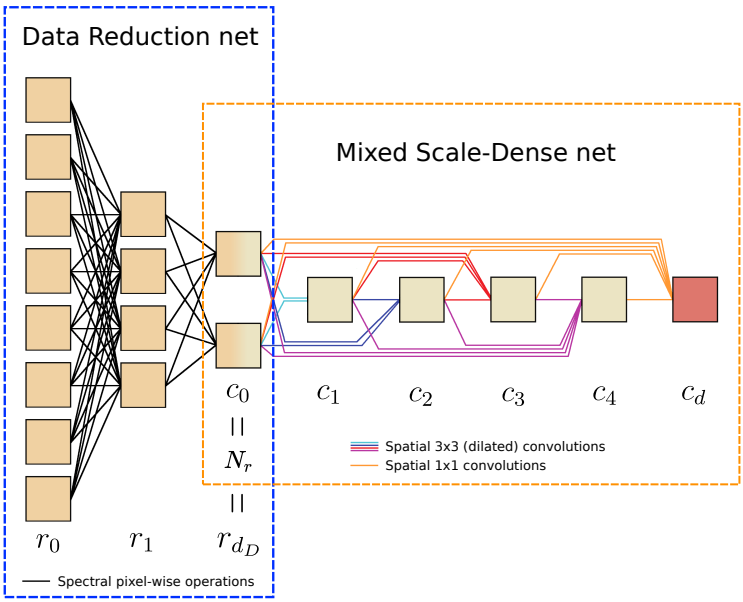


Figure 3.4: Example of a Data Reduction Mixed-Scale Dense (MSD) network structure. The number of channels are indicated with the feature maps. Since $w = 1$ is chosen, $c_i = 1$ for $i > 0$. The data are reduced from $r_0 = 8$ input images to $r_{d_D} = N_r = c_0 = 2$ feature map channels in the data reduction net, while the segmentation task is performed by an MSD net of depth $d = 5$. Each 3×3 convolution is followed by a Rectified Linear Unit (ReLU) operation.

The dilations in each convolutional layer range from 1 to 10, repeatedly increasing from 1 to 10 over the depth of $d = 100$. All dilated convolutions are followed by a ReLU operation. All bias and weight parameters are initialized to zero. For the convolution weights, Xavier initialization is used. During training, ADAM optimization [146] is used on the cross-entropy loss between the data and the predictions. We use the CPU and GPU implementations in Python of [225, 227], with additional CPU and GPU implementation for the data reduction component. Each network is trained on one GPU core of a GeForce GTX TITAN X with CUDA version 10.1.243.

Data reduction U-Net

A second CNN architecture that is used for the experiments is the commonly used U-Net. An example of a Data Reduction U-Net (DRUNet) with reduction scheme $D = [8, 4, 2]$ is given in Figure 3.5. In this example, the data are reduced to $N_r = 2$ feature map channels, which in turn is the input for the U-Net subnetwork. In the U-Net architecture used in our experiments, the feature maps are downsampled twice, with a stride of 2, and the initial number of feature map channels is $c_1 = 128$. Bilinear interpolation is used for upsampling. The number of feature map channels doubles in each downsampling layer, which gives $c_2 = c_5 = 256$, $c_3 = 512$, $c_4 = 768$, $c_6 = 384$ and $c_7 = c_1 = 128$. All downsampling and upsampling operations are preceded and followed by a spatial 3×3 convolution operation with zero padding, each of which is followed by a ReLU activation function. All biases and weights in the data reduction layers are initialized to zero, whereas the biases and convolution weights are initialized by sampling from $\mathcal{U}(-\sqrt{k}, \sqrt{k})$, where $k = \frac{1}{c_{\text{in}} \cdot a^2}$ is the range,

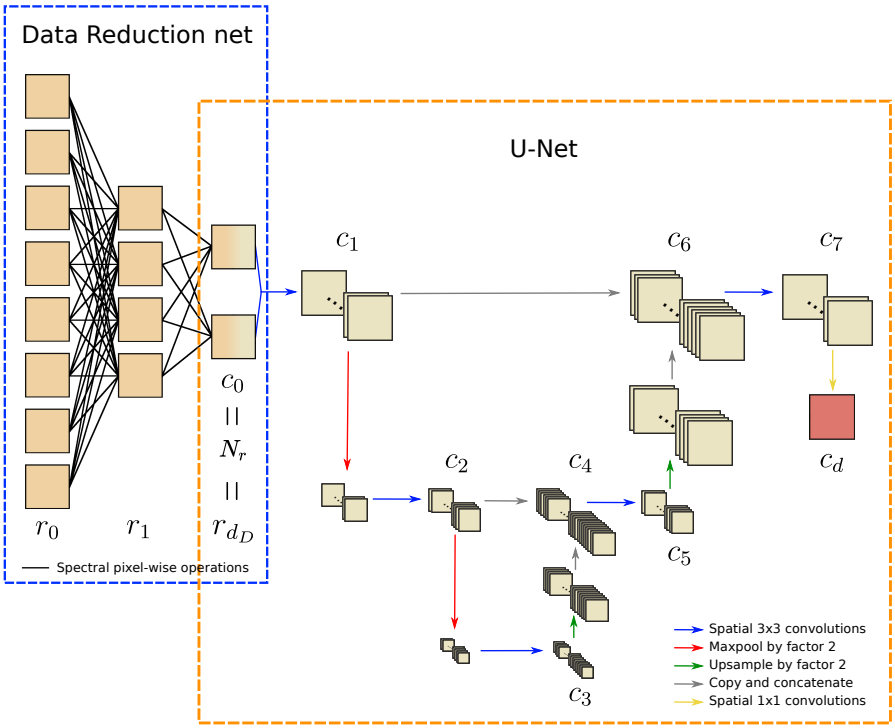


Figure 3.5: Example of a Data Reduction U-Net structure. The number of channels are indicated with the feature maps. The data are reduced from $r_0 = 8$ input images to $r_{d_D} = N_r = c_0 = 2$ feature map channels in the data reduction net. Each 3×3 convolution is followed by a ReLU operation. The number of channels is shown after each convolution and concatenation operations. The network is designed to have $c_1 = c_7$, $c_2 = c_5 = 2c_1$, $c_3 = 4c_1$, $c_4 = 6c_1$ and $c_6 = 3c_1$. The value of c_d is equal to the number of segmentation classes $|\mathcal{C}|$.

c_{in} is the number of input channels and a is the kernel size. During training, ADAM optimization is used on the average of the binary cross entropy loss and the dice loss [135, 270] between the data and the predictions. The network is implemented using PyTorch [222, 223] and is trained on one GeForce GTX TITAN X GPU core with CUDA version 10.1.243.

3.4.2 Datasets

In this section, we will introduce the datasets that are used in the experiments. Hyperspectral data reduction methods are commonly compared using satellite datasets which consist of one hyperspectral image of a certain location (Pavia University, Indian Pines, and Salinas for example [104]), with annotated ground truth segmentation values, some of which are very rare in the image. The use of only one image may cause information leaks between training and test sets when evaluating convolutional neural networks on the same image [206], since overlapping spatial information is likely to be used for classifying pixels in both sets. The availability of other hyperspectral datasets with multiple labelled training samples is limited [259, 306]. We opt to use generated artificial hyperspectral X-ray and remote sensing datasets to resolve this problem, since they consist of multiple images that can be divided into different independent sets.

Simulated attenuation-based hyperspectral X-ray dataset

The first simulated dataset on which we test the method is based on the physical properties in hyperspectral X-ray imaging, including the geometric setup, source spectrum and attenuation properties of various materials. We leave out other effects such as scattering and detector responses as they do not substantially contribute to the understanding of the data reduction network properties. The dataset contains 100 2D images of size 512×512 consisting of $N_b = 300$ spectral bins each. These are simulated X-ray projections of 3D volumes of $1024 \times 1024 \times 1024$ voxels containing 120 cylinders with randomized lengths, thicknesses, angles and positions. A schematic overview of the simulated X-ray setup is given in Figure 3.6. A virtual source and a virtual detector of size 1536×1536 are placed in front and behind the object, respectively, and we use the ASTRA toolbox [1, 2] to compute the projections of size 512×512 from this geometric setup. An example of a projection of 120 cylinders and one cylinder is given in Figure 3.7.

For the experiments, we assign materials to these 120 cylinders, by means of assigning atomic numbers in two different setups. In the first *few-material* setup, we assign atomic number 47 (silver) to two randomly chosen cylinders whereas the remainder is assigned 48 (cadmium). In the second *many-material* setup, each material from atomic numbers 30 (zinc) up to 89 (actinium) is uniquely assigned to two randomly chosen cylinders. To prevent the cylinders with high atomic numbers to be too highly attenuating, the cylinders consist of a mix of 99% polyethylene and 1% of the assigned material. An overview of a selection of the attenuation

spectra is given in Figure 3.8a. The spectra are taken from the National Institute for Standards and Technology (NIST) [130]. Further details on the setup and on computing the projections are given in Appendix B.2.

Poisson noise is applied to both the projection images and the flatfield images, i.e. projection images without objects. In this case, the flatfield images are averaged over 50 separate flatfield images. As a result of the shape of the source spectrum $\overline{I}_0$, shown in Figure 3.8b, the flux of photons is lowest at low and high energies. Therefore, the bins corresponding to energies close to 13kV and 70 kV are more noisy than the others. Example images of noisy and clean data from the few-material datasets are given in Figure 3.9. We combine the clean and noisy setups with the few-material and the many-material settings, resulting in four combinations of datasets. The data are 31.5 GB in size for every combination.

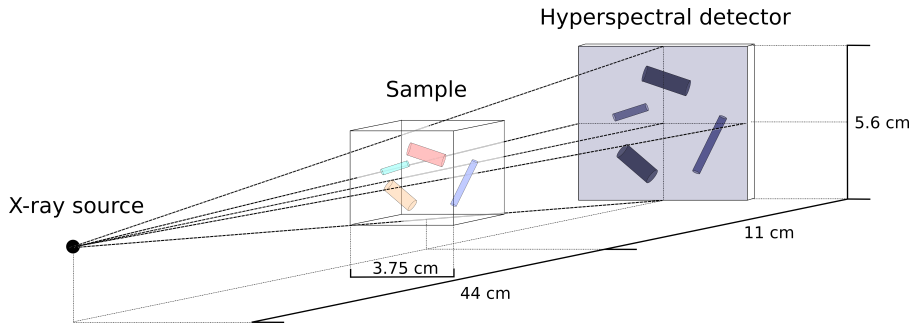
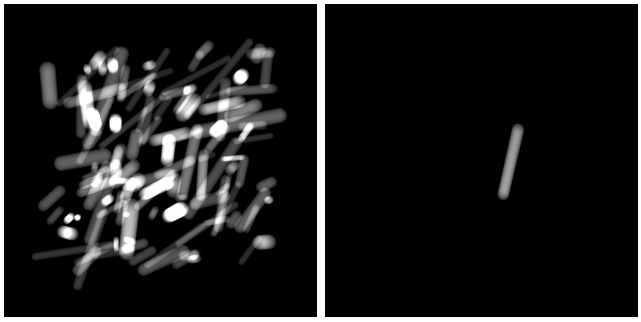
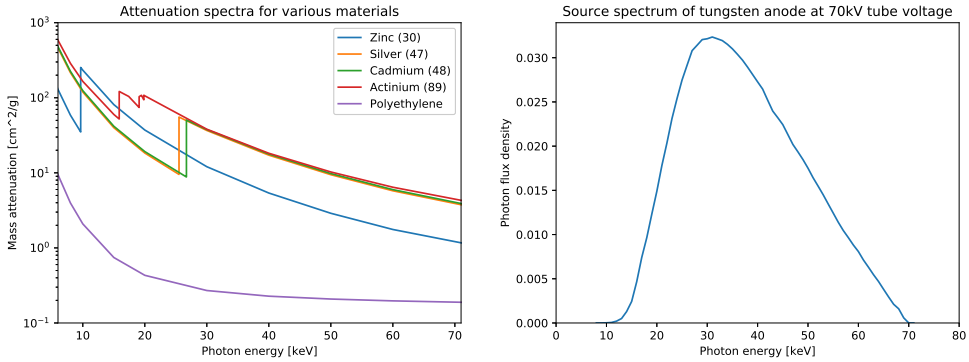


Figure 3.6: Schematic overview of the hyperspectral X-ray projection setup with a cone beam geometry.



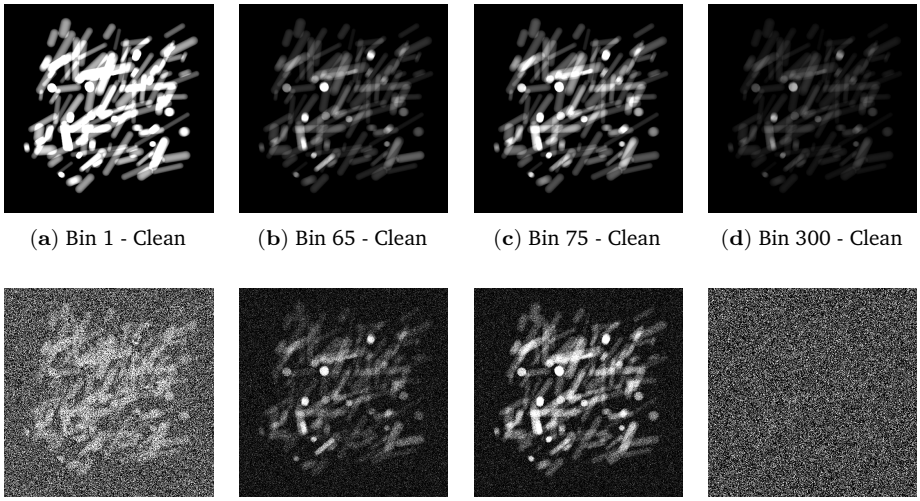
(**a**) All cylinders (**b**) One cylinder

Figure 3.7: Example of the simulated material projections before material designation. The cylinders are shown to be all combined in one image (**a**), and separately (**b**).



(**a**) Mass attenuation spectra (**b**) Source spectrum

Figure 3.8: Mass attenuation spectra for zinc, silver, cadmium, actinium and polyethylene from 6 keV to 71 keV (**a**). In this spectral region, zinc, cadmium and actinium have one K-edge, polyethylene has none, while actinium has multiple edges. Note that the K-edges of silver and cadmium are relatively close to each other. This holds for all adjacent atomic numbers (not shown in this figure). (**b**) The normalized plot of the source spectrum $\overline{I}_0$ used for generating the hyperspectral X-ray projections.



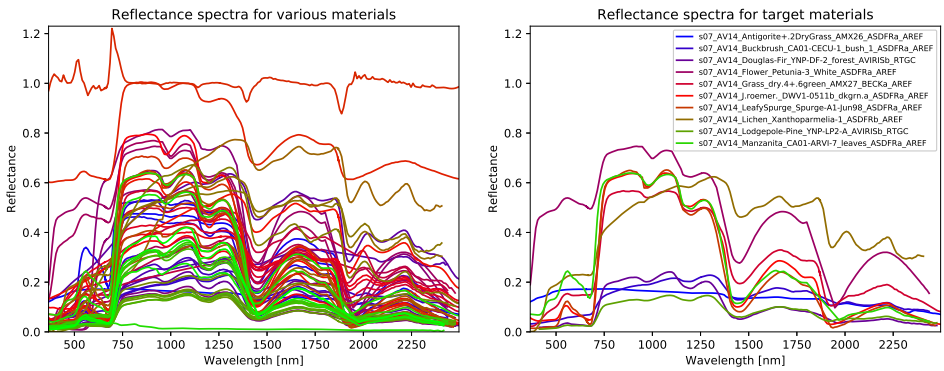
(**a**) Bin 1 - Clean (**b**) Bin 65 - Clean (**c**) Bin 75 - Clean (**d**) Bin 300 - Clean

(**e**) Bin 1 - Noisy (**f**) Bin 65 - Noisy (**g**) Bin 75 - Noisy (**h**) Bin 300 - Noisy

Figure 3.9: Visualization of the simulated X-ray data at different bins. The K-edge transition of cadmium is visible between bins 65 and 75 (among (**a**)-(**d**), compare (**b**,**c**)). The data in bins 1 and 300 (**e**,**h**) are much more noisy than in bins 65 and 75 (**f**,**g**), due to low source spectrum values at bin 1 and 300.

Simulated reflectance-based hyperspectral remote sensing dataset

In addition to the previous dataset, we created a simulated dataset where the spectral properties are taken from remote sensing settings. This dataset again contains 100 2D images of size 512×512, now consisting of $N_b = 200$ spectral bins. We create 360 cylinders of different sizes and place these in the images such that none of these overlap. Each of these cylinders is assigned a material in such a way that there are 60 different materials with 6 cylinders each. The reflectance spectra are taken from the United States Geological Survey (USGS) High Resolution Spectral Library [148, 149], which contains a wide variety of reflectance spectra for liquids, minerals, soils and vegetation, among other categories. The spectra used for this dataset (Figure 3.10a) are randomly drawn from the vegetation section. For the experiments, we consistently choose 10 labels out of the 60 which need to be detected. The spectra of these materials are given in Figure 3.10b. The spectral range we use is from 450 nm to 2400 nm, and we only use materials for which the full spectrum in this range is included in the library.



(**a**) All 60 spectra (**b**) 10 target spectra

Figure 3.10: Reflectance spectra used for this dataset (**a**). The 10 target spectra on the right (**b**) are a subset of the 60 spectra. The filenames of these target spectra in the USGS Library are added.

The reflectance spectra are multiplied by the solar irradiance spectrum, which is the base intensity received from the Sun. We use the spectrum AM1.5 (G-173-03 International standard) global [268] from the American Society for Testing and Materials (ASTM) [19] that gives terrestrial solar spectral irradiance on a surface under certain conditions such as orientation towards the Sun, temperature, pressure and atmosphere composition, among other conditions [268]. The solar irradiance spectrum is given in Figure 3.11. At certain wavelengths there is low transmittance through the atmosphere of the Earth due to presence of certain substances, for instance carbon dioxide, oxygen and most importantly water vapor. On the resulting images, we apply Gaussian noise where the standard deviation is $\frac{1}{1000}$ of

the maximum signal in the dataset. The resulting images are the measured signals. There are many ways to normalize the radiance data, and we apply a flatfield correction [107] using the solar irradiance spectrum given in Figure 3.11. The bins located in regions where solar irradiance is blocked have a very low signal-to-noise ratio.

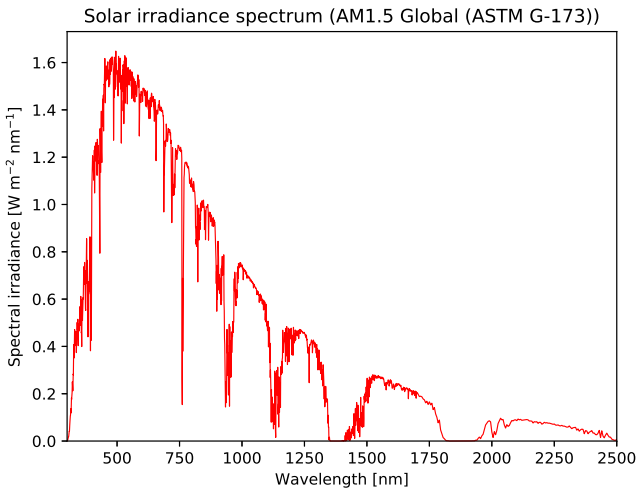
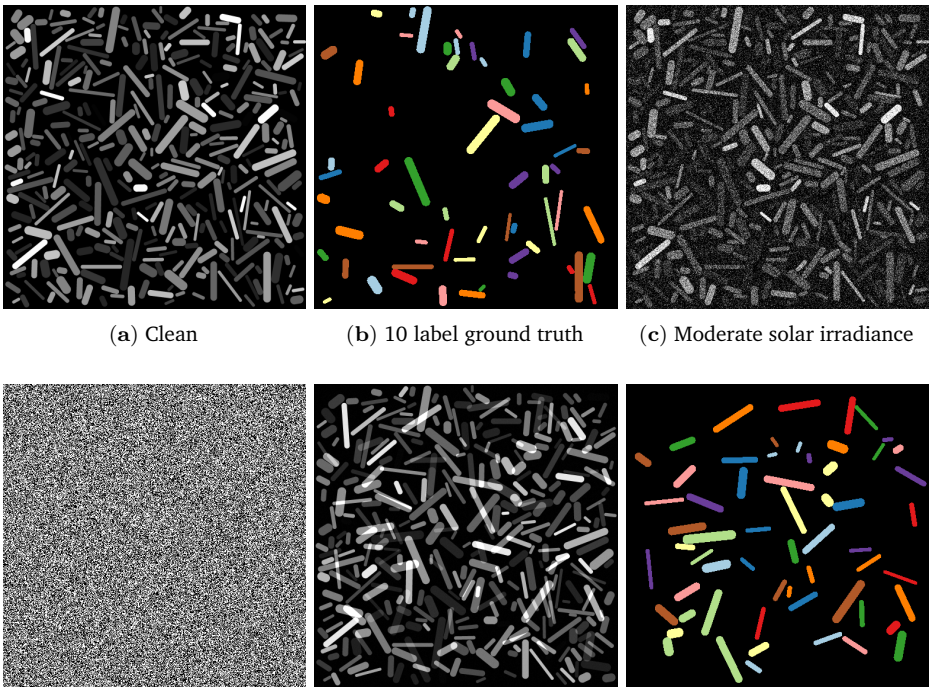


Figure 3.11: The solar irradiance spectrum used for the remote sensing experiments. Note that the drops to a value close to 0 in the graph, particularly at wavelengths 1350–1400 nm and 1800–1950 nm, are mostly due to absorption by water vapor.

In addition to the data described above, we created a dataset where the reflectance images of the 10 target materials are imposed on those of the 50 remaining materials. This creates overlap between these two material sets but keeps the materials within these sets non-overlapping. This simulates mixed material reflectance signals that are likely to occur in realistic remote sensing data. Figure 3.12 shows visualized examples of the data. Note that the difference with the hyperspectral X-ray dataset is that we now have 11 classes instead of 2 to classify pixels into.



(**a**) Clean (**b**) 10 label ground truth (**c**) Moderate solar irradiance

(**d**) Low solar irradiance (**e**) Overlapping data (**f**) Ground truth of overlapping data

Figure 3.12: Visualization of the simulated remote sensing data. The clean data and the ground truth are shown in (**a**) and (**b**), respectively. When the described Gaussian noise is added to this data, many bins still resemble the clean data, but (**c**) shows a moderately noisy bin and (**d**) shows an extremely noisy bin, resulting from differences in solar irradiance. The data with overlapping cylinders and their representation as ground truths are given in (**e**,**f**).

3.4.3 Implementation of standard data reduction methods

In the experiments, we will compare the proposed data reduction method with standard data reduction methods (PCA, NMF and LDA). These methods are implemented using Scikit-learn (version 0.22.1) [224]. In all cases, the default settings have been used. For memory limitation reasons, a subset of the data points is used to compute each transformation, where every k data points in the images are included. For computing the standard data reduction transformation, we sample every $k = 2$ data points when employing PCA, every $k = 5$ for NMF and every $k = 6$ for LDA. When reducing to 200 feature map channels, we use $k = 3$ data points for PCA, $k = 6$ for NMF and $k = 6$ for LDA.

3.4.4 Results

In this section, we describe all experiments, and outline the results. For each experiment, we use $N^{\text{train}} = 70$ images for training, $N^{\text{val}} = 20$ for validation and $N^{\text{test}} = 10$ for testing from the relevant datasets. During training, data augmentation is applied by rotating and flipping the input images, resulting in a total of $70 \cdot 8 = 560$ training images. The chosen training time of MSD is 2 days on the hyperspectral X-ray data (ranging from ca. 300 to 6500 epochs), and 3 days on the generated remote sensing data (ranging from ca. 1500 to 9000 epochs), based on the results on the validation sets. After these time durations most networks did not show any more improvement on the validation sets. The training lengths for U-Net are fixed to 3000 epochs (roughly 1.5 days on average), as the U-Nets converged faster. Despite the possibility to employ multilayered data reduction networks, we experienced that a depth of $d_D = 1$ yielded the best segmentation results. The additional advantage of this is that tuning the sizes of intermediate data reduction layers can be avoided. In addition, the number of trainable parameters and the processing and training times are also slightly lower as a consequence. For all experiments, we take the network with the best performance on the validation set and measure its average class accuracy over all 10 segmentation images in the test set. The average class accuracy is the average number of correctly classified pixels per class relative to the total number of pixels in that class, given by:

$$\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \frac{TP_c}{TP_c + FN_c}$$

Here, $\mathcal{C}$ is the set of classes, TP_c is the number of correctly classified (true positive) pixels with true class c and FN_c is the number of incorrectly classified (false negative) pixels with true class c. To test the robustness of our method, we present averages and standard deviations over 8 runs for a few selected experiments in Appendix B.4.

Noise and multiple materials

For the assessment of the segmentation accuracies of DRCNN compared to those using other data reduction methods, we apply reduction to $N_r = 2$ feature map channels while varying the data reduction method on both the simulated hyperspectral X-ray dataset and the remote sensing datasets. To determine what properties of the data contribute to the performance of the methods, we vary the inclusion of noise and the number of materials (2 or 60) for the X-ray dataset (Section B.2). For the remote sensing dataset, we vary the inclusion of noise and overlapping materials (Section 3.4.2). MSD and U-Net networks without data reduction are trained on the full hyperspectral data as well. Note that since there are only two target classes in the X-ray dataset, LDA will reduce the data to one channel.

				PCA	NMF	LDA	DRCNN	No red.
X-ray	**MSD**	Clean	2 mat.	99.70	**99.73**	**99.73**	**99.73**	99.37
			60 mat.	*60.36*	*82.69*	**99.69**	99.68	99.42
		Noisy	2 mat.	*50.00*	*57.31*	*90.39*	**99.11**	98.60
			60 mat.	*50.00*	*52.31*	*75.53*	**99.16**	98.77
	U-Net	Clean	No overlap	99.66	99.67	99.67	99.69	**99.72**
			Overlap	*79.64*	*82.92*	**99.72**	99.69	99.69
		Noisy	No overlap	*50.00*	*50.00*	*79.49*	**98.92**	98.40
			Overlap	*50.00*	*54.79*	*85.11*	98.69	**98.86**
Remote sensing	**MSD**	Clean	2 mat.	99.87	99.85	**99.94**	99.75	99.90
			60 mat.	98.29	97.99	**99.97**	98.74	99.33
		Noisy	2 mat.	*9.09*	*9.09*	*91.15*	99.76	**99.86**
			60 mat.	*9.09*	*9.09*	*90.53*	97.98	**99.17**
	U-Net	Clean	No overlap	97.97	99.81	99.87	99.95	**99.97**
			Overlap	*95.39*	*96.78*	**99.98**	98.82	99.28
		Noisy	No overlap	*9.09*	*9.09*	*92.10*	99.76	**99.87**
			Overlap	*9.09*	*9.09*	*90.20*	**98.76**	98.13

Table 3.1: Average class accuracies for various datasets and data reduction methods for reductions to two feature map channels (except for Linear Discriminant Analysis (LDA) on the X-ray dataset, which is reduced to one feature map channel, since there are only two target labels). Accuracies below 97.5% are indicated in *italic red*, and the highest value(s) per row for each dataset and CNN are indicated in **bold**.

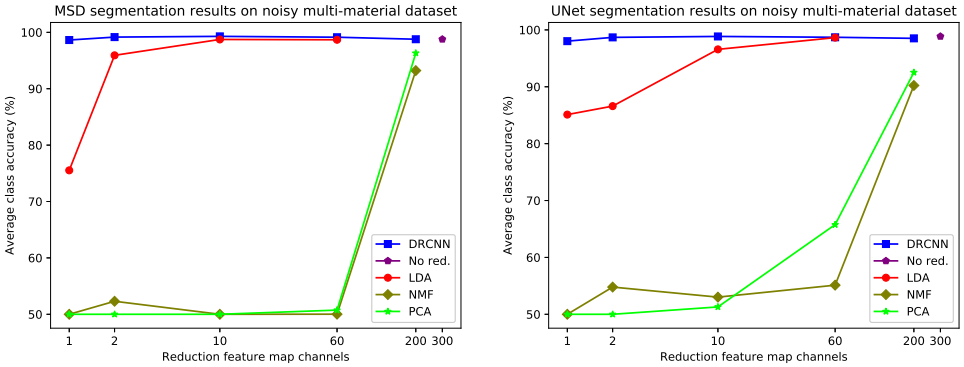
The results of the experiments on the X-ray datasets and the remote sensing datasets are summarized in Table 3.1 for MSD and U-Net. All data reduction methods obtain high accuracy ($> 99\%$) in the case of clean data and two materials. However, PCA and NMF obtain a significantly lower accuracy when multiple materials are introduced ($< 83\%$), while LDA, DRCNN and CNN without reduction retain the high accuracy ($> 99\%$). For two-material data, when noise is introduced, the average class accuracies for NMF, PCA and to a lesser extent LDA decrease ($< 91\%$), while DRMSD, DRUNet, MSD and U-Net still maintain high accuracy ($> 98\%$). This difference for LDA is amplified with MSD when dealing with many materials in a noisy setting, with LDA having a notably lower accuracy than in the two-material setting (at ca. 76% accuracy), although this trend is not seen with U-Net (ca. 85%). The other two data reduction methods have a significantly reduced averaged class accuracy ($< 58\%$). Both DRMSD and DRUNet have a high performance ($> 98.5\%$) in this case, showing that they are robust to both noise

and the inclusion of multiple materials. The experiments on the remote sensing datasets show a similar trend. Table 3.1 shows that the robustness of DRMSD and DRUNet remains (ca. 98% or higher), but the differences with LDA when including noise are smaller ($> 90\%$). For PCA and NMF the accuracy decreases only slightly when overlap is included, but does decrease significantly when noise is introduced (down to ca. 9%). To summarize, DRMSD and DRUNet are shown to be able to remain mostly robust to noisy and multiple materials (ca. 98% or higher accuracies) in these simulated datasets.

Number of reduction feature map channels

To assess the image quality of the DRCNN as a function of the number of spectrally reduced feature map channels, we carried out experiments for both datasets. For the X-ray dataset, we use 1, 2, 10, 60 and 200 feature map channels. For the remote sensing dataset we focus on small feature maps, using 1, 2 and 10 channels. For the latter dataset, we also vary the properties of the dataset (clean/noise and overlap) to assess the influence of those properties on the performance, and a layered reduction to 2 and subsequently 1 map is added. For comparison, the performance using the other standard data reduction methods is also assessed, and for reduction of LDA to more than one feature map channel with the X-ray data, we add prior knowledge about the presence of other materials in the ground truth. For reduction to 2 feature map channels, the 59 other materials are added as one additional class. For reduction to 10 channels, these other materials are added as 9 classes consisting of 6 to 7 materials each, grouped by their atomic numbers. For reduction to 60 channels, each remaining material is added as a separate class. Note that this prior knowledge is in many practical cases not available, so this constitutes an artificial comparison.

Figure 3.13 outlines the results of the experiment for the noisy multi-material X-ray dataset. First of all, DRMSD and DRUNet have high average class accuracies for all numbers of reduction feature map channels. The accuracy is highest when reducing to 2 or 10 feature map channels, and the reduction to only 1 feature map channel is only slightly lower in comparison. In any case, all accuracies for DRMSD and DRUNet are higher than 98%. By contrast, PCA and NMF reach an accuracy of more than 90% when 200 or more reduction map channels are used, but for 60 or fewer reduction channels the accuracies remain below 70%, showing that both data reduction methods are not suitable when 60 or fewer channels are required to reduce the data to. For LDA, the data are reduced to 1 feature map channel and the accuracy is lower than both DRMSD and DRUNet ($< 85\%$). It is only when prior knowledge about the 59 other materials is included that the accuracy approaches that of the DRCNN methods, and even then it may only attain a similar accuracy when the data are reduced to more than 2 feature map channels.



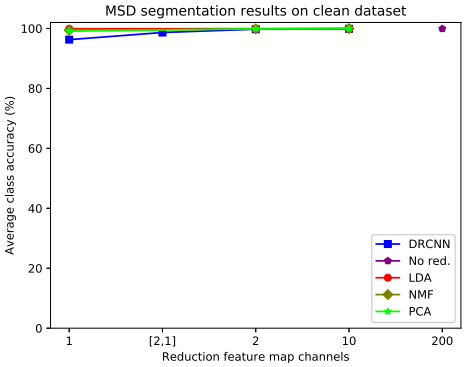
(**a**) MSD average class accuracy

(**b**) U-Net average class accuracy

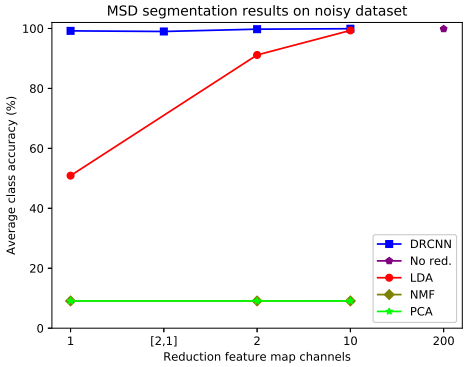
Figure 3.13: Average class accuracies for different data reduction methods using MSD (**a**) and U-Net (**b**) on the noisy multi-material X-ray dataset. As a reference, the results for standard MSD and U-Net networks are included, which act directly on all 300 spectral bins.

For the remote sensing datasets, the quantitative segmentation results of applying the trained networks to the test sets for DRMSD and DRUNet are given in Figures 3.14 and 3.15 respectively, broken down into the four combinations of noise and overlap properties. The visual results for MSD are given in Figure 3.16. DRMSD and DRUNet show high accuracies ($> 99\%$) on the standard noisy dataset for all numbers of reduction map channels, but when the target cylinders are overlapping other cylinders, the accuracies for DRMSD and DRUNet become slightly lower (about 95.5%) when compressing to one feature map channel. In this case, both DRCNNs do not miss any cylinders but there are artifacts on some detected cylinders. However, when the number of reduction feature map channels increases, the accuracy increases rapidly (to ca. 98% and higher). The layered reduction $D = [2, 1]$ images show no strong additional value over direct reductions to 1 image in any experiment. On the other hand, the LDA accuracy is considerably lower for reductions to 1 or 2 feature map channels for all noisy datasets ($< 70\%$ and $< 93\%$, respectively) and reaches a comparable accuracy with reductions to between 3 and 10 feature map channels. From the visual results it can be observed that segmentation from LDA reduced data causes the network to completely miss certain cylinders. In all experiments with noisy data reduced with PCA and NMF, the accuracy is ca. 9%.

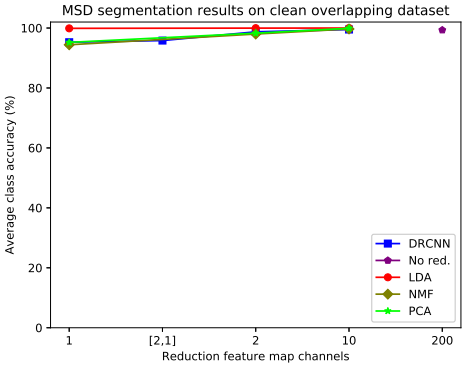
We conclude that for both datasets and most of the properties we investigated, the image quality of the DRCNNs is acceptable for reductions to one feature map channel and higher accuracies are attained when the number of reduction feature map channels is slightly increased. The methods perform favorably compared to the common data reduction methods.



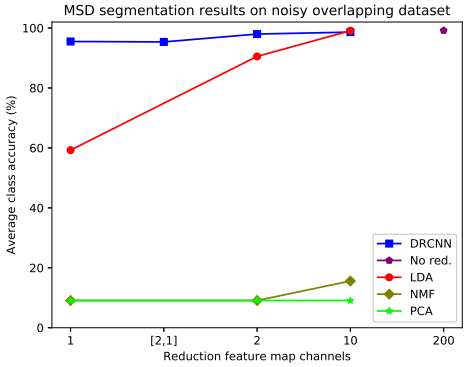
(**a**) Clean dataset



(**b**) Noisy dataset

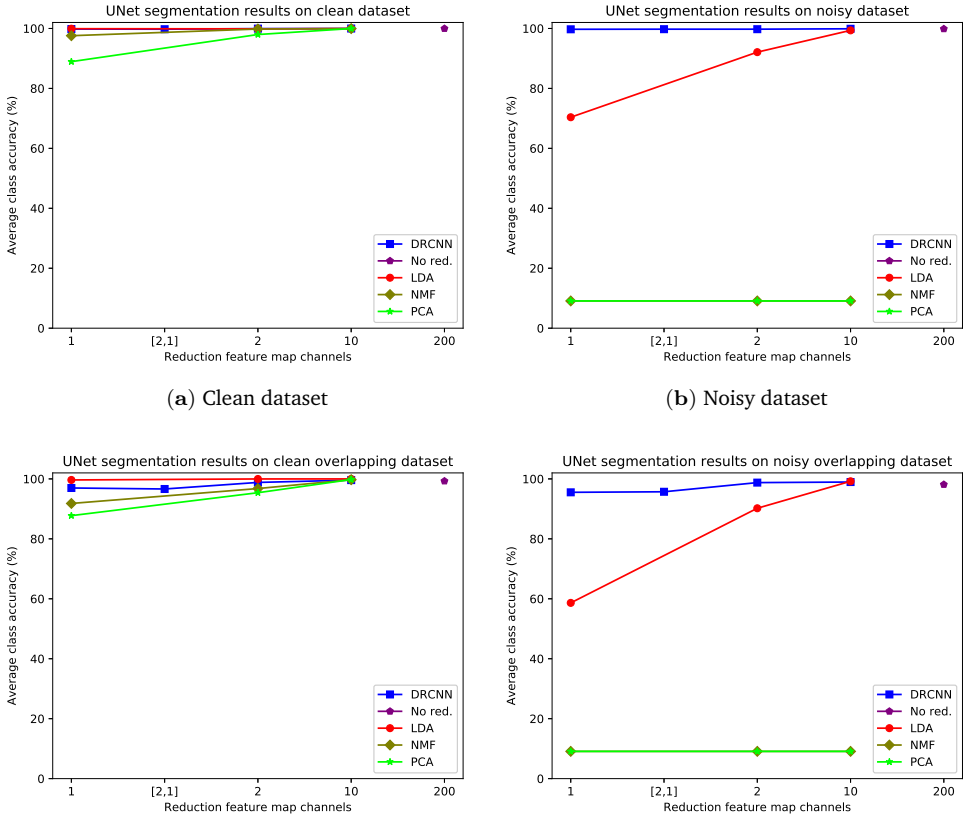


(**c**) Clean dataset with overlapping cylinders



(**d**) Noisy dataset with overlapping cylinders

Figure 3.14: Average class accuracies for different reduction schemes with MSD as CNN for different simulated remote sensing datasets: clean dataset (**a**), noisy dataset (**b**), clean overlapping dataset (**c**) and noisy overlapping dataset (**d**). The layered reductions to 2 and then 1 feature map channel(s) are indicated by $[2, 1]$.



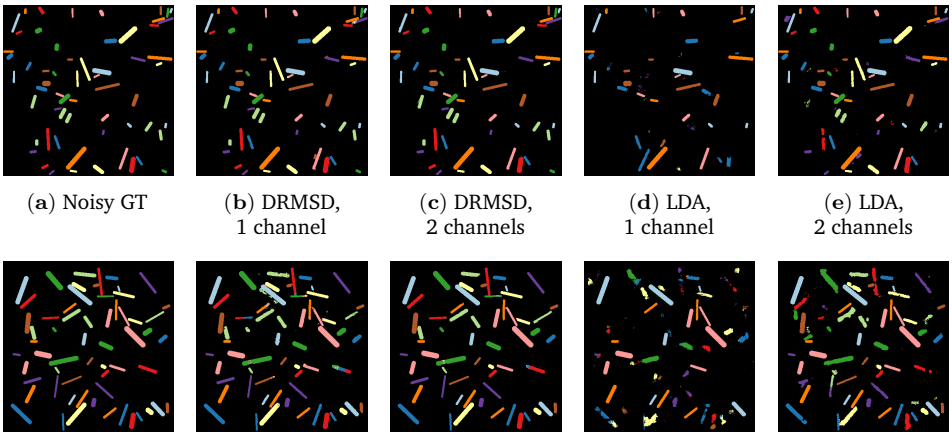
(**a**) Clean dataset

(**b**) Noisy dataset

(**c**) Clean dataset with overlapping cylinders

(**d**) Noisy dataset with overlapping cylinders

Figure 3.15: Average class accuracies for different reduction schemes with U-Net as CNN for different simulated remote sensing datasets: clean dataset (**a**), noisy dataset (**b**), clean overlapping dataset (**c**) and noisy overlapping dataset (**d**). The layered reduction to 2 and then 1 feature map channel(s) is indicated by $[2, 1]$.



(**a**) Noisy GT (**b**) DRMSD, 1 channel (**c**) DRMSD, 2 channels (**d**) LDA, 1 channel (**e**) LDA, 2 channels

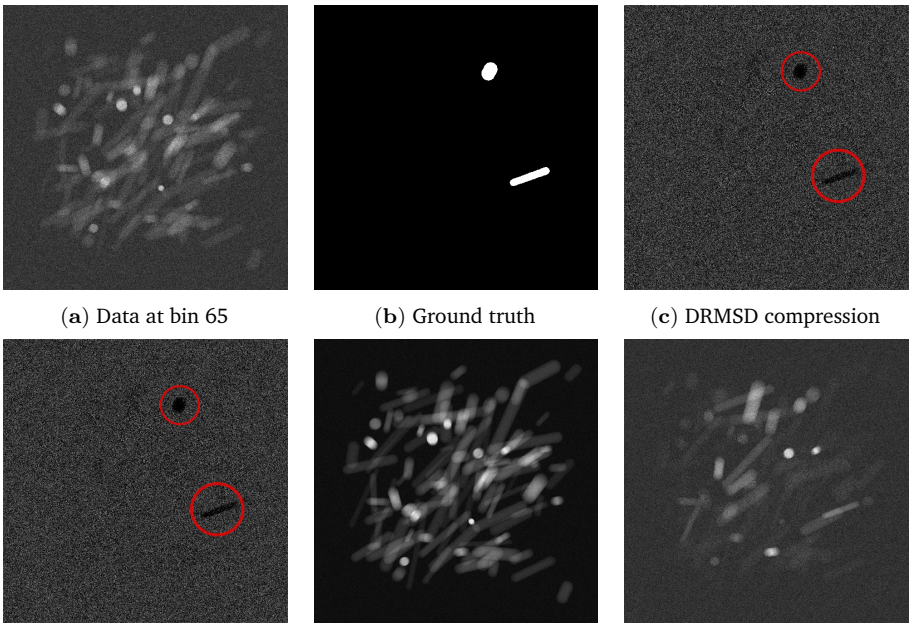
(**f**) Noisy overlap GT (**g**) DRMSD, 1 channel (**h**) DRMSD, 2 channels (**i**) LDA, 1 channel (**j**) LDA, 2 channels

Figure 3.16: Visual results for LDA and DRMSD reduction schemes for reductions to 1 and 2 feature map channels on the noisy dataset (**a–e**) and the noisy overlapping dataset (**f–j**).

Dependence of feature map properties on data reduction method

In the final experiment, we look at the properties of trained networks on the datasets when the number of reduced feature map channels is set to $N_r = 1$. For the X-ray dataset, we compare the feature maps with standard reduction images and assess whether the network makes use of the distinguishing attenuation properties (Figure 3.8a) of the materials to identify the target cylinders. Along with the X-ray dataset, we look at the weights as a function of spectral bins, to assess the differences in learned weights by MSD and U-Net.

Figure 3.17 shows the reduction images relative to the ground truth, for all different reduction methods (including DRMSD and DRUNet) to one feature map channel. While the standard reduction methods do not yield very distinctive reductions, the DRMSD and DRUNet methods give a reduction that, although noisy, already gives a clear indication in black what the locations of the ground truth objects are. In Figure 3.18a, the output weights for the first layer in the DRMSD network are plotted as a function of the energy bins, which gives an indication of what the network learns during the training process. First of all, there is a clear peak at bins 60–63 and a valley at bins 64–67. The K-edge transition of the ground truth material silver is located between bins 63 and 64. Since the attenuation of silver changes between these bins, the network learns to take combinations of bins with energy slightly lower than that of the edge and bins with energy slightly higher than that of the edge, with positive and negative weights, respectively, to make the silver objects stand out. The other bins have decidedly lower weight magnitudes that revolve around zero, roughly cancelling out their contribution to the compressed image. It shows that the network can learn that



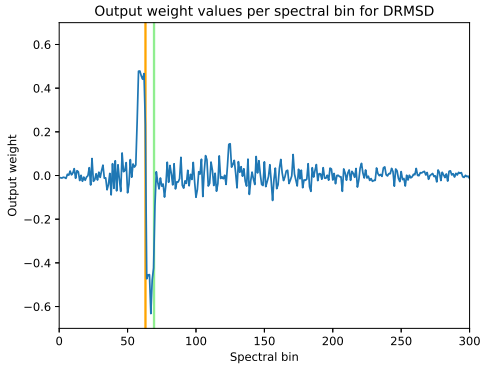
(**a**) Data at bin 65 (**b**) Ground truth (**c**) DRMSD compression

(**d**) DRUNet compression (**e**) PCA compression (**f**) LDA compression

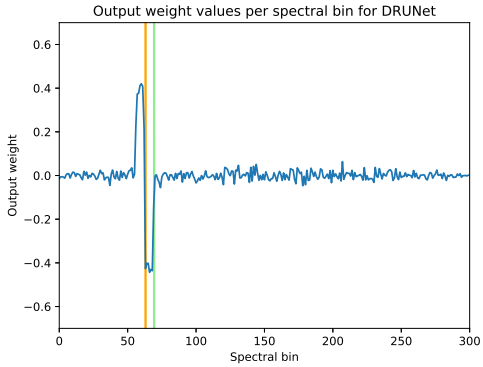
Figure 3.17: Visual comparison of the data reduction methods for reduction of the noisy many material dataset (**a**,**b**) to 1 image. Despite the high noise, DRMSD (**c**) and DRUNet (**d**) create the most distinctive images with respect to the ground truth (note the dark shapes at the target cylinder locations, indicated by red circles). The PCA (**e**) and LDA (**f**) compressions are included, but the Nonnegative Matrix Factorization (NMF) compression is omitted, as it is highly similar to the PCA compression.

only the aforementioned bins are critical to performing this segmentation task. Additionally, the weights become even smaller when bin 1 or 300 is approached, showing that the network learns to disregard very noisy bins. In Figure 3.18b the same quantities are plotted for a DRUNet trained network. The peaks and sharp transitions are visible in this case as well, but the absolute values of the non-peak weight values are generally lower than those of DRMSD.

For the noisy remote sensing dataset without overlap, Figure 3.19 shows an example image with its corresponding DRMSD and DRUNet compression and weight values in the data reduction layer when reducing to one feature map channel. While the shape of the graphs are different, there are some similarities. First of all, the weights of the bins 94–98 and 141–153, corresponding to the wavelengths from which noisy data arise, are zero. Thus, both networks learn to leave out noisy bins. Apart from this, despite the differences in network architecture, there are some similarities in the shapes of the graphs of the weights, mostly in the bin ranges 0–30 and 141–200.



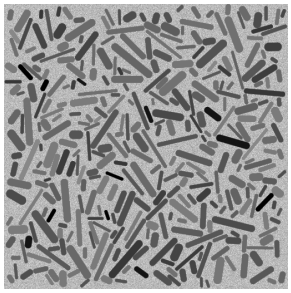
(**a**) Data reduction weights after training with DRMSD



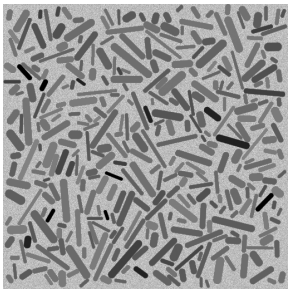
(**b**) Data reduction weights after training with DRUNet

Figure 3.18: Data reduction weights per bin after training with DRMSD (**a**) and DRUNet (**b**) on the noisy few-material X-ray dataset. The K-edge of the material of the objects to be detected (silver) is located between bins 63 and 64 (indicated in orange), which is the location of the drop. The K-edge of cadmium is indicated in green. Additionally, note that the absolute value of the weights decreases when approaching bin 1 or 300.

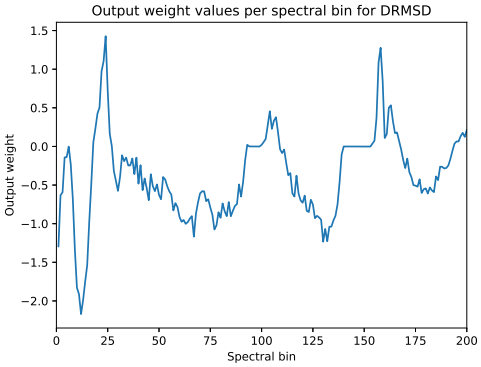
(**a**) Ground truth



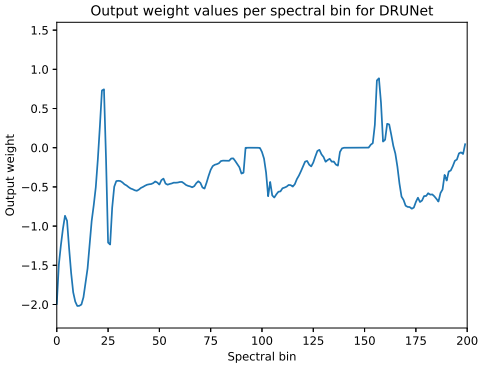
(**b**) DRMSD compression



(**c**) DRUNet compression



(**d**) Data reduction weights after training with DRMSD



(**e**) Data reduction weights after training with DRUNet

Figure 3.19: Example of data reduction weights (**d**,**e**) and resulting compressions (compare with ground truth (**a**)) for DRMSD (**b**) and DRUNet (**c**) with reduction to one feature map channel for the noisy remote dataset without overlap.

3.5 Discussion

We have applied the proposed Data Reduction CNN approach to two simulated datasets. We expected these datasets to be challenging since they contain noisy bins that have to be identified and left out. Furthermore, for the X-ray dataset one material has to be identified among 60 others, all of which may be superimposed on the same location in the projection images. For the remote sensing dataset, we expected that high spectral similarities between materials make identifying 10 different classes simultaneously a challenging task, especially when reducing to only one or two feature map channels. The Data Reduction CNN is able to obtain high accuracy with the proposed approach of simultaneous data-driven compression and training (Section 3.4.4). Note that we observe some slightly higher accuracies for DRCNN compared to the standard CNN approach in some cases and some slightly lower accuracies in other cases. We expect that this is because of statistical deviations due to the random nature of training CNNs. As can be seen in Section 3.4.4, the data reduction subnetwork learns to map spectral properties of the relevant materials to reduction feature maps, while the CNN simultaneously extracts the spatial properties from these feature maps for accurate segmentation. Whereas other common data reduction methods are expected to need considerably more feature map channels for a high accuracy, the DRCNN method is able to compress the data without any (hyper)parameter tuning to a very limited number of feature map channels (Section 3.4.4). The data reduction layer can, in principle, be successfully combined with any CNN, provided that this CNN without data reduction can also solve the imaging task.

Eventually, the compression method can be easily extracted from a trained network, such that the compression procedure and the classification task can be carried out at separate locations. In addition, the training procedure has to be carried out only once after which the task it has been trained for can be performed at high-speed throughput. Depending on the CNN architecture, the data reduction approach may speed up the training and application process as the reduction to few feature map channels may significantly decrease the number of trainable parameters in the network. An initial foray into the time reduction of the DRCNN approach is given in Appendix B.3, which compares the application of DRMSD and the MSD networks after training in GPUs and CPUs.

In future work, we plan to apply this approach to real datasets and in practical problem settings and assess whether the satisfactory accuracies and robustness results (Appendix B.4) carry over. The data reduction method shows accurate preliminary results on common benchmark datasets such as Pavia University, Indian Pines and Salinas. Therefore, it will be interesting what DRCNN can achieve on large and challenging experimental hyperspectral datasets.

3.6 Conclusions

In this work, we have proposed a task-driven end-to-end approach for supervised deep learning in hyperspectral imaging problems by adding a data reduction component to a convolutional neural network. The method is designed to work with any CNN and the combined Data Reduction CNN (DRCNN) network learns to effectively spectrally reduce the data for a given task, without the need for prior knowledge or network and parameter tuning. The data reduction subnetwork is directly connected with the chosen CNN and learns to combine data from hyperspectral bins, which is done simultaneously with training of the CNN for the imaging task. Using a simulated hyperspectral X-ray dataset and a simulated hyperspectral remote sensing dataset, we have demonstrated with a Multi-Scale Dense (MSD) and a U-Net network that a DRCNN can learn complex reductions from a typically large spectral dimension to a very limited number of feature map channels. As opposed to standard reduction methods such as PCA, NMF or LDA, this learned data reduction method finds essential distinctive task-specific features in the hyperspectral data while retaining high imaging task accuracies when compressing these features into a very limited number of feature map channels. We have shown that, despite noise and the presence of multiple overlapping material properties, high compression can be achieved with Data Reduction CNNs, resulting in significant advantages for high-compression and high-throughput applications.

Code and data availability

The source code to reproduce all results in this chapter are available on `https://github.com/mzeegers/DRCNN` [310]. These Python scripts make use of open-source toolboxes, in particular the ASTRA toolbox [1, 2], MSD-net [227] and PyTorch [222, 223]. The NIST attenuation spectra and (the links to) the USGS reflection spectra are available in the DRCNN code repository. The datasets generated for this chapter are also part of the DRCNN source code.

4

Multi-channel discrete tomography

In this chapter, we will look at discrete tomography where the number of materials in an object is limited. This can be exploited to make accurate reconstructions with a limited number of projections. The DART algorithm has proven to be a successful algorithm in this regard, but its usage is limited to objects with a few different materials. To push the number of materials that DART can handle, multi-channel data recorded with advanced tomographic imaging techniques – such as spectral X-ray tomography – can be used.

To effectively deal with multi-channel data, we present Multi-Channel DART (MC-DART). This class of algorithms is a generalization of DART to multiple channels and combines the information for each separate channel-reconstruction in a multi-channel segmentation step. We demonstrate in a range of simulation experiments with spectral X-ray imaging that MC-DART is capable of producing more accurate reconstructions compared to single-channel DART.

This chapter is based on:

M. T. Zeegers, F. Lucka, and K. J. Batenburg. "A Multi-Channel DART algorithm". In: *International Workshop on Combinatorial Image Analysis*. (Porto, Portugal). Ed. by R. P. Barneva, V. E. Brimkov, and J. M. R. S. Tavares. Springer, 2018, pp. 164–178.

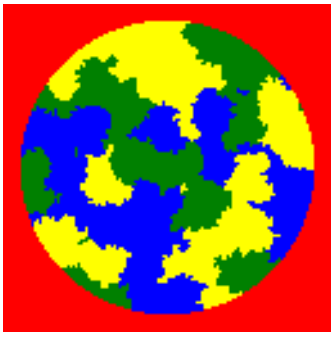
4.1 Introduction

Tomography is a non-invasive technique for creating 2D or 3D images of the inner structure of an object. Projections of the object are acquired by sending photonic or particle beams (e.g. X-rays, electrons, neutrons) through the object in a particular direction and measuring the signal resulting from interaction of the beam and the object at a detector. By acquiring this data from multiple positions and under various angles, a collection of projections is obtained. An image of the interior of the object is then reconstructed by applying a reconstruction algorithm to this projection data. Tomography is successfully used in many fields, including medical imaging [127] and electron tomography in materials science [93, 196]. If a large number of accurate projection images are available, solving the reconstruction problem is straightforward by a closed-form inversion formula [51]. Practical constraints on the dose, acquisition time or available space can impose limitations on the number of projections that can be taken, the angular range, or the noise level of the data, resulting in artefacts in the reconstructed images if standard reconstruction methods are used [127].

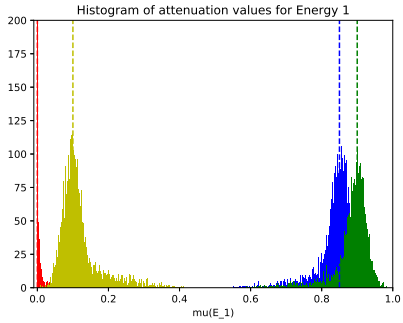
Discrete tomography is a powerful technique for dealing with such limited tomographic data. It can be applied if the object consists of only a limited number of materials with homogeneous densities. The Discrete Algebraic Reconstruction Technique (DART) [30, 31] is an algebraic reconstruction method for discrete tomography that alternates between continuous reconstruction steps and discretization of the image intensities by segmentation. The DART algorithm has demonstrated to obtain higher image quality reconstructions with limited projections and angles compared to standard reconstruction methods. Numerous successive studies have improved the DART algorithm, which include automatic parameter estimation (PDM-DART [3] and TVR-DART [328]), multi-resolution reconstruction (MDART [65]), relaxing voxel constraints (SDART [42]) and adaptive boundary reconstructions (ADART [181]). Nevertheless, a key limitation of DART is that it can only improve reconstruction quality if the number of different materials in the object is relatively small. The main reason is that for a larger number of materials, the segmentation step is no longer effective [30, 31].

In some cases it is possible to obtain tomographic information in *multiple measurement channels*. For instance, in X-ray imaging the beams are typically polychromatic, i.e. X-ray photon energies are distributed over a spectrum. Each material in the object has different attenuation properties for different X-ray energy levels. Whenever a single X-ray energy value is desired the range of energies within the beam can be narrowed by applying filters at the X-ray source [51]. Some detectors are capable of separating the incoming photons into energy bins while counting (e.g. HEXITEC [295]). In these cases spectral multi-channel projection data are acquired, providing additional information about the object at different energies. Compared to the single-channel setting, where each material has a single

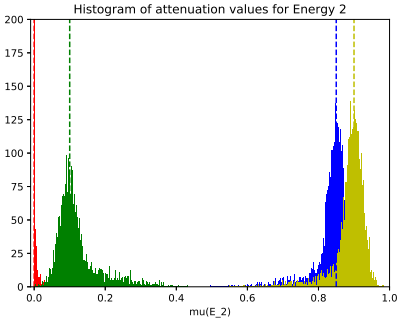
attenuation value in the reconstructed image, in the multi-channel setting the attenuation value for each material varies along the channels. In this way, a tomographic dataset of the object is acquired for each channel, where the attenuation values of the materials change throughout these datasets. This multi-channel imaging can potentially yield extra information about the materials. With more materials in the object, especially with similar attenuation features at a fixed energy, having data from multiple channels enables a better separation during segmentation. A conceptual example of this is shown in Figure 4.1. It is hard to separate points based on their attenuation values in a one-dimensional energy space. For instance, the right side of the blue area might as well be assigned to the green or yellow material during segmentation. With two energy dimensions the points are easily separable, since each voxel value lies close to its attenuation cloud center. Note



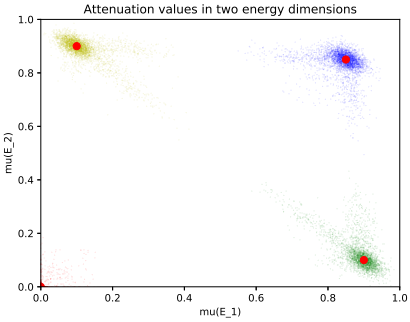
(**a**) Material distribution



(**b**) One-dimensional attenuations for energy channel E_1



(**c**) One-dimensional attenuations for energy channel E_2



(**d**) Two-dimensional attenuations

Figure 4.1: Elementary example of separation difficulties during segmentation. (a) Distribution of the three materials (blue, yellow, green) in the object. The background is indicated in red. (b) Histogram of attenuation values of pixels at energy E_1 (above) and E_2 (below). Vertical lines show true material attenuations. (c) Attenuations of the materials (red dots) and computed attenuations by a reconstruction algorithm for each voxel (colors indicate the materials these belong to).

that these spectra are artificial and not likely to occur in real-world examples.

In this chapter we present a new class of algorithms that combines DART with multi-channel imaging for solving discrete multi-channel reconstruction problems. Our method can combine the information from multiple channels to produce a final segmentation that is superior to that of the (single-channel) DART algorithm. Note that since this new method is designed by means of modules or subroutines that are interchangeable (as with DART), the method is essentially a class of algorithms providing a framework for dealing with multi-channel data. For simplicity, however, we will frequently call this framework an algorithm.

This chapter is structured as follows. Section 2 introduces the multi-channel discrete tomography problem. In Section 3 the DART algorithm is restated and the *Multi-Channel DART* (MC-DART) algorithm is introduced. Results of experiments with this algorithm are reported in Section 4. Finally, Section 5 presents the conclusions of this study.

4.2 Problem formulation

The standard (single-channel) tomography problem can be modelled as a system of linear equations. The image is characterized by a vector of voxel attenuation values $\boldsymbol{x} \in \mathbb{R}^n$, where n is the number of voxels. We will work with 2D images, but the problem formulation and methods in this work can easily be extended to the 3D setting. We will refer to the image pixels as *voxels* to distinguish these from detector pixels. We will often interchangeably speak of voxels and their corresponding indices. The projection values (also called data) are given as the vector $\boldsymbol{p} \in \mathbb{R}^l$, where l is the number of projection angles times the number of detector pixels. The reconstruction problem can then be described by solving the following set of linear equations for $\boldsymbol{x}$:

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}. \tag{4.1}$$

Here $\boldsymbol{W}$ is the projection matrix, also called the forward operator [140]. This matrix incorporates the contribution of each voxel to each projection, where element w_{ij} indicates the contribution of voxel j to projection i. Applying the operator $\boldsymbol{W}$ on a vector $\boldsymbol{x}$ results in the forward projection (the sinogram). Since inverting the matrix $\boldsymbol{W}$ is computationally too expensive (or not even possible, for example when the problem is ill-posed) the reconstruction problem is to find a solution $\boldsymbol{x}^*$ whose forward projection $\boldsymbol{W}\boldsymbol{x}^*$ matches the projection data best with respect to some norm $||\cdot||$.

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}|| \tag{4.2}$$

Since this is a least squares problem over $\mathbb{R}^n$, a solution always exists. For simplicity of notation we also assume that it is unique. A vector that encapsulates noise

from real-world examples can also be modelled with (4.2). In our experiments with phantom examples in Section 4.4 there is no noise.

In the *discrete tomography problem*, the image to be reconstructed consists of a limited number of materials with homogeneous densities, each having an attenuation which is known beforehand by means of the set $\mathcal{R} = \{\rho_1, \ldots, \rho_m\}$, where m is the number of different materials in the object. Therefore the problem to be solved becomes finding a vector $\boldsymbol{x} \in \mathcal{R}^n$ that matches the data best:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathcal{R}^n} ||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||. \tag{4.3}$$

Note that this is a minimization problem over a non-empty finite set. Hence, a minimum always exists. Again, the minimizer does not need to be unique but we use this notation throughout the chapter for simplicity.

In the *multi-channel* setting different properties of the target can be individually interrogated and measured. The information of each property is obtained through a separate channel. An example of a channel is an energy level, as in the example in Section 4.1. In Figures 4.1b and 4.1c, the channels are the two energy levels revealing attenuations of the object at different energies. In a more abstract way the object is described as set a of voxels with labels instead of attenuation values, since each material has different attenuation values in different channels. The material labels are values in the set $\mathcal{M} = \{1, 2, \ldots, m\}$. The channel indices are given by $\mathcal{E} = \{E_1, E_2, \ldots, E_C\}$ where the number of channels is given by C. Again, the attenuations are known beforehand in the sets $\mathcal{R}_{E_1} = \{\rho_{1,1}, \ldots, \rho_{1,m}\}, \mathcal{R}_{E_2} = \{\rho_{2,1}, \ldots, \rho_{2,m}\}, \ldots, \mathcal{R}_{E_C} = \{\rho_{C,1}, \ldots, \rho_{C,m}\}$. In this setting, let $\boldsymbol{\mathcal{R}} = \cup \mathcal{R}_{E_c}$. The function $\mu : \mathcal{M} \times \mathcal{E} \mapsto \boldsymbol{\mathcal{R}}$ maps the label-channel combinations to their attenuation value, so the attenuation of a material with label s at channel E_c is given by $\mu(s, E_c)$. Note that there is not necessarily a one-to-one correspondence between the attenuation values and the material-channel combinations, because some combinations can have the same attenuation value. In this multi-channel case the projection data are given by a vector of projection data vectors at various channels:

$$\boldsymbol{P} = (\boldsymbol{p}_{E_1}, \ldots, \boldsymbol{p}_{E_C}) \in \mathbb{R}^{n \times C}. \tag{4.4}$$

For each channel E_c the reconstruction problem for $\boldsymbol{x}_{E_c}$ is given by the following set of linear equations:

$$\boldsymbol{W}\boldsymbol{x}_{E_c} = \boldsymbol{p}_{E_c}, \qquad E_c \in \{E_1, \ldots, E_C\}. \tag{4.5}$$

For $\boldsymbol{y} \in \mathcal{M}^n$, define $\mu(\boldsymbol{y}, E_c) = (\mu(\boldsymbol{y}_1, E_c), \ldots, \mu(\boldsymbol{y}_n, E_c))^\top$ as the vector of voxel attenuation values at channel E_c. The multi-channel problem is now defined as follows. Given data vector $\boldsymbol{P}$ and projection matrix $\boldsymbol{W}$, find a labelling vector $\boldsymbol{y}^* \in \mathcal{M}^n$ such that for each channel E_c the difference between forward projection $\boldsymbol{W}\mu(\boldsymbol{y}^*, E_c)$ and data is minimal with respect to some norm $||\cdot||$:

$$\boldsymbol{y}^* = \underset{\boldsymbol{y}\in\mathcal{M}^n}{\arg\min} \sum_{h=1}^{C} ||\boldsymbol{W}\mu(\boldsymbol{y}, E_c) - \boldsymbol{p}_{E_c}||. \tag{4.6}$$

Note that for one channel the minimization problem is equivalent to (4.3) where the labelling is given by the attenuation values $\boldsymbol{x}$, by setting $\mu(\boldsymbol{y}, E_1) = \boldsymbol{x}$ and $\mathcal{M}^n = \mathcal{R}^n$ and $\boldsymbol{p}_{E_1} = \boldsymbol{p}$:

$$\boldsymbol{y}^* = \underset{\boldsymbol{y}\in\mathcal{M}^n}{\arg\min} ||\boldsymbol{W}\mu(\boldsymbol{y}, E_1) - \boldsymbol{p}_{E_1}|| \tag{4.7}$$

$$= \underset{\boldsymbol{x}\in\mathcal{R}^n}{\arg\min} ||\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}||. \tag{4.8}$$

4.3 Algorithms

In this section, the Multi-Channel DART (MC-DART) framework for solving the minimization problem of Eq. (4.6) is introduced. We first explain the DART algorithm as given in [30] by discussing the overall structure and its building blocks. We then describe each building block of the MC-DART algorithm separately in more detail. Note that ASTRA [1, 215] provides an implementation for numerically computing all projection matrices in these algorithms, either by storing the full matrix or doing all necessary computations in a matrix-free way.

4.3.1 DART

The DART algorithm attempts to solve the optimization problem of Eq. (4.3) by iteratively alternating between continuous reconstruction steps and discrete segmentation steps. The number of materials in the object to be reconstructed and their attenuation values should be known beforehand, given by the function μ. The algorithm consists of several phases, which are indicated in the flow-chart in Figure 4.2. The pseudocode of DART is given in Algorithm 1.

Algorithm 1 DART

Input: $\boldsymbol{W}$, $\boldsymbol{p}$, $\mathcal{R}$
$\boldsymbol{x}^0 \leftarrow$ **Mask-ARM**$(\boldsymbol{W}, \boldsymbol{p}, \mathbf{1_n}, \mathbf{0_n})$
for $k = 1$ **to** K **do**
  $\boldsymbol{y}^k \leftarrow$ **Seg**$(\boldsymbol{x}^k, \mathcal{R})$
  $\boldsymbol{M}^k \leftarrow$ **Mask**$(\boldsymbol{y}^p)$
  $\boldsymbol{x}^k \leftarrow$ **Mask-ARM**$(\boldsymbol{W}, \boldsymbol{p}, \boldsymbol{M}^k, \boldsymbol{x}^{k-1})$
Output: $\boldsymbol{x}^K$, **Seg**$(\boldsymbol{x}^K, \mathcal{R})$

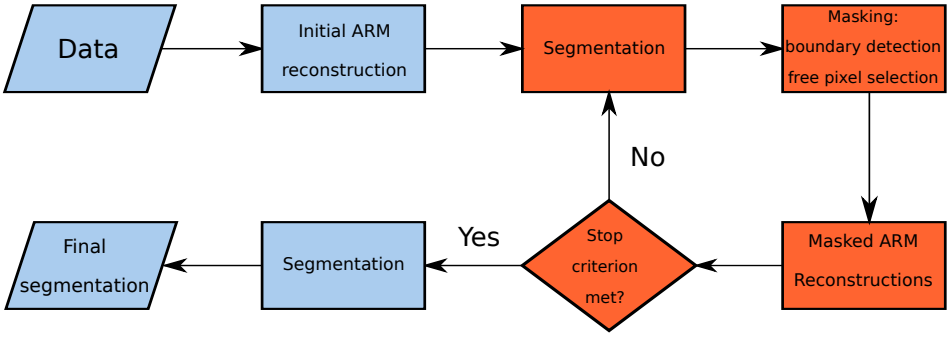


Figure 4.2: Flow chart of the DART algorithm. The DART iteration activities are indicated in red and the initialization and post-segmentation activities are indicated in blue.

Initialization

In the initialization phase, given the projection data $\boldsymbol{p}$ and the projection properties by means of $\boldsymbol{W}$, an initial reconstruction $\boldsymbol{x}^0$ is calculated using an Algebraic Reconstruction Method of choice (hereafter referred to as the ARM), for example ART, SART or SIRT [140]. With the initial reconstruction $\boldsymbol{x}^0$ at hand, the main loop of the DART algorithm begins.

Segmentation

In this main loop, in iteration k the image $\boldsymbol{x}^{k-1}$ is segmented using a simple thresholding scheme, forming the image $\boldsymbol{y}^k \in \mathcal{R}^n$, by computing for every voxel j the closest material attenuation value:

$$\boldsymbol{y}_j^k = \begin{cases} \rho_1, & \boldsymbol{x}_j^{k-1} < \frac{1}{2}(\rho_1 + \rho_2) \\ \rho_2, & \frac{1}{2}(\rho_1 + \rho_2) \leq \boldsymbol{x}_j^{k-1} < \frac{1}{2}(\rho_2 + \rho_3) \\ \quad\vdots & \\ \rho_m, & \frac{1}{2}(\rho_{m-1} + \rho_m) \leq \boldsymbol{x}_j^{k-1} \end{cases} \tag{4.9}$$

$$= \arg\min_{\rho \in \mathcal{R}} ||\boldsymbol{x}_j^{k-1} - \rho||_2. \tag{4.10}$$

The second expression is easier to generalize to a higher-dimensional setting, which will be done in Section 4.3.2.

Boundary detection and masking

A set of voxels in the segmentation is selected for a new reconstruction to refine the resulting image. First, the set $B^k \subset \{1, \ldots, n\}$ of boundary voxel indices is determined based on the segmentation. Various schemes can be applied for boundary detection. Additionally, a set $U^k \subset \{1, \ldots, n\}$ of free voxel indices is

determined, where each voxel is included with a certain probability $1-\beta$, with $0 \leq \beta \leq 1$. The process of selecting the voxels $U^k \cup B^k$ to be reconstructed and the voxels to be left out is called masking. Note that in the initialization phase all voxels are included in the mask.

Masked ARM reconstructions

The set of free voxel indices $U^k \cup B^k$ are subjected to a new ARM reconstruction. This is done by computing the forward projection of the voxels $(\boldsymbol{y}_j^k)$ with $j \notin U^k \cup B^k$, and subtracting this from the input data $\boldsymbol{p}$ to obtain the residual sinogram $\overline{\boldsymbol{p}}^k$. The subproblem that has to be solved in this phase is:

$$\overline{\boldsymbol{W}}^k \overline{\boldsymbol{x}}^k = \overline{\boldsymbol{p}}^k. \tag{4.11}$$

In Eq. (4.11) matrix $\overline{\boldsymbol{W}}^k$ is defined by $\overline{\boldsymbol{W}}^k = (w_{ij})_{j \in U^k \cup B^k}$ and vector $\overline{\boldsymbol{x}}^k$ to be found has length $|U^k \cup B^k|$. Thus, the system of equations contains the same number of equations as Eq. (4.2) but has fewer unknowns. The system is solved using a fixed number of ARM iterations, taking the values of $(\boldsymbol{x}_j^{k-1})_{j \in U^k \cup B^k}$ as the starting condition. The complete reconstruction $\boldsymbol{x}^k$ at the end of iteration k is then formed by merging $\overline{\boldsymbol{x}}^k$ with $\boldsymbol{y}^k$.

Some DART implementations also include a smoothing step at this point. The entire loop is repeated a predefined number of times. After the loop ends, the image is segmented one more time. An example of the DART algorithm on a two-material phantom is given in Figure 4.3. Note that the DART algorithm has many degrees of freedom. This includes the number of ARM iterations in the initialization phase, the number of DART iterations, the number of ARM iterations during these DART iterations, the fixing probability β, and possibly parameters in the smoothing operation. The quality of the reconstructions also depends on the tomographic setup, such as the number of projections and the number of projection angles, and on the complexity of the object, including the number of materials and different attenuation values. Despite the DART algorithm performing well in practice, it is a heuristic method for which no solution guarantees exist [28]. The DART algorithm is also highly modular. Approaches for segmentation, boundary detection, reconstruction (ARM) and possible smoothing can easily be changed without sacrificing the overall structure of the algorithm. For the multi-channel algorithm proposed in this section, the segmentation phase is adapted to using all multi-channel reconstructions as input.

The complexity of the framework depends on the algorithms that are used for reconstruction and segmentation. In this work, we use SIRT as the reconstruction algorithm and the thresholding segmentation as described above. Therefore, in this case, the DART algorithm has a time complexity of $O(Kn(m+l))$. The space complexity of our implementation is $O(ln)$.

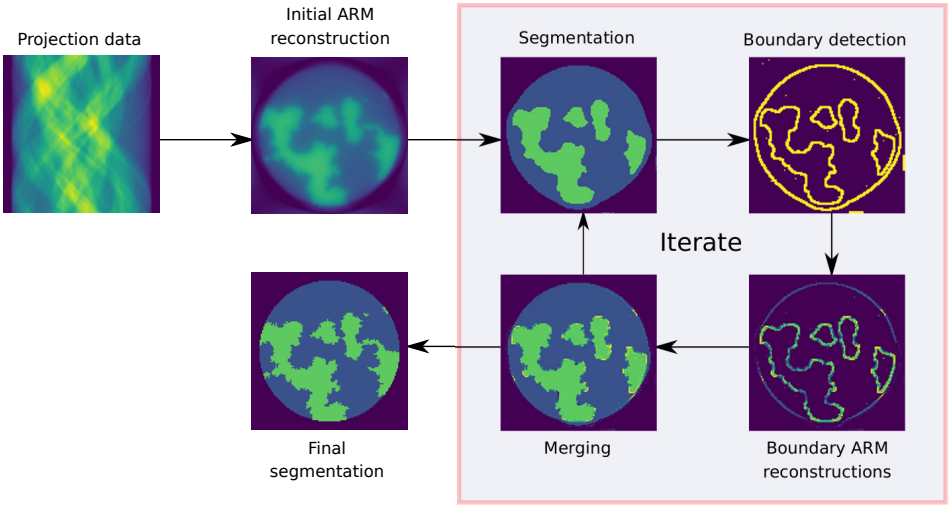


Figure 4.3: An example of DART applied to simulated X-ray data from a two-material phantom.

4.3.2 Multi-channel DART

We now present the *Multi-Channel DART* (MC-DART) algorithm and outline its separate building blocks. Most focus will be on the multi-channel segmentation. Note that labelling single-channel images separately by attenuation values does not work here, since across multiple channels different materials can have the same attenuation. Therefore, there are some slight changes in the other blocks as well due to a new labelling mechanism. The algorithm structure is shown in the flow-chart in Figure 4.4. The pseudocode of MC-DART is given in Algorithm 2.

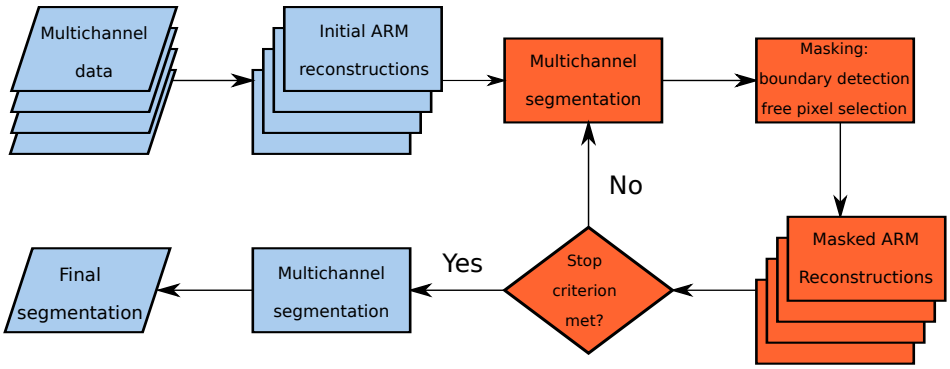


Figure 4.4: Flow chart of the MC-DART algorithm. A stacked number of activities indicate that these are applied at different channels simultaneously.

Algorithm 2 MC-DART

Input: $\boldsymbol{W}$, $\mathcal{E}$, $\boldsymbol{P}$, $\boldsymbol{\mathcal{R}}$, $\mathcal{M}$, μ
for $c = 1$ **to** C **do**
  $\boldsymbol{x}^0_{E_c} \leftarrow$ **Mask-ARM**$(\boldsymbol{W}, \boldsymbol{p}_{E_c}, \mathbf{1_n}, \mathbf{0_n})$
for $k = 1$ **to** K **do**
  $\boldsymbol{y}^k \leftarrow$ **MCSeg**$(\boldsymbol{X}^k, \boldsymbol{\mathcal{R}}, \mathcal{M}, \mu)$
  $\boldsymbol{M}^k \leftarrow$ **Mask**$(\boldsymbol{y}^k)$
  for $c = 1$ **to** C **do**
    $\boldsymbol{x}^k_{E_c} \leftarrow$ **Mask-ARM**$(\boldsymbol{W}, \boldsymbol{p}_{E_c}, \boldsymbol{M}^k, \boldsymbol{x}^{k-1}_{E_c})$
Output: $\boldsymbol{X}^K$, **MCSeg**$(\boldsymbol{X}^K, \boldsymbol{\mathcal{R}}, \mathcal{M}, \mu)$

Initialization

In the multi-channel setting we start out with a vector of projection data $\boldsymbol{P}$ at various channels and the matrix $\boldsymbol{W}$ as before. For each channel E_c a reconstruction $\boldsymbol{x}^0_{E_c}$ is computed using the selected ARM. This results in C initial reconstructions for the MC-DART loop.

Multi-Channel segmentation

Given the reconstructions for all channels, similar to the DART segmentation, the multi-channel segmentation will determine a label image $\boldsymbol{y}^k \in \mathcal{M}^n$. Let $\boldsymbol{\mu}(s) = (\mu(s, E_1), \ldots, \mu(s, E_C)) \in \mathcal{R}_{E_1} \times \mathcal{R}_{E_2} \times \ldots \times \mathcal{R}_{E_C}$ be the vector of all attenuation values at each energy for material $s \in \mathcal{M}$, and let $\boldsymbol{X}^k(\cdot, j) = (\boldsymbol{x}^k_{j,E_1}, \ldots, \boldsymbol{x}^k_{j,E_C})$ be the vector of all attenuation values of voxel j at each channel. We compute the segmented image by computing for each voxel $j \in \{1, \ldots, n\}$ the label using a basic thresholding scheme:

$$\boldsymbol{y}^k_j = \arg\min_{s \in \mathcal{M}} ||\boldsymbol{X}^k(\cdot, j) - \boldsymbol{\mu}(s)||_2. \tag{4.12}$$

Essentially, this operation selects the material label for which the multi-dimensional difference between the material attenuation and voxel attenuations is smallest.

Masking and boundary detection

The masking works exactly the same as in the single-channel case. Given the segmentation $\boldsymbol{y}^k$ the masking produces a set $U^k \cup B^k$ of voxel indices to be included in the multi-channel reconstructions.

Multi-Channel reconstructions

In the MC-DART algorithm the reconstructions are handled separately for each energy. Thus, in MC-DART iteration k the ARM is invoked C times to find $\overline{\boldsymbol{x}}^k_{E_c}$

for each channel c in

$$\overline{\boldsymbol{W}}^k \overline{\boldsymbol{x}}^k_{E_c} = \overline{\boldsymbol{p}}^k_{E_c}. \tag{4.13}$$

The resulting (merged) reconstructions are then given by $\boldsymbol{X}^k := (\boldsymbol{x}^k_{E_1}, \ldots, \boldsymbol{x}^k_{E_C}) \in \mathbb{R}^{n \times C}$.

An example of the MC-DART algorithm on a two-material phantom with two channels is given in Figure 4.5. As with DART, the complexity of this framework depends on the reconstruction and segmentation methods that are chosen, as well as the extent of parallelization. If we use SIRT and the multi-channel segmentation method as described above and use a completely sequential implementation, the time complexity of MC-DART is $O(CKn(l+m))$. Because of the dependencies on the methods, we rather speak of a *relative complexity* of MC-DART to DART, which we define as the ratio of the sequential MC-DART complexity to that of DART, irrespective of the subroutines used. This relative time complexity is $O(C)$. The space complexity of this algorithm instance of MC-DART is $O(Cn)$, resulting in a relative space complexity of $O(C)$ as well.

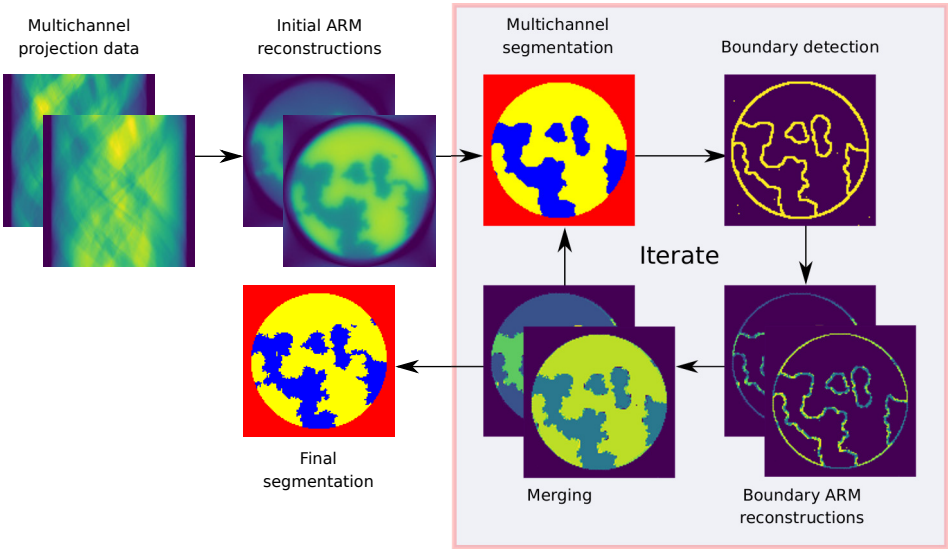


Figure 4.5: An example of MC-DART applied to simulated two-channel X-ray data from a two-material phantom.

4.4 Experiments and results

In this section the performance of the described MC-DART framework in terms of reconstruction and segmentation is presented. A series of experiments have been designed in which the number of channels C and different materials m are varied. For each experiment, multiple random phantoms are created. The size of these two-dimensional phantoms is 128×128 pixels, and each phantom consists of a circular disk containing a random parcellation among m materials in such a way that the total surface is approximately equal for each material. An example of this random phantom is given in Figure 4.1a, where $m = 3$. Given the number of materials and channels, random attenuation spectra are generated by assigning a random number $\mu(s, E_c) \sim \mathcal{U}(0, 1)$ for each channel-material combination, where $s \in \{1, \ldots, m\}$ and $E_c \in \{E_1, \ldots, E_C\}$. With this way of generating spectra no dependencies between channels are established. Note that in most practical applications such dependencies do exist, as materials all have their own attenuation spectrum. For each phantom, reconstructions are made. The reference values for the tomographic setup and the parameter values of the MC-DART reconstruction algorithm for these reconstructions are summarized in Table 4.1. For multi-channel segmentation the method as described in Section 4.3.2 is used.

Parameter	Reference value
Angles	32 (equidistant)
ARM	SIRT
Start iterations	10
MC-DART iterations K	10
ARM iterations	10
Fix probability β	0.99

Table 4.1: Reference values for the parameters of the tomographic setup and the reconstructions algorithm for all experiments.

We vary the number of channels $C \in \{1, \ldots, 10\}$ and materials $m \in \{2, \ldots, 10\}$ independently. For each combination, a random phantom $\boldsymbol{y}_{\text{init}}$ is created, after which data $\boldsymbol{P}$ is generated by applying the forward projection as described in Section 4.2 on the phantom by applying μ and $\boldsymbol{W}$ on $\boldsymbol{y}_{\text{init}}$. In all experiments parallel-beam geometries are used and the detector size is 128 pixels. After this, the MC-DART algorithm as described in Section 4.3.2 is applied with $K = 10$ MC-DART iterations. The final segmentation is compared to the original phantom and the pixel error is computed, which is defined as the number of pixels in the final segmentation $\boldsymbol{y}^K$ that are labelled differently compared to the corresponding pixels in the original phantom $\boldsymbol{y}_{\text{init}}$. Only the pixels in the inner disk of the phantoms are taken into account. All experiments are repeated for and averaged over 100 runs with different phantoms.

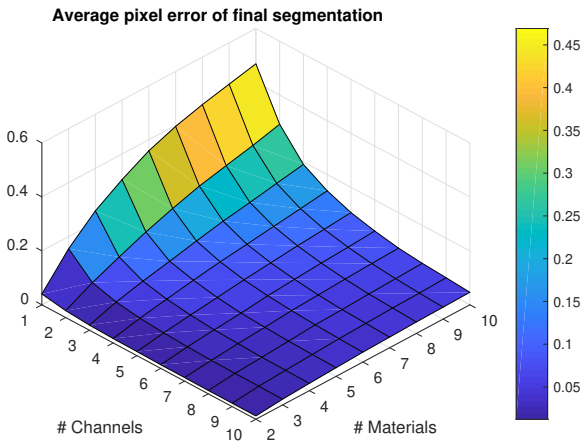
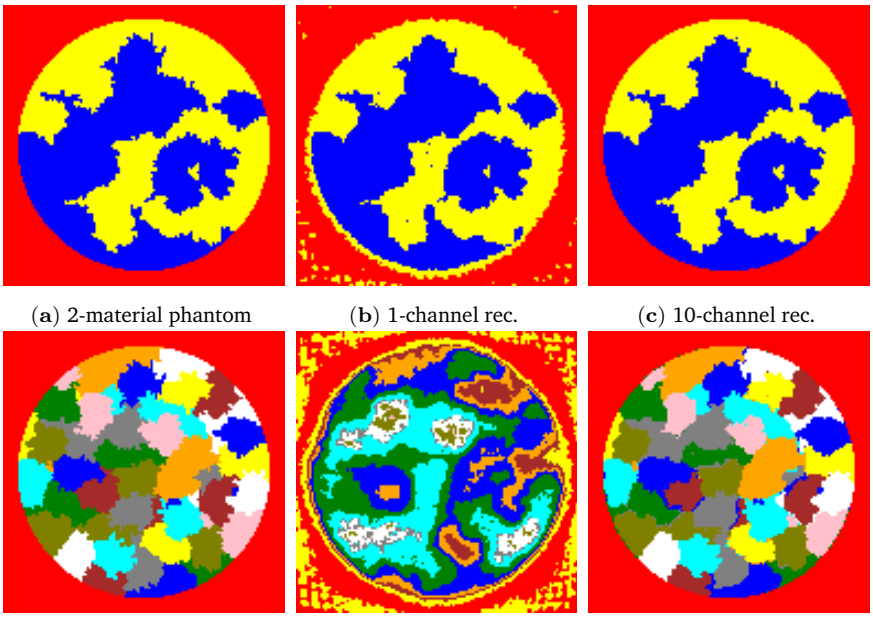


Figure 4.6: Pixel error percentage for different number of material-channel combinations.

The creation of random phantoms is implemented in MATLAB. The remainder of the experiment setup scripts are implemented in Python. The reconstruction algorithms, including the MC-DART algorithm, are implemented in Python, where the ASTRA Toolbox [1, 2] is used to take care of the ARM invocations and forward projections, including the masking in each MC-DART iteration and the creation of matrices $\boldsymbol{W}$ and $\overline{\boldsymbol{W}}^k$ based on the geometric properties.

Figure 4.6 shows the percentage of misclassified pixels with respect to the number of pixels in the inner disk. The percentage is lowest when the number of materials is low and the number of channels is high, while the percentage is highest when the number of channels is low and the number of materials is high. Given a number of channels, the percentage seems to scale logarithmically with the number of materials. On the other hand, given a number of materials, the percentage seems to scale exponentially with the number of channels for larger number of materials. Therefore, in this setup, the addition of only a few channels improves the reconstruction quality considerably. Figure 4.7 shows examples of the reconstructions at the corners of the curved plane of Figure 4.6.

We have investigated the effect of changing the parameters that are shown in Table 4.1. The number of starting iterations has no effect on the pixel error percentage curve. For these parameters, we found that increasing the number of MC-DART iterations further than 4 had no significant effect on the reconstructions. This threshold depends on the number of ARM iterations in each MC-DART iteration. Also, the quality of the reconstructions increases only marginally when β is increased. However, the pixel error percentage drops considerably as the number of ARM iterations during an MC-DART iteration increases. Also, when scanning



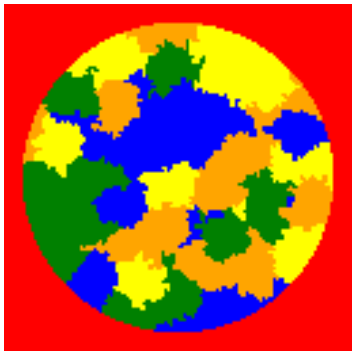
(**a**) 2-material phantom (**b**) 1-channel rec. (**c**) 10-channel rec.

(**d**) 10-material phantom (**e**) 1-channel rec. (**f**) 10-channel rec.

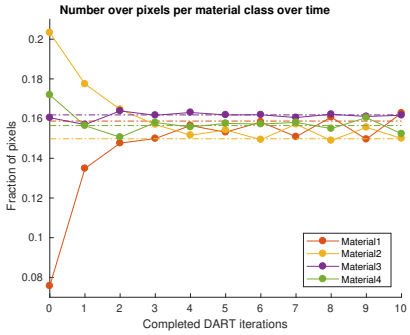
Figure 4.7: Reconstructions for various setups. (a, d) Phantoms used with two and ten materials respectively. (b, e) Reconstructions using one channel. The mislabelled yellow pixels are because the attenuation of the yellow material is very close to zero. (c, f) Reconstructions using ten channels.

data from many angles are available, the reconstruction quality improvements with multiple materials become much better. For only 2 angles, the reconstruction between $C = 1$ and $C = 10$ channels improves from pixel error percentage 27% to 23% for two materials and from 55% to 41% for ten materials. In comparison, for as much as 128 angles the reconstructions between $C = 1$ and $C = 10$ channels improve by from 3% to less than 1% for two materials and from 46% to 4% for ten materials. We conclude that in all these cases the MC-DART algorithm gives better results when more channels are available.

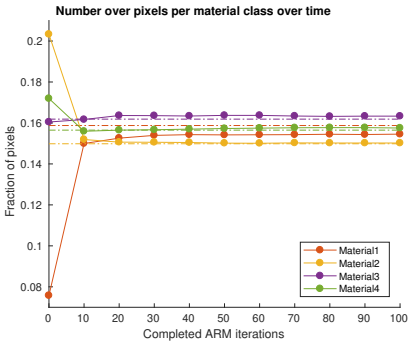
Additionally, apart from the pixel error, we investigate how the number of assigned pixels per material class behaves as the MC-DART reconstruction proceeds. The results are shown in Figure 4.8. A random phantom with four different materials and background is used. The number of channels is set to $C = 10$, and for each channel c and material m a random attenuation value $\mu(s, E_c) \sim \mathcal{U}(0, 1)$ is generated. Then the MC-DART algorithm is applied to this phantom in two different experimental setups. In the first experiment, the number of MC-DART iterations is set to 10 and the number of ARM iterations per MC-DART iteration is set to 10. After each MC-DART iteration, the number of pixels assigned are calculated for each class. During the first four MC-DART iterations the number



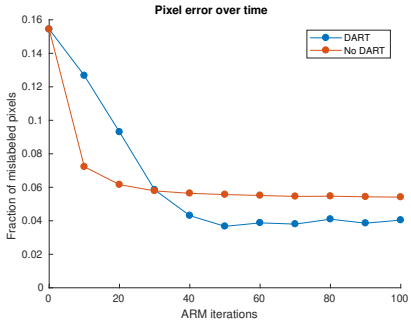
(**a**) Phantom



(**b**) Number of pixels per class for DART



(**c**) Number of pixels per class for non-DART



(**d**) Pixel error over time

Figure 4.8: Convergence behavior for material classes with $C = 10$ channels. (a) Phantom that has been used with $m = 4$ materials (b) Graph showing the behavior for each material class in a DART routine for this phantom. The number of iterations for the initial reconstruction is set to 2, the number of DART iterations is 10, the number of ARM iterations is set to 10, the number of angles is 8, fixing probability is set to $\beta = 0.99$. The chosen ARM is SIRT. Shown are the number of pixels assigned per class during segmentation after each DART step, with the true value of these indicated by dashed lines. (c) Number of pixel assigned per class over number of ARM iterations. The number of DART iterations is 0, and instead we apply 100 ARM iterations with 2 initial iterations. The results are based on intermediate segmentations after each 10 ARM iterations, but these segmentations are not used in further iterations. Background pixels are excluded from the results (d) Pixel error over number of ARM iterations for both approaches.

of assigned pixels is converging towards their real values. After this, the graphs enter an oscillatory phase in which for each class the number of assigned pixels alternates between two values whose average is not necessarily the real number of pixels for that class. For comparison, in the second experiment the same setup is used, but without using MC-DART iterations and applying the same ARM for 100 iterations instead. In this way the ARM is effectively invoked equally often. After each 10 iterations a segmentation is made based on the current reconstruction and the pixels per class are measured, but no new forward projections are calculated from these segmentations and used in subsequent ARM iterations. In this case the number of pixels converges much more quickly for each class. Also, there is no oscillatory phase and the number of pixels are just as close to their true values as with the DART approach. However, plotting the total pixel error over time reveals that the pixel error in the non-DART case is higher. The pixel error for the MC-DART case needs more time to stabilize to its oscillatory phase, but the values are eventually lower than in the non-MC-DART case.

4.5 Discussion

This chapter presents the first steps to implement a multi-channel reconstruction technique using multi-channel segmentation. Currently, there are no standard approaches for the discrete multi-channel problem presented in Section 4.2. We propose a framework in which reconstruction and segmentation techniques can be exchanged. The modules in the framework can be adjusted to the problem to be solved. For instance, segmentation can be performed with neural network based methods. The proposed method is not aimed at optimizing reconstructions with state-of-the-art ARMs or segmentation techniques but at presenting a framework to work with multi-channel data. If more data from different channels are available, this implementation outperforms DART but it does not mean that the problem is optimally solved. To further develop this technique and transfer it to real-world settings, real-data properties should be taken into account. These properties include the correlation of attenuation values between channels and noise contained in the projection data. In our study we only make use of the multi-channel data during segmentation. Another approach could be to use the multi-channel data during reconstruction, modelling the reconstruction problem as a large inverse problem where the unknowns are the material concentrations in each pixel (e.g. see [143, 275]). However, solving this problem is much more involved and the MC-DART framework presented in this chapter provides a simple but effective alternative of separating materials using multi-channel data.

4.6 Conclusions

A new class of algorithms for solving discrete multi-channel reconstruction problems has been proposed. This framework uses the strength of DART regarding dealing with limited data in a multi-channel setting by using a multi-channel segmentation method. The experiments have shown that combining information from different channels by a multi-channel segmentation method increases the reconstruction quality compared to the single-channel DART algorithm. Therefore, we conclude that the MC-DART framework is a promising approach for dealing with multi-channel data.

Code and data availability

The source code to reproduce all results in this chapter are available on `https://github.com/mzeegers/DRCNN` [311]. These Python scripts make use of the ASTRA toolbox [1, 2]. The data configurations generated for this chapter are available in the MC-DART source code repository.

5

Dictionary-based spectral tomography

Spectral X-ray imaging has the potential to significantly improve the accuracy of material decomposition in tomography. Nevertheless, the problem of volumetric material decomposition in spectral tomography is often highly ill-posed. Therefore, most state-of-the-art methods rely on strict prior information, and are able to provide material decomposition with a limited number of materials, with limited accuracy, or a limited set of spectral configurations.

To resolve these issues, we propose 'A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography' (ADJUST). Its formulation relies on forming a dictionary of spectral signatures of materials common in CT and prior knowledge of the number of materials present in an object. In particular, we decompose the spectral volume linearly in terms of spatial material maps, a spectral dictionary, and the indicator of materials for the dictionary elements. We propose a memory-efficient accelerated alternating proximal gradient method to find an approximate solution to the resulting bi-convex problem. From numerical demonstrations on several synthetic phantoms, we observe that ADJUST performs exceedingly well compared to other state-of-the-art methods. Additionally, we address the robustness of ADJUST against limited and noisy measurement patterns. The demonstration of the proposed approach on a spectral micro-CT dataset shows its potential for real-world applications.

This chapter is based on:

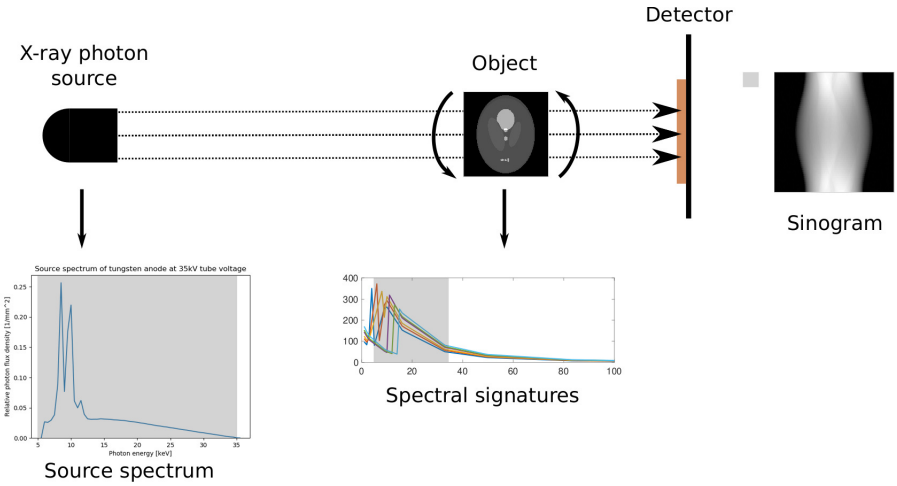
M. T. Zeegers, A. Kadu, T. van Leeuwen, and K. J. Batenburg. "ADJUST: A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography". *Inverse Problems* 38.12 (2022), p. 125002.

5.1 Introduction

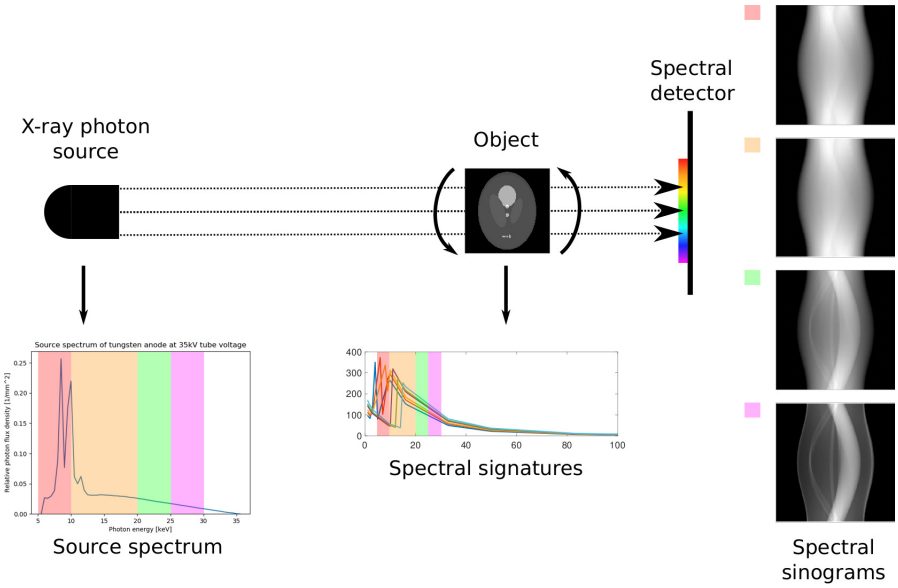
X-ray Computed Tomography (CT) estimates the spatial attenuation map of the object of interest from its measured X-ray projections obtained from different angles. The conventional tomography acquisition setup consists of a polychromatic X-ray source and an X-ray detector that collects the transmitted X-rays. However, these conventional detectors do not discriminate between different incident photon energies and collect the attenuated X-rays solely in one energy bin (Figure 5.1). Images of conventional detectors are reconstructed as grey scale volumes, representing the aggregate attenuation coefficients of the materials. Several materials may correspond to the same grey level, which makes it difficult to determine the material composition of an object. Nonetheless, the attenuations of materials are energy-dependent and their spectral attenuation curves are mutually different. Therefore, by probing multiple energy levels, additional information is obtained for discriminating the materials. A typical approach is dual-energy CT, which allows for more accurate *material decomposition* [40, 274], also known as material unmixing. It uses two polychromatic sources with different peak voltages, and correspondingly two sets of conventional detector panels to measure attenuated X-rays, each from one source. Dual-energy CT is commonly used in clinical settings for separating high-attenuating materials from low-attenuating materials, for example, determining the location of contrast agents [199] such as iodine in the body [142]. However, more accurate material decomposition and concentration determination requires multi-energy CT, enabled by multi-energy X-ray (photon counting) detectors [40, 274].

Multi-energy X-ray detectors allow for collecting X-ray projection data in multiple energy bins with a very high spectral resolution. We can divide these detectors into two classes: (*i*) detectors that measure all X-ray photons simultaneously and directly categorize these into various spectral channels, and (*ii*) detectors that indirectly do this. Detectors from the first category are often used for hyperspectral imaging, and some examples include the Hexitec detector [78, 134, 282], Amptek X-123 CdTe detector [236] and the SLcam [213, 247]. On the other hand, the spectroscopic X-ray detectors from the second category indirectly measure many spectral channels by applying threshold scans. Detectors from the Medipix and Pixirad families are examples of this [24, 25]. The number of energy bins that can be recorded simultaneously is limited (usually up to 10), but different energy thresholds, with which X-ray photons with higher energies than that threshold can be detected, can be set between measurements. If the object of interest is static, hyperspectral images can, in principle, be easily obtained using these devices.

The measurements from these detectors, *i.e.* the tomographic projections after their preprocessing and log-correction, are linearly related to the spectral characteristics of the materials present in the object. Although these detectors can now measure the X-ray projections in multiple energy bins (combined referred to as



(**a**) Data acquisition in conventional X-ray CT



(**b**) Data acquisition in spectral X-ray CT

Figure 5.1: Comparison of data acquisitions in conventional X-ray CT (**a**) and spectral X-ray CT (**b**). In conventional X-ray CT, the energy window is dictated by the source spectrum (grey). None of the detected X-ray photons are distinguished by energy, leading to one sinogram (containing the projections ordered by angle). On the other hand, in spectral CT, the X-ray photons are distinguished according to their energy from multiple windows (colors), yielding multiple sinograms with different characteristics.

spectral projections), it is not straightforward to optimally use the available spectral data for material decomposition. The conventional strategies for reconstruction and material decomposition are not designed to simultaneously process this type of spectral measurements and find accurate material maps (refer to Figure 5.3). Hence, advanced computational techniques are required to infer the material composition from these spectral projections.

5.1.1 Motivation and problem description

Spectral tomographic projections are X-ray measurements at multiple energies that result from the spatial and spectral properties of the materials present in the object. As noted in Section 1.5, *spectral computed tomography* is the process of estimating spatial maps and spectral signatures of materials present in the object from spectral tomographic projections. In general, the spectral CT workflow consists of two steps: (*i*) *reconstruction*: computing spatial maps from their tomographic projections, (*ii*) *unmixing*: decomposition of the spectral volume into spatial material maps and their spectral signatures. Mathematically, reconstruction boils down to solving an inverse problem, while decomposition involves matrix factorization, an unsupervised learning approach. These two steps can be performed in a serial fashion to obtain the material maps from the spectral projections (refer to Figure 5.2). However, these two steps can also be combined into a one-step approach, which we refer to as *joint*.

To illustrate the performance of these two-step and one-step approaches, we consider a Shepp-Logan phantom, as described in Figure 5.3. The numerical phantom consists of five materials (vanadium, chromium, manganese, iron and cobalt). For the tomographic projections, a full-view setting is chosen where 180 projections between 0 to π are acquired. Figure 5.3 shows that the two-step approaches do not yield clear material decompositions or spectral signature reconstructions. Although the two-step methods are computationally efficient, these generally do not yield the same solution due to the ill-posedness of the problem. For the same reason, the joint approach is not expected to work well in all cases either. In particular, the current algorithms for joint methods suffer from ill-conditioning of the spectral profiles (refer to Section 5.4.2). The suboptimal performance of these methods motivates us to develop novel reconstruction methods to improve the spatial resolution and precise characterization of materials. To address the ill-posedness, we incorporate spatial and spectral prior information.

5.1.2 Contributions and outline

In this chapter, we propose a new technique called '*A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography*' (ADJUST) to reconstruct spatial material maps from their spectral tomographic measurements. ADJUST is a novel bi-convex optimization formulation that incorporates an effective *spatiospectral* prior. This prior includes (*i*) *spatial*: the contribution of materials at

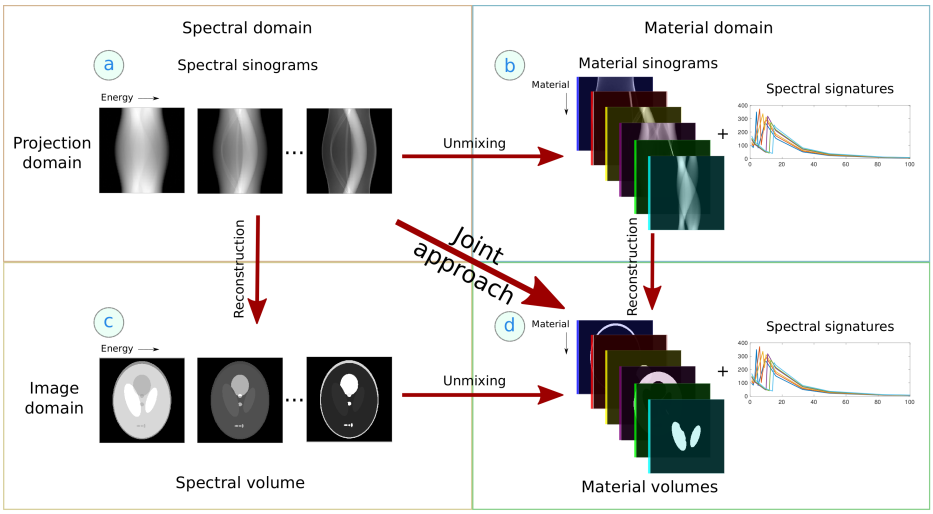


Figure 5.2: A schematic overview of various approaches to spectral computed tomography: The spectral sinograms (projections from various angles) obtained at different energy levels make up the input of these algorithms (**a**). The spectral sinograms can be decomposed first in the projection domain to obtain material sinograms and the spectral signatures (**b**). The spatial material maps (**d**) are then obtained by reconstructing each material sinogram separately. Alternatively, reconstruction can be done before the material decomposition by first making a CT reconstruction of every spectral channel using the associated spectral sinograms to obtain the spectral volumes (**c**). Then, in the image domain, these spectral volumes can be decomposed to obtain the spatial material volumes. These two-step approaches can also be combined into a one-step approach in which the decomposition and reconstruction are carried out as a joint approach.

each location should sum to 1, and (*ii*) *spectral*: the spectral signature of material should be a linear combination of the elements of a spectral dictionary. To solve the constrained optimization problem, we design a memory-efficient alternating proximal iterative scheme. In particular, we develop numerical methods to compute the required proximal operators quickly. In numerical experiments, we demonstrate that ADJUST performs better than the existing state-of-the-art methods. Furthermore, we show that ADJUST is also applicable to limited-angle problems found in industrial X-ray tomography, optical tomography and electron tomography.

The remainder of the chapter is organized as follows. Section 5.2 discusses the existing work on spectral CT and material decomposition. Section 5.3 introduces the forward modelling of spectral X-ray tomography. In particular, we derive the linear map from the spatial map of materials to the spectral tomographic measurements. Then, we introduce the inverse problem in Section 5.4 that estimates the spatial material maps and spectral signatures from the spectral tomographic measurements. Here, we also discuss the ill-posedness involved in the inversion process. To reduce this ill-posedness, we introduce ADJUST in Section 5.5. Moreover, we propose an iterative scheme that finds an approximate solution to the resulting bi-convex

Figure 5.3: Comparison of various methods for spectral CT on a five-material spectral Shepp-Logan phantom. The tomographic projections are gathered in 100 equally-sized bins between 5 to 35 keV. The top row shows the true material maps and the spectral signatures of each material. The second and third row show the material maps retrieved from two-step methods, reconstruction-then-unmixing (RU) and unmixing-then-reconstruction (UR), respectively. The fourth row shows the results of the classical joint method (see Section (5.6)). In contrast, the fifth row shows results of the proposed method (see Section 5.5).

formulation. In Section 5.6, we numerically compare ADJUST with other methods on various synthetic phantoms. We also demonstrate the robustness of the method on limited measurement patterns such as sparse-angle tomography and limited view tomography, as well as in situations with limited spectral resolution and higher noise levels. In addition, we apply ADJUST to an experimental X-ray micro-CT dataset. Finally, we discuss the possibilities and limitations of the approach in Section 5.7 and conclude the chapter.

5.2 Related work

For the sake of convenience, we categorize the previous work on spectral CT into (i) (two-step) sequential approaches and (ii) (one-step) joint approaches to reconstruction and material decomposition. Since we focus on multi-spectral CT, we do not discuss the advances in dual-energy CT. However, we refer the reader to the comprehensive review paper [199] that covers dual-energy CT.

For spectral CT, sequential approaches (also known as two-step methods) are mainly (i) reconstruction followed by decomposition [60, 77, 95, 298, 299, 318], and (ii) decomposition followed by reconstruction [125, 190, 248]. In the former category, material decomposition is carried out in the image domain, while in the

latter category, it is carried out in the projection domain (see Figure 5.2). In both approaches, independent methods for material decomposition in the projection domain [75], (multi-channel) spectral reconstruction [312] with various forms of structural or spectral regularization [143, 244, 246], and material decomposition in the image domain [95] can be plugged in. Although these sequential two-step methods are computationally inexpensive, separating the reconstruction and decomposition steps causes information loss [202, 292].

Joint methods (also known as one-step methods) that simultaneously reconstruct multi-channel spectral images and perform material decomposition have been developed to address the issues associated with sequential methods [136, 202]. All current one-step methods are iterative in nature and allow for the incorporation of prior knowledge through regularization. For example, (*i*) structure on material maps is imposed through various penalties [125, 202], (*ii*) material maps are constrained using simplex constraints [95, 298, 299], and (*iii*) structure on spectral signatures is enforced using a spectral dictionary [299, 316]. In particular, when the materials present in the object of interest are precisely known, various techniques improve the quality of the reconstructions. These tailored methods can image contrast agents, such as iodine [318], gold, gadolinium in angiography [190], in the presence of bone and tissue. Sometimes, knowledge about the materials can be used to choose the spectral bins effectively [95]. Moreover, a deep learning approach has been proposed to perform joint decomposition and reconstruction task by generating a training set based on synthetic phantoms [5]. Most existing methods have been developed for, or demonstrated on, a limited number of materials (for example, two [136, 166, 189], three [27, 316], and six [95]), sometimes heavily relying on prior information about the materials, spectral signatures and energy bins. Some of these methods are extendable to more materials [5], but this can be difficult for each additional parameter that may need to be estimated with each new material [27]. However, these methods suffer when (*i*) the number of materials present in the object is larger than 3, (*ii*) the number of projections is smaller than the conventional criterion, (*iii*) the measurements are corrupted with high noise [118].

5.3 Spectral forward model

In this section, we provide the spectral X-ray forward model, by describing how we represent the objects and how spectral X-ray projections are obtained from this. The object is characterised by its attenuation coefficients $\mu(x, E) \in \mathbb{R}_+$, where $x \in \mathbb{R}^d$ is the location and $E > 0$ the X-ray photon energy, with $d \in \{2, 3\}$ being the dimension of our space depending on whether we consider a slice-based or full 3D reconstruction. Given a polychromatic X-ray source with a source spectrum $I_0(E)$ at energy level E, and C energy windows from the set $\mathcal{E} = \{\mathcal{E}_c\}_{c\in\mathcal{C}} = \left\{\left[E_c^{\min}, E_c^{\max}\right]\right\}_{c\in\mathcal{C}}$ (with $\mathcal{C}$ being its index set with $|\mathcal{C}|= C$) in which a spectral detector captures associated X-ray photons, we model the total X-ray photons captured by a detector pixel in

energy bin $\mathcal{E}_c$ as follows:

$$I(\mathcal{E}_c) = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E) \exp\left(-\int_\ell \mu(x,E)\,\mathrm{d}x\right) \mathrm{d}E. \tag{5.1}$$

Here, the inner integral is taken over the line ℓ from the X-ray source to a detector pixel. The maximum and minimum energy range depend on detector specifications. We model the energy-dependent attenuation as a linear combination of energy-dependent material attenuations and their spatial contributions. We represent it mathematically as

$$\mu(x,E) = \sum_{m\in\mathcal{M}} \mu_m(E)\alpha_m(x), \tag{5.2}$$

where the material attenuation coefficient μ_m is a function of energy. The material m must be contained in a set of considered materials $\mathcal{M}$ of size M. The proportion of material m at location x is given by α_m. From Equations (5.1) and (5.2), we arrive at the following continuous relationship of measured photons in terms of the material spatial distributions and their attenuations:

$$I(\mathcal{E}_c) = \int_{E_c^{\min}}^{E_c^{\max}} I_0(E) \exp\left(-\sum_{m\in\mathcal{M}} \mu_m(E) \int_\ell \alpha_m(x)\,\mathrm{d}x\right) \mathrm{d}E.$$

If the spectral bins are sufficiently narrow then the source spectrum and the material-attenuation values can be approximated by their representative (average) values. These values are respectively $\overline{I}_0(\mathcal{E}_c)$ and $\overline{\mu}_m(\mathcal{E}_c)$, and form the following relation for the photon count in energy bin $\mathcal{E}_c$ corresponding to the cth channel:

$$\begin{aligned} I(\mathcal{E}_c) &\approx \overline{I}(\mathcal{E}_c) \\ &= \overline{I}_0(\mathcal{E}_c) \exp\left(-\sum_{m\in\mathcal{M}} \overline{\mu}_m(\mathcal{E}_c) \int_\ell \alpha_m(x)\,\mathrm{d}x\right). \end{aligned}$$

In general, the photon count is perturbed by an energy dependent noise distribution. Moreover, the spectral X-ray detector is a photon-counting detector, for which the noise in energy bin $\mathcal{E}_c$ can be modelled using a Poisson distribution with parameter $I(\mathcal{E}_c)$ [221]. If the mean in each energy bin is sufficiently high, a realization $I^{\text{meas}}(\mathcal{E}_c)$ of $I(\mathcal{E}_c)$ measured by a spectral detector can be approximated using Gaussian distribution $\mathcal{N}(0,\sigma^2)$, with variance σ being inversely proportional to $I(\mathcal{E}_c)$. In our experiments, we assume that the average photon count in each bin of detectors are sufficient for this approximation.

5.4 Spectral inverse problem

From the measured spectral X-ray projections, *spectral computed tomography* aims to retrieve the energy-dependent attenuation values $\mu_m(E_c)$ for each material m

and energy channel E_c and the distribution α_m throughout the object for each material m. It boils down to solving the bilinear system, in terms of $\overline{\mu}_m$ and α_m,

$$\sum_{m\in\mathcal{M}} \overline{\mu}_m(\mathcal{E}_c) \int_{\ell_j} \alpha_m(x)dx = -\ln\left(\frac{I_j^{\text{meas}}(\mathcal{E}_c)}{\overline{I}_0(\mathcal{E}_c)}\right), \quad \forall\, \ell_j \in L,\, \mathcal{E}_c \in \mathcal{E}, \tag{5.3}$$

with L and $\mathcal{E}$ denoting the set of rays (with size $|L|= J$) and the set of energy channels, respectively. In tomography, the material distributions are determined by discretizing the object space into the grid of either pixels (2D) or voxels (3D). For now, we consider a three-dimensional scene with N voxels. The proportion of material m in the ith voxel is then given by a_{im}. For each ray over line ℓ_j corresponding to the jth measurement, the quantity w_{ji} determines the contribution of ith voxel to the jth measurement. Usually, the quantity $\overline{I}_0(\mathcal{E}_c)$ can be determined accurately by performing a flatfield measurement (i.e. measurement without object). Therefore, the right-hand side of Equation (5.3) is known. By expressing $-\ln\left(I_j^{\text{meas}}(\mathcal{E}_c)/\overline{I}_0(\mathcal{E}_c)\right)$ by y_{jc} and $\overline{\mu}_m(\mathcal{E}_c)$ by f_{mc}, we arrive at the following expression:

$$\sum_{m=1}^{M}\left(\sum_{i=1}^{N} w_{ji}a_{im}f_{mc}\right) = y_{jc}, \qquad j = 1,\ldots,J, \text{ and } c = 1,\ldots,C. \tag{5.4}$$

Here, the total number of measurements for each channel is given by J, and C denotes the total number of (energy) channels. Subsequently, we can write the expression in Equation (5.4) in the following matrix notation

$$\boldsymbol{WAF} = \boldsymbol{Y}, \tag{5.5}$$

where $\boldsymbol{Y} \in \mathbb{R}^{J\times C}$ represents tomographic measurements for C number of channels, and $\boldsymbol{W} \in \mathbb{R}^{J\times N}$ is a projection matrix containing the weights w_{ji} described in Equation (5.4), $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ consists of M columns of size N, with each column representing a spatial map corresponding to the material present in the object, while $\boldsymbol{F} \in \mathbb{R}^{M\times C}$ consists of M rows with each row denoting the channel attenuation information of the material. It is important to note that matrices $\boldsymbol{W}$ and $\boldsymbol{Y}$ are known and matrices $\boldsymbol{A}$ and $\boldsymbol{F}$ are unknown. We formulate the joint spectral tomographic imaging and decomposition problem in a constrained least-squares form as

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{F}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{A},\boldsymbol{F}) \triangleq \tfrac{1}{2}\|\boldsymbol{Y} - \boldsymbol{WAF}\|_F^2, && \text{(least-squares misfit)} \\ \text{subject to} \quad & \boldsymbol{A} \geq 0, && \text{(non-negativity of material maps)} \\ & \boldsymbol{F} \geq 0. && \text{(non-negativity of attenuation coefficients)} \end{aligned} \tag{5.6}$$

Here, we impose non-negativity constraints on both $\boldsymbol{A}$ and $\boldsymbol{F}$. The function $\mathcal{J}(\boldsymbol{A}, \boldsymbol{F}) : \mathbb{R}^{N\times M} \times \mathbb{R}^{M\times C} \mapsto \mathbb{R}$ defines the misfit between the true measurements $\boldsymbol{Y}$ and the simulated measurements $\boldsymbol{WAF}$ using the Frobenius norm. This norm is valid if the noise in the measurements is approximately Gaussian. We note that this joint formulation is a bi-convex optimization problem since the misfit function is bi-convex, and the constraint set is a bi-convex set [103]. We denote the solution set of the joint formulation by

$$\mathcal{B} = \left\{ (\boldsymbol{A}_{\text{joint}}, \boldsymbol{F}_{\text{joint}}) \right\} = \underset{\boldsymbol{A}\geq 0, \boldsymbol{F}\geq 0}{\arg\min}\ \mathcal{J}(\boldsymbol{A}, \boldsymbol{F}).$$

This solution set $\mathcal{B}$ may contain more than one solution if the misfit function $\mathcal{J}$ is not strongly bi-convex. The solution set $\mathcal{B}$ cannot be determined trivially. To find the elements in set $\mathcal{B}$, we need to solve the optimization problem (5.6) using an iterative scheme with an initial estimate of the solution [290].

5.4.1 Practical methods

As outlined in Section 5.2, joint methods aim to simultaneously estimate the spatial material distribution and the energy-dependent attenuation coefficients. However, these one-step methods are not practical due to their high computational cost. In practice, two-step methods are popular, where reconstruction and decomposition are performed separately, because of their modular nature. For each step, tailored solvers are readily available for different platforms. The first category of two-step methods, which we call *RU* (short for Reconstruction-then-Unmixing), estimates a spectral volume from the spectral tomographic measurements, and then decomposes the resulting spectral volume to obtain the material maps and spectral signatures. That is, *RU* solves the following problems in a serial fashion:

$$\begin{aligned}
\boldsymbol{V}_{\text{RU}} &= \underset{\boldsymbol{V}\geq 0}{\arg\min}\ \frac{1}{2}\|\boldsymbol{WV} - \boldsymbol{Y}\|_F^2 + \lambda\mathcal{R}_1(\boldsymbol{V}) \\
(\boldsymbol{A}_{\text{RU}}, \boldsymbol{F}_{\text{RU}}) &= \underset{\boldsymbol{A}\geq 0, \boldsymbol{F}\geq 0}{\arg\min}\ \frac{1}{2}\|\boldsymbol{AF} - \boldsymbol{V}_{\text{RU}}\|_F^2,
\end{aligned}$$

where $\boldsymbol{V}_{\text{RU}} \in \mathbb{R}^{N\times C}$ is a spectral volume, and $\boldsymbol{A}_{\text{RU}}, \boldsymbol{F}_{\text{RU}}$ are the material maps and spectral signatures respectively reconstructed by this method. $\mathcal{R}_1 : \mathbb{R}^{N\times C} \mapsto \mathbb{R}$ is a regularization function that incorporates prior information about the spectral volumes and $\lambda \geq 0$ is a regularization parameter. Contrarily, *UR* (short for Unmixing-then-Reconstruction), the other class of two-step methods, separates the spectral tomographic measurements into projections and spectral signatures. These projections then lead to the material maps. *UR* mathematically reads

$$\begin{aligned}
(\boldsymbol{P}_{\text{UR}}, \boldsymbol{F}_{\text{UR}}) &= \underset{\boldsymbol{P}\geq 0, \boldsymbol{F}\geq 0}{\arg\min}\ \frac{1}{2}\|\boldsymbol{PF} - \boldsymbol{Y}\|_F^2 \\
\boldsymbol{A}_{\text{UR}} &= \underset{\boldsymbol{A}\geq 0}{\arg\min}\ \frac{1}{2}\|\boldsymbol{WA} - \boldsymbol{P}_{\text{UR}}\|_F^2 + \gamma\mathcal{R}_2(\boldsymbol{A})
\end{aligned}$$

where $\boldsymbol{P}_{\mathrm{UR}} \in \mathbb{R}^{J \times M}$ is a material volume, and $\boldsymbol{F}_{\mathrm{UR}}, \boldsymbol{A}_{\mathrm{UR}}$ are the material maps and spectral signatures respectively reconstructed by this method. Here, $\mathcal{R}_2 : \mathbb{R}^{N \times M} \mapsto \mathbb{R}$ integrates prior information about the material maps with $\gamma \geq 0$ being the regularization parameter. These practical methods only work well when a complete projection series is available and the object is composed of materials whose spectra are clearly separable. However, the spectra of materials as measured by industrial spectral tomography equipment can be very similar. Although these practical methods do not give accurate solutions, the results can be used as an initial guess for advanced reconstruction methods.

5.4.2 Ill-posedness

The Hadamard conditions to define a *well-posed* problem consist of three criteria: (*i*) *existence*: There must be an $\boldsymbol{A}^\star$ and $\boldsymbol{F}^\star$ that satisfy $\boldsymbol{W}\boldsymbol{A}^\star\boldsymbol{F}^\star = \boldsymbol{Y}$. (*ii*) *uniqueness*: The solution $\boldsymbol{A}^\star$ and $\boldsymbol{F}^\star$ must be unique. (*iii*) *stability*: small perturbations in the measurements $\boldsymbol{Y}$ should not lead to significant deviations in $\boldsymbol{A}^\star$ and $\boldsymbol{F}^\star$. If any of these conditions is violated, we call the problem *ill-posed*. In general, we assume the existence of a solution to the least-squares problem since we use the Euclidean norm to measure the misfit in the discrete setting. However, the uniqueness condition needs to be verified. Moreover, the stability of the solution relies on the conditioning of projection matrix $\boldsymbol{W}$ and the measurements $\boldsymbol{Y}$.

In general, the spectral inverse problem has multiple solutions if no prior information is incorporated. To see this, suppose $(\boldsymbol{A}^\star, \boldsymbol{F}^\star)$ is a solution to Equation (5.5), then $(\alpha\boldsymbol{A}^\star, (1/\alpha)\boldsymbol{F}^\star)$ is also a solution to (5.5) for any $\alpha > 0$. Hence, the practical reconstruction methods and classical joint method solve an ill-posed problem. To reduce the non-uniqueness, and hence make the problem less ill-posed, we need to incorporate appropriate spatiospectral prior information.

5.5 Proposed method - ADJUST

Since the conventional spectral inverse problem remains ill-posed due to the non-uniqueness of solutions, we propose to incorporate spectral information through a spectral dictionary. The spectral profiles of many materials are readily available [130, 301], and the spectral responses in each spectral channel can easily be computed from these spectral profiles. We model the spectral response for material m as the binary combination of the dictionary elements. That is,

$$\boldsymbol{f}_m = \widehat{r}_{m1}\boldsymbol{t}_1 + \widehat{r}_{m2}\boldsymbol{t}_2 + \cdots + \widehat{r}_{mD}\boldsymbol{t}_D,$$

where $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_D$ correspond to the spectral responses of D distinct dictionary materials, and $\widehat{r}_{m1}, \widehat{r}_{m2}, \ldots, \widehat{r}_{mD}$ are the coefficients of material m that take the value of either 0 or 1. Suppose the jth material in the dictionary corresponds to the material m, then $\widehat{r}_{mj} = 1$, and the other coefficients will be zero. Hence, we

can represent the spectral matrix $\boldsymbol{F} \in \mathbb{R}^{M \times C}$ as

$$\underbrace{\begin{bmatrix} - & \boldsymbol{f}_1 & - \\ & \vdots & \\ - & \boldsymbol{f}_M & - \end{bmatrix}}_{\boldsymbol{F}} = \underbrace{\begin{bmatrix} - & \widehat{\boldsymbol{r}}_1 & - \\ & \vdots & \\ - & \widehat{\boldsymbol{r}}_M & - \end{bmatrix}}_{\widehat{\boldsymbol{R}}} \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{t}_1 & \dots & \boldsymbol{t}_D \\ | & & | \end{bmatrix}}_{\boldsymbol{T}},$$

where $\boldsymbol{T} \in \mathbb{R}^{D \times C}$ is a dictionary of D materials (with $D \geq M$) with spectral information for C channels, and $\widehat{\boldsymbol{R}} \in \{0,1\}^{M \times D}$ is a spectral coefficient matrix.

Due to the binary constraints, finding such a matrix $\widehat{\boldsymbol{R}}$ jointly with $\boldsymbol{A}$ is a non-convex problem. To make it convex for fixed $\boldsymbol{A}$, we relax the binary nature of the variable $\widehat{\boldsymbol{R}}$. Moreover, we apply additional constraints on the material maps to ensure that the total contribution of materials at every voxel does not exceed 1. Thus, the resulting formulation, termed as *A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography* (ADJUST), is phrased as

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{R}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{A},\boldsymbol{R}) \triangleq \tfrac{1}{2}\|\boldsymbol{Y} - \boldsymbol{WART}\|_F^2, && \text{(least-squares misfit)} \\ \text{subject to} \quad & \boldsymbol{A} \in \mathcal{C}_A, && \text{(constraints on spatial map)} \\ & \boldsymbol{R} \in \mathcal{C}_R, && \text{(constraints on dictionary coefficients)} \end{aligned} \tag{5.7}$$

where $\boldsymbol{Y} \in \mathbb{R}^{J \times C}$ represents spectral tomographic measurements, $\boldsymbol{W} \in \mathbb{R}^{J \times N}$ is a projection matrix containing the weights w_{ji} described in Equation (5.4), $\boldsymbol{A} \in \mathbb{R}^{N \times M}$ is the matrix that constitutes the spatial contributions of the materials, $\boldsymbol{T} \in \mathbb{R}^{D \times C}$ represents the fixed dictionary matrix containing attenuation spectra of many materials, and $\boldsymbol{R} \in \mathbb{R}^{M \times D}$ is the dictionary coefficient matrix that represents the *continuous* version of $\widehat{\boldsymbol{R}}$. The constraint sets are

$$\mathcal{C}_R \triangleq \left\{ \boldsymbol{X} \in \mathbb{R}^{M \times D} \,|\, \underbrace{x_{ij} \geq 0}_{(a)}, \underbrace{\sum_{j=1}^{D} x_{ij} \leq 1}_{(b)}, \underbrace{\sum_{i=1}^{M} x_{ij} \leq 1}_{(c)} \right\},$$

$$\mathcal{C}_A \triangleq \left\{ \boldsymbol{X} \in \mathbb{R}^{N \times M} \,|\, \underbrace{x_{ij} \geq 0}_{(d)}, \underbrace{\sum_{j=1}^{M} x_{ij} \leq 1}_{(e)} \right\}.$$

We provide the details of each constraint below:

(a) *Non-negativity of* $\boldsymbol{R}$: Since $\boldsymbol{R}$ is a dictionary-coefficient matrix that is a convex proxy for $\widehat{\boldsymbol{R}}$, the values must be greater than or equal to 0.

(b) *Row-sum constraints for* $\boldsymbol{R}$: In principle, we would like to impose that each material present in the object must be a part of the dictionary. The convex approximation of this condition is that the total contribution of the dictionary elements to represent material should not exceed 1.

(c) *Column-sum constraints for* $\boldsymbol{R}$: Each column of $\boldsymbol{R}$ represents the contribution of the dictionary element to generate materials in the object. Since the number of materials in the object is smaller than the total number of dictionary elements, the contribution of many dictionary elements will be 0. Moreover, the materials present in the object must be distinct. Hence, the contribution of dictionary elements must not exceed 1. Hence, the column-sum constraints impose these conditions.

(d) *Non-negativity of* $\boldsymbol{A}$: Each material should have a non-negative contribution to every voxel.

(e) *Row-sum constraints for* $\boldsymbol{A}$: The total contribution of materials in each voxel must not exceed 1.

We enumerate the benefits of ADJUST as follows. (*i*) The ADJUST formulation is less ill-posed when compared with the two-step methods or the Joint formulation given in (5.6). This is due to the fact that the incorporation of the spectral dictionary resolves the scaling issue. (*ii*) It is a *parameter-free* approach since the constraints are simplex and do not involve any parameters that need to be estimated. The only parameter ADJUST requires is the number of materials present in the object. However, this prior knowledge is generally available to the user. (*iii*) The optimization problem (5.7) is bi-convex (see Appendix C.2 for the proof). Hence, it can be solved efficiently using the iterative minimization method.

5.5.1 Numerical optimization

To obtain an approximate solution to (5.7), many alternating minimization schemes exist [20, 43, 156]. However, these schemes rely on complete minimization with respect to at least one variable in every step. Moreover, their convergence to a partially optimal solution is slow. Hence, such schemes might not be computationally feasible for large-scale problems. For practical applications, we propose an accelerated variant of Proximal Alternating Linearized Minimization (PALM) [43], by combining it with the acceleration strategy in alternating direction method of multipliers (ADMM) [220]. We term this variant as 'Alternating Accelerated Proximal Minimization' (AAPM):

$$
\begin{aligned}
&\textbf{for } k = 0, \ldots, K-1: \\
&\quad \boldsymbol{R}_{k+1} = \mathbf{proj}_{\mathcal{C}_R}\left(\boldsymbol{R}_k - \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k)\right) && \text{(update dictionary coefficients)} \\
&\quad \boldsymbol{A}_{k+1} = \mathbf{proj}_{\mathcal{C}_A}\left(\boldsymbol{A}_k - \beta \nabla_{\boldsymbol{A}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_{k+1}, \boldsymbol{U}_k)\right) && \text{(update material maps)} \\
&\quad \boldsymbol{U}_{k+1} = \boldsymbol{U}_k + \rho\left(\boldsymbol{W}\boldsymbol{A}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{T} - \boldsymbol{Y}\right) && \text{(running-sum-of-errors)}
\end{aligned}
\tag{5.8}
$$

where α and β are estimated using a line-search method (*e.g.*, backtracking), ρ is an acceleration parameter chosen from the range $[10^{-3}, 1)$, and $\widetilde{\mathcal{J}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U}) = \mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) + \langle \boldsymbol{U}, \boldsymbol{Y} - \boldsymbol{WART} \rangle$. The variable U contains the running sum of errors (*i.e.* residuals):

$$
\boldsymbol{U}_{k+1} = \boldsymbol{U}_0 + \rho \sum_{i=1}^{k} \boldsymbol{R}_i \quad \text{where} \quad \boldsymbol{R}_i = \boldsymbol{W}\boldsymbol{A}_i\boldsymbol{R}_i\boldsymbol{T} - \boldsymbol{Y}.
$$

For $\rho = 0$, AAPM is equivalent to PALM. When $\rho > 0$, the error signal U is driven to zero by feeding back the integral of the error to its input. In Figure 5.4, we plot the residuals versus iterations for various values of ρ. From these results, we conclude that acceleration can be achieved by including the running-sum-of-errors into an alternative iteration scheme. As the ρ values are increased, the residual decreases faster. However, for higher values of ρ, the monotonic decrease of the residual disappears. We consider ρ as a hyperparameter for which a reasonable value can be determined heuristically. The derivation of the method is given in Appendix C.3.

It is easy to compute the partial derivatives from basic linear algebra and calculus rules (refer to Appendix C.4). The partial derivatives are

$$
\begin{aligned}
\nabla_{\boldsymbol{A}} \widetilde{\mathcal{J}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U}) &= \boldsymbol{W}^T(\boldsymbol{WART} - \boldsymbol{Y} - \boldsymbol{U})\,\boldsymbol{T}^T\boldsymbol{R}^T, \\
\nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U}) &= \boldsymbol{A}^T\boldsymbol{W}^T(\boldsymbol{WART} - \boldsymbol{Y} - \boldsymbol{U})\,\boldsymbol{T}^T.
\end{aligned}
$$

The proximal operators are derived in Section 5.5.2. We use the following parameters to determine the stopping criteria:

$$
\begin{aligned}
\epsilon_k^{\text{abs}} &= \|\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{A}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{T}\|_F / \|\boldsymbol{Y}\|_F \\
\epsilon_k^{\text{rel}} &= \|\boldsymbol{A}_{k+1} - \boldsymbol{A}_k\|_F + \|\boldsymbol{R}_{k+1} - \boldsymbol{R}_k\|_F
\end{aligned}
$$

where k is an iteration of the optimization scheme (5.8). The benefits of AAPM are the following:

- *Simple gradient computations*: The gradients have explicit expressions and can be computed using simple matrix-matrix multiplications.

- *Fast proximal operations*: The proximal operations are just orthogonal projections onto convex sets. These operations have either an explicit expression or a simple alternating routine to find the proximal point efficiently.
- *Backtracking line-search*: The improved line-search (*i.e.* finding α, β) makes sure that the progress in the descent direction is appropriate for every iterate.
- *Acceleration through running-sum-of-errors*: The regular update of variable $\boldsymbol{U}$, which contains the running-sum-of-errors, helps in accelerating the convergence to the partial optimal solution.
- *Memory-efficiency*: The method relies only on forward and adjoint operations with the tomography operator $\boldsymbol{W}$. Hence, it saves memory to explicitly store the tomography operator in either single or double precision.

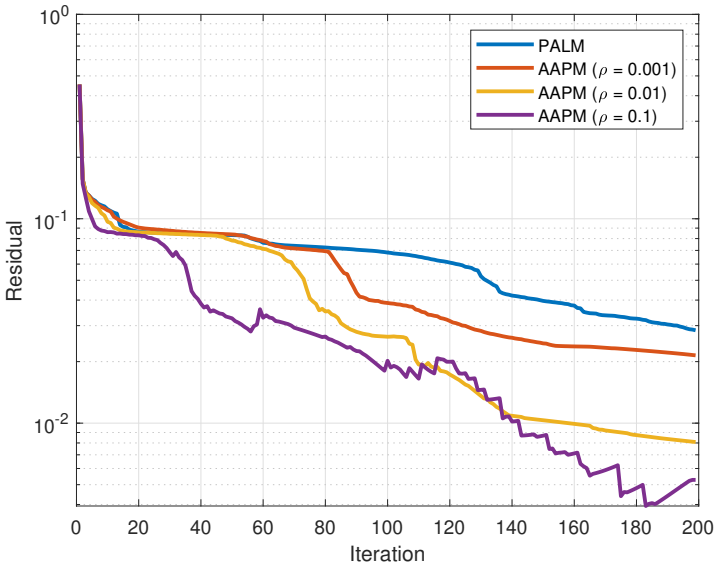


Figure 5.4: Numerical demonstration of proposed acceleration scheme on the Shepp-Logan phantom for various ρ values.

5.5.2 Orthogonal projections

In this section, we derive the projections onto the convex sets $\mathcal{C}_R$ and $\mathcal{C}_A$. For $\mathcal{C}_A = \left\{\boldsymbol{X} \in \mathbb{R}^{N\times M} \,|\, \boldsymbol{X} \geq 0,\ \boldsymbol{X}\mathbf{1} \leq \mathbf{1}\right\}$, its orthogonal projection takes the following form:

$$\mathbf{proj}_{\mathcal{C}_A}(\boldsymbol{Z}) = \max\left(\boldsymbol{Z} - \boldsymbol{\Lambda}\mathbf{1}, \mathbf{0}\right), \tag{5.9}$$

where $\boldsymbol{\Lambda} = \mathbf{diag}\left(\lambda_1, \ldots, \lambda_N\right)$ is the diagonal matrix with weights computed from solving the inequalities

$$\left\langle \mathbf{1}, \max\left(\boldsymbol{z}_{i:}^T - \lambda_i \mathbf{1}, \mathbf{0}\right)\right\rangle \leq 1, \quad i = 1, \ldots, N, \tag{5.10}$$

with $\boldsymbol{z}_{i:}$ being the ith row of the matrix $\boldsymbol{Z}$ [34, Theorem 6.27]. To find the optimal weight λ_i, we carry out bisection on λ_i for which $\max\left(\left\langle \mathbf{1}, \max\left(\boldsymbol{z}_{i:}^T - \lambda_i \mathbf{1}, \mathbf{0}\right)\right\rangle - 1, 0\right) = 0$ holds, starting with the initial interval $[0, \max(\boldsymbol{z}_{i:})]$. The function given by $\left\langle \mathbf{1}, \max\left(\boldsymbol{z}_{i:}^T - \lambda_i \mathbf{1}, \mathbf{0}\right)\right\rangle = 1$ is piecewise linear, with breakpoints at the values $z_{i1}, \ldots, z_{iM}$. Hence, once we have localized λ_i to be between two adjacent values, we can immediately compute the optimal value $\lambda_i^\star$. Furthermore, the following theorem entails the projection onto set $\mathcal{C}_R$:

Theorem 1. *The convex set* $\mathcal{C} = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}\mathbf{1} \leq \mathbf{1}, \boldsymbol{X}^T\mathbf{1} \leq \mathbf{1}\right\}$ *is composed of convex set* $\mathcal{C}_1 = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}\mathbf{1} \leq \mathbf{1}\right\}$ *and convex set* $\mathcal{C}_2 = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}^T\mathbf{1} \leq \mathbf{1}\right\}$. *The projection of point* $\boldsymbol{Z} \in \mathbb{R}^{M\times D}$ *onto set* $\mathcal{C}$ *is given by the fixed-point iteration scheme*

$$\boldsymbol{X}_{t+1} = \mathbf{proj}_{\mathcal{C}_1}\left(\mathbf{proj}_{\mathcal{C}_2}\left(\frac{\boldsymbol{X}_t + \boldsymbol{Z}}{2}\right)\right), \quad t = 0, \ldots, D \tag{5.11}$$

with $\boldsymbol{X}_0 = \boldsymbol{Z}$.

Proof. The proof is given in Appendix C.1. □

For the set $\mathcal{C}_2 = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}^T\mathbf{1} \leq \mathbf{1}\right\}$, the orthogonal projection, derived from [34, Theorem 6.27], takes the form

$$\mathbf{proj}_{\mathcal{C}_2}(\boldsymbol{Z}) = \max\left(\boldsymbol{Z} - \mathbf{1}\boldsymbol{\Omega}, \mathbf{0}\right), \tag{5.12}$$

with $\boldsymbol{\Omega} = \mathbf{diag}\left(\omega_1, \ldots, \omega_D\right)$ is the diagonal matrix with weights computed from solving the equation

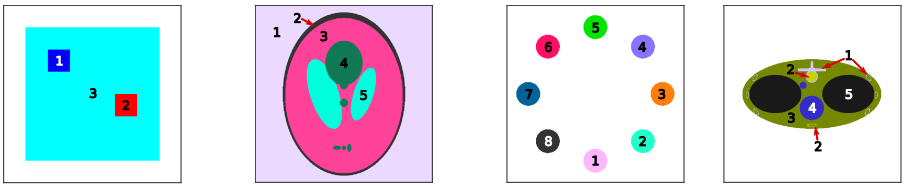
$$\left\langle \mathbf{1}, \max\left(\boldsymbol{z}_i - \omega_i \mathbf{1}, \mathbf{0}\right)\right\rangle = 1, \quad i = 1, \ldots, D. \tag{5.13}$$

5.6 Experiments and results

This section compares ADJUST with the sequential methods (RU, UR), the classical joint method (cJoint), and five state-of-the-art joint methods on a benchmark synthetic spectral phantom. After this, we compare ADJUST with the sequential methods and the classical joint method on more advanced spectral phantoms (in terms of number of materials and material shapes). Next, we examine the robustness of ADJUST against various limited measurement patterns. Finally, we show the application of ADJUST on spectral X-ray micro-CT dataset. Additional numerical experiments are presented in Appendices C.7, C.8, C.9, and C.10.

5.6.1 Experimental setup

Before presenting the results, we first outline the experimental setup. We describe the phantoms that are used, the settings for the attenuation spectra and the source spectrum, the chosen materials for each phantom and a discussion on the implementation of the algorithms.



Mory phantom **Shepp-Logan Phantom** **Disks phantom** **Thorax phantom**

Figure 5.5: Visualizations of the numerical phantoms used in the studies.

Spectral phantoms

A number of phantoms are used in our numerical studies, some of which are standardized while others are custom-made. All these phantoms are shown in Figure 5.5.

Mory phantom We use a slightly modified version of the phantom provided in the work by Mory et al. [202] for comparing new one-step methods against the five one-step approaches addressed in their work. The phantom contains three different materials on a 128×128 grid. As opposed to the original phantom, each location contains only one material.

Shepp-Logan phantom This standard phantom is commonly used in tomography for benchmarking. We modify this phantom to have five unique grey values. We discretize it on a 512×512 uniform grid.

Disks phantom In this custom phantom, several disks with different materials are placed on a circle. The phantom is created so that we can place up to 15 different disks on this circle. However, for our numerics, we have taken eight disks and discretized the resulting phantom on a 512×512 pixel grid.

Thorax phantom We use a modified thorax phantom provided in the CONRAD software framework [182]. We created a thorax phantom of 512^3 voxels, took slice $z = 255$, and removed a few ribs. The resulting 512×512 phantom has eight different material candidates, on which we assign five different materials.

Inverse crime refers to the process of using the same forward operator for the generation of synthetic measurements as for the subsequent reconstruction process [139]. To avoid inverse crime in all of our numerical experiments, we generate measurements by increasing the spatial resolution of the spectral phantoms by a factor of 2. For example, the Shepp-Logan phantom is discretized on a grid of 1024×1024 pixels to generate the spectral tomographic measurements. These measurements are, however, acquired on 512 (equally-spaced) detector pixels for 180 projection angles in $[0, \pi)$. The spectral tomographic inversion is then performed on a grid of 512×512.

Attenuation spectra, source spectrum and selected materials

To generate the spectral sinograms and construct the dictionary matrix $\boldsymbol{T}$, we use attenuation spectra provided by the National Institute for Standards and Technology (NIST) [130, 301]. We perform linear interpolation to approximate the real spectra and discretize these such that the energetic centers are located at 100 equidistant values ranging from 5 to 35 keV. We use the corresponding interpolated attenuation values as representative attenuation values in the bins. For experiments with the Thorax phantom, we use a spectral range of $[20, 80]$ keV. Regarding the experiments with the Mory phantom, we use the attenuation spectra provided by Mory et al. in their implementation [100], providing 100 equidistant bins with energetic centers between 20 and 119 keV. In accordance with their data preprocessing procedure, we have scaled both the material maps and the dictionary entries for iodine and gadolinium by their densities and a value of 0.01 to obtain a concentration of 10 mg/ml.

The materials chosen for each phantom (Figure 5.5) are given in Table 5.1. For the up to 42 chosen materials included in the dictionary matrix, we refer to Appendix C.5. For the Thorax phantom, the dictionary consists of the materials that appear in the phantom. This also holds for the Mory phantom.

To generate the source spectrum, we make use of the SpekPy software [47, 230]. With this, we have simulated an X-ray source spectrum from a molybdenum source with a peak voltage of 35kV. This source material provides us a with low energy bound of 5 keV for a positive flux, enabling us to include more material absorption edges (mostly K-edges) in our simulations than with other source materials such as tungsten. However, for experiments with the Thorax phantom, we use a tungsten source with a peak voltage of 80kV. Both source spectra are shown in Figure 5.6. For the Mory phantom, we use the source spectrum provided in their implementation [100], and have changed the matrix for detector response to an identity matrix.

	Shepp-Logan	Disks	Thorax	Mory
1	Vanadium	Arsenic	Bone	Iodine
2	Chromium	Selenium	90% Blood + 10% Iodine	Gadolinium
3	Manganese	Bromine	Soft tissue	Water
4	Iron	Krypton	Blood	-
5	Cobalt	Rubidium	Lung tissue	-
6	-	Strontium	-	-
7	-	Yttrium	-	-
8	-	Zirconium	-	-

Table 5.1: Selection of materials for each phantom.

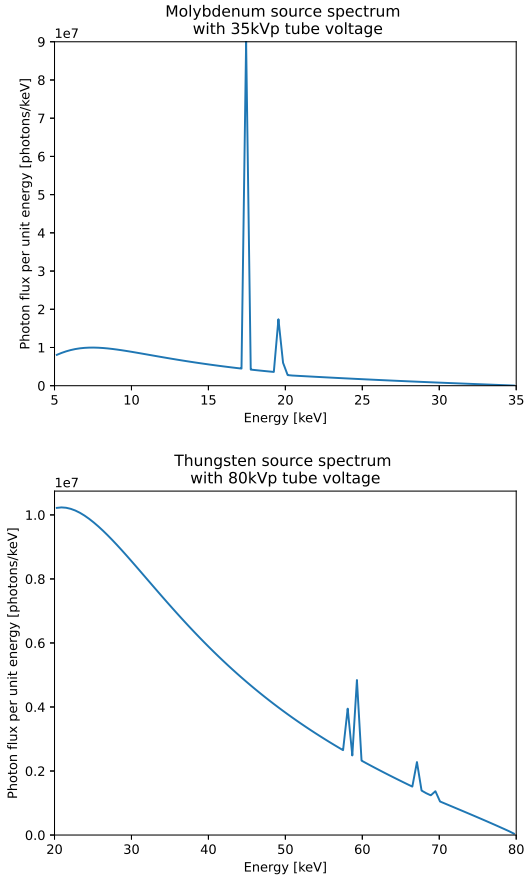


Figure 5.6: The simulated molybdenum-based and tungsten-based source spectra with respectively peak voltages of 35kVp and 80kVp used in the experiments.

Implementation of the numerical algorithms

We briefly describe the implementation of various algorithms used in our studies below. For the existing one-step methods, we keep a similar naming convention as in the work of Mory et al. [202] by using the last name of the first author of the associated paper. We use the ASTRA toolbox [1, 2] for the simulation of the X-ray projections and implementation of forward and the adjoint operator of tomography [41]. For compatibility with the implementations of the existing methods in [100], we use the AIR Toolbox [113] for the computation of the forward operator and the adjoint instead of ASTRA whenever the Mory phantom is considered. For the last five listed existing one-step methods, we use the implementation and parameter values provided in [100]. Table 5.2 summarizes the prior information used in these methods.

RU For the reconstruction, we solve the Tikhonov-regularized optimization problem with regularization parameter λ set to 10^{-3}. We perform a maximum of 20 conjugate-gradient iterations on the resulting normal equations with the tolerance of 10^{-6} [123]. We use non-negative matrix factorization (NMF) with an alternating least-squares algorithm [39] for 100 maximum iterations in the decomposition step. Since NMF is a non-convex problem, we use ten different initializations to determine the solution.

UR We use the same settings for the decomposition and the reconstruction steps as described in RU.

cJoint We solve the problem using an alternating minimization scheme. The maximum number of iterations is set to 2000 with tolerance, defined as the relative residual, of 10^{-4}. In each iterate, we solve the minimization with a spectral projected gradient scheme [250].

ADJUST We use the AAPM scheme described in (5.8) to find the solution. For all the experiments, we choose ρ value of 10^{-2} and set $\epsilon_k^{\text{abs}}, \epsilon_k^{\text{rel}}$ to 10^{-4} and 10^{-6}, respectively. We run AAPM for maximum 1000 iterations.

Cai This Bayesian reconstruction approach solves a minimization problem with a non-quadratic cost function using a monotone conjugate gradient algorithm with heuristic descent steps [52]. We perform 5000 iterations to find the solution.

Long This is a regularized approach that uses Separable Quadratic Surrogates to minimize Kullback-Leibler cost function with edge-preserving regularization [172]. We run this method with 5000 iterations.

Weidinger This approach is similar to Long with modification in the approximation of regularization function by the Green potential, and leaving out Ordered Subsets that Long uses to speed up convergence [289]. We run the algorithm for 5000 iterations.

Mechlem This approach builds upon Weidinger while replacing the regularization using the Huber function. However, it uses Ordered Subsets and Nesterov acceleration to find the solution [189]. We run the algorithm for 200 iterations.

Barber This approach solves a constrained optimization problem using a primal-dual algorithm where constraints are composition of total-variation and simplex [27]. We run a maximum of 10000 primal-dual iterations.

	Spatial prior information	Spectral prior information
RU	Non-negativity	Non-negativity
UR	Non-negativity	Non-negativity
cJoint	Non-negativity	Non-negativity
Cai	(Modified) TV	Spectral signatures of present materials
Long	Simplex and (modified) TV	Spectral signatures of present materials
Weidinger	(Weak) non-negativity and (modified) TV	Spectral signatures of present materials
Mechlem	(Modified) TV	Spectral signatures of present materials
Barber	Constrained TV	Spectral signatures of present materials
ADJUST	Simplex	Spectral dictionary

Table 5.2: Spatial and spectral information for each method. TV stands for Total-Variation regularization [241]. Note that each method has prior information on the number of materials. The spectral dictionary contains signatures of a large superset of present materials.

5.6.2 Comparison of ADJUST with other methods

Results on the Mory phantom

To compare the ADJUST method with the other approaches listed in Section 5.6.1, we perform numerical studies on the Mory phantom. We take projections from 363 equidistant angles in $[0, \pi)$ using 181 detectors. Moreover, we apply Poisson noise to the resulting projections, with the incident photons being proportional to the source spectrum. We show the reconstruction results in Figure 5.7, and tabulate the performance measures in Table 5.3. We use Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) to assess the reconstruction results with respect to the ground truth phantom (details of the measures are given in Appendix C.6). We list the results as the measures averaged over all material maps. We observe that ADJUST obtains the best values for PSNR, SSIM and MSE.

The two-step methods (RU and UR) find the iodine and the gadolinium locations, but the results appear to be spatially smeared out, which is also reflected in the relatively low PSNR and MSE values. However, the UR method still recovers the shapes reasonably well, as shown in the high value for the SSIM. The cJoint method finds the water location, but fails to recover the locations of iodine and gadolinium. The state-of-the-art one-step methods overall perform better than UR, RU, and cJoint. However, these methods gives rise to edge artefacts (e.g. water edges partially appearing in iodine or gadolinium maps, and blurring near the edges), resulting in suboptimal PSNR and SSIM values. These artefacts hardly

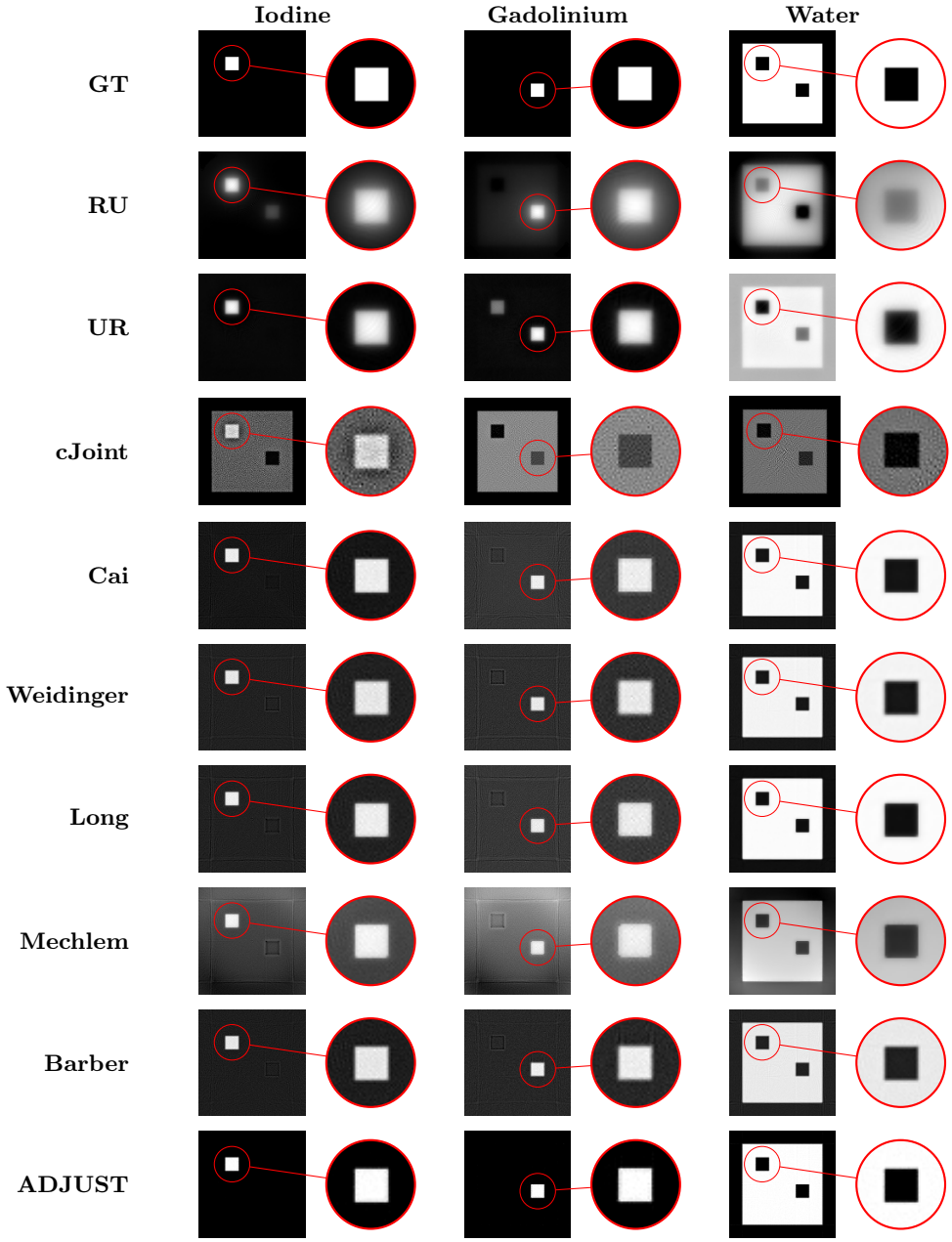


Figure 5.7: Visual comparison of various methods on Mory phantom. GT refers to ground truth.

appear in ADJUST reconstructions, resulting in significantly higher values of the PSNR and SSIM.

RU, UR and cJoint method do not incorporate any prior information about the object other than non-negativity constraints. Hence, these methods struggle to find the optimal solution. However, one-step methods strongly assume the knowledge of materials present in the object. Although they perform better than RU, UR and cJoint, they cannot be applied when the composition of object is not known. In contrast, ADJUST does not know about the material composition of the object. It solely relies on the number of materials present in the object (this number can be estimated through trial and error).

	RU	UR	cJoint	Cai	Weidinger	Long	Mechlem	Barber	ADJUST
MSE	0.0267	0.0790	0.0860	0.0115	0.0117	0.0112	0.0406	0.0115	**0.0103**
PSNR	20.35	16.44	16.03	19.54	19.33	19.62	14.45	19.41	**21.94**
SSIM	0.4413	0.6720	0.4119	0.6034	0.6131	0.6187	0.2159	0.6460	**0.9616**

Table 5.3: Reconstruction error in terms of Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) of various methods for the Mory phantom.

Results on the remaining phantoms

We compare the proposed method ADJUST with RU, UR, and Joint method on the Shepp-Logan, Disks and Thorax phantoms. Comparison with the other methods is not possible as those are designed for handling only a limited number of materials. The first two phantoms consist of only hard materials, and hence K-edges are present in the spectra. All these phantoms are more advanced compared to the Mory phantom. The Shepp-Logan and Thorax phantoms are structurally more complicated. On other hand, the Disks phantom contains up to eight hard materials. For the Thorax phantom, we also include soft materials. Nevertheless, we aim to reconstruct the bone, the iodine-blood mixture and the remaining soft materials into three separate classes. For all three phantoms, we measure tomographic projections for 180 equidistant angles from 0 to π. These measurements consist of Poisson noise that is proportional to the incoming photons on the detector. We tabulate the measures on the solutions produced by the RU, UR, cJoint and ADJUST algorithms for the three phantoms in Table 5.4. Moreover, for the Thorax phantom, we show the reconstructed material maps and the recovered spectra in Figure 5.8.

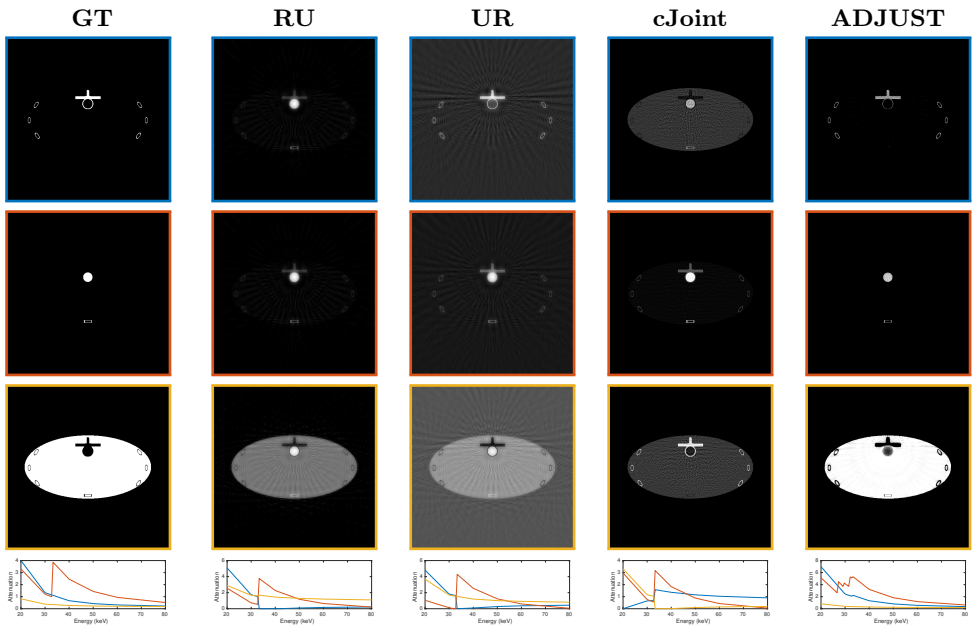


Figure 5.8: Visual comparison of ADJUST with RU, UR, and cJoint method on the Thorax phantom. The top row shows the ground truth and reconstructed material maps and spectral signatures for bone, the second row shows these for iodine and the third row shows (averages of) these for soft materials. We match the colors of the bounding box for material maps with the (recovered) spectral signatures shown on the bottom row.

Phantom		RU	UR	cJoint	ADJUST
Shepp-Logan	MSE	0.0711	0.0598	0.0548	**0.0061**
	PSNR	16.41	16.66	13.74	**23.12**
	SSIM	0.2433	0.4497	0.1077	**0.9599**
Disks	MSE	0.0125	0.0082	0.0063	**0.0030**
	PSNR	19.50	21.22	23.72	**33.32**
	SSIM	0.8970	0.8975	0.8882	**0.9925**
Thorax	MSE	0.0587	0.0622	0.0525	**0.0020**
	PSNR	19.92	20.96	19.11	**36.68**
	SSIM	0.6902	0.6094	0.7628	**0.9198**

Table 5.4: Reconstruction error in terms of MSE, PSNR and SSIM of various methods for various phantoms.

We see that for all measures, ADJUST outperforms the other three methods. The results for the other methods are similar to each other. In general, the Disks phantom leads to very good similarity measures for ADJUST, indicating the capability of dealing with eight materials. For the other two phantoms, where the proportions of materials are different, the measures are slightly worse, but we observe that the reconstructions for the Shepp-Logan and Disks phantoms are visually satisfactory (the visual results for the Disks phantom are given in Appendix C.7, while visual results for the Shepp-Logan phantom are given in Figure 5.3). However, for the Thorax phantom we see a striped pattern in the material map of the soft tissues. A possible reason is that the spectral signature of the combined tissue materials may not be present in the dictionary, and therefore produces visually suboptimal results. On the other hand, it can also be observed that the spectrum of bone is not fully correctly recovered. This may be because it does not have a discontinuity in the chosen spectral range, and is therefore too similar in shape to the tissue spectra. Despite this, ADJUST outperforms the other three methods on all phantoms.

5.6.3 Limited measurement patterns

Through the numerical experiments in this section, we demonstrate that ADJUST is robust. We consider three scenarios: (*i*) *Sparse-angle tomography*, where the number of measurements is reduced by sampling fewer projections angles, (*ii*) *Limited-view tomography*, where measurements from a particular range of angles are missing (representing the case of hardware limitations), and (*iii*) *Sparse channels*, where the number of spectral bins of the detector is limited. We apply these settings to the Shepp-Logan phantom and the Disks phantom, and report the results in Table 5.5.

For the sparse-angle tomography setup, we consider tomographic projections from 10 equidistant angles in the range of 0 to π. For the Shepp-Logan phantom, the spectral signatures are determined correctly and we observed very minor artefacts. For the Disks phantom, the material maps and the spectral signatures are precisely reconstructed, as reflected in the very low MSE and very high PSNR and SSIM values. Therefore, for these phantoms and the selected angles, ADJUST performs well.

For the limited-view tomography setup, we limit the projection angle range from 0 to $2\pi/3$. Restricting the angle range for the projections results in a well-known missing-wedge artefact. We take projections for 60 equidistant projection angles and add Poisson noise. For both phantoms, we observed no missing wedge artefacts when reconstructed with ADJUST. The PSNR and SSIM measures remain very high and the MSE measure remains low. We conclude that for these phantoms, ADJUST can deal well limited-view measurements.

In the sparse channel setting, we reduce the number of spectral bins based on the spectral dictionary. Since we consider a dictionary of 42 hard materials, we

Phantom		Full sampling	Sparse-angle	Limited-view	Sparse channels
Shepp-Logan	MSE	0.0061	0.0113	0.0330	0.0066
	PSNR	23.12	20.27	18.41	23.04
	SSIM	0.9599	0.9435	0.9112	0.9670
Disks	MSE	0.0030	0.0028	0.0057	0.0002
	PSNR	33.32	33.05	26.31	36.76
	SSIM	0.9925	0.9924	0.9807	0.9929

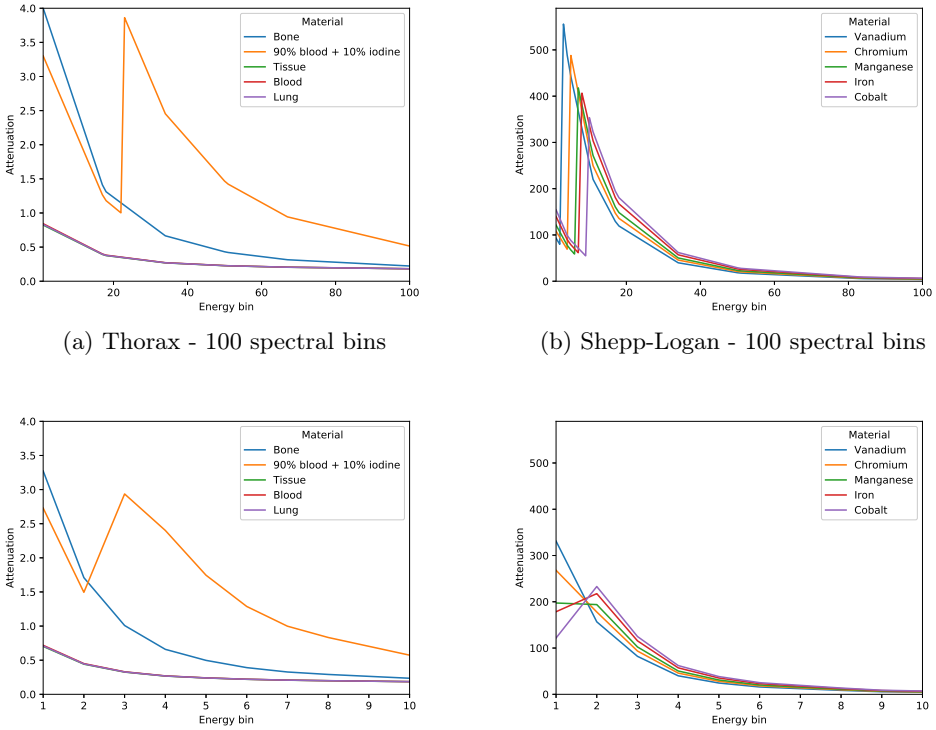
Table 5.5: Reconstruction error in terms of MSE, PSNR, SSIM with ADJUST for various phantoms with limited measurement pattern experiments.

reduce the spectral channels from 100 to 42. These 42 spectral channels are chosen based on the independent columns of the spectral dictionary. For tomography, we choose 60 equidistant angles between 0 and π. We observed that the Shepp-Logan phantom has been reconstructed precisely, but a few artefacts are visible on the edges of the disks for the Disks phantom. These artefacts are reflected in the slightly higher MSE and slightly lower PSNR and SSIM. For the Shepp-Logan phantom there is no obvious decrease of quality in terms of the measurements. So ADJUST appears to be capable of dealing with a sparse channel setting with the given phantoms and the spectral setup.

5.6.4 Limited spectral resolution

In this section, we investigate the performance of the algorithm when each spectral bin spans a wider energy range. This problem is also called *limited spectral resolution*. To this end, we have simulated data with 10 spectral bins instead of 100 bins for the Thorax and Shepp-Logan phantoms over the same energy range as before. The selected number of energy bins is in the same order as the number of bins resulting from the use of multi-spectral X-ray photon-counting detectors, rather than with the hundreds of energy bins of hyperspectral X-ray detectors. This results in a coarser energy resolution for both the source spectra and the attenuation spectra, and the dictionary has been updated accordingly. Apart from the different energy bins, we apply the same settings and configuration as the experiments in Section 5.6.1.

To illustrate the spectral differences, Figure 5.9 shows the attenuation spectra for the materials in the phantoms with the two different energy resolutions. The soft materials, bone and iodine present in the Thorax phantom are easily identifiable since their attenuation spectra are very different from each other in the entire spectrum (*i.e.* $20 - 80$ keV). The materials in the Thorax phantom will still be separable with decrease in the spectral resolution. In the Shepp-Logan phantom, however, all included materials have K-edges very close to each other. Hence, differences resulting from the K-edge of the materials in the Shepp-Logan phantom



(a) Thorax - 100 spectral bins

(b) Shepp-Logan - 100 spectral bins

(c) Thorax - 10 spectral bins

(d) Shepp-Logan - 10 spectral bins

Figure 5.9: Attenuation spectra of different materials in the Thorax phantom (left) and the Shepp-Logan phantom right (right), partitioned over 100 energy bins (top) and 10 energy bins (bottom) over the same energy range.

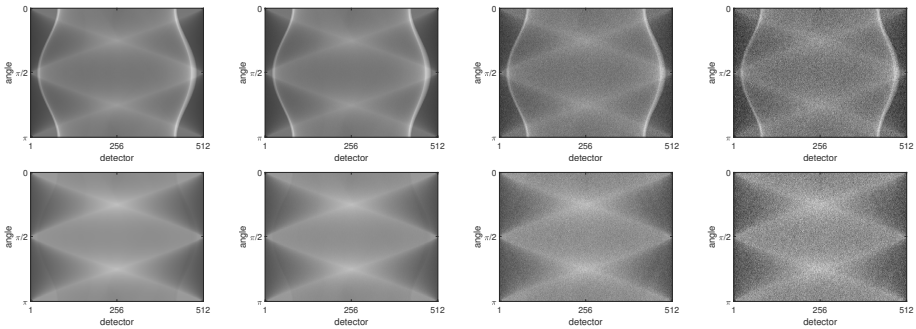
Phantom		RU	UR	cJoint	ADJUST
Shepp-Logan	MSE	0.1446	0.1856	0.1980	**0.0976**
	PSNR	9.042	9.392	7.259	**10.94**
	SSIM	0.0213	0.3773	0.0209	**0.6228**
Thorax	MSE	0.0543	0.0715	0.0533	**0.0010**
	PSNR	21.22	19.64	17.95	**33.05**
	SSIM	0.6820	0.6579	0.7608	**0.9707**

Table 5.6: Reconstruction error in terms of MSE, PSNR, SSIM for various phantoms with 10 times smaller spectral resolution (10 spectral bins in total).

largely disappear. This is also reflected in the performance of various algorithms in Table 5.6, with considerably lower SSIM and PSNR compared to the full spectral resolution experiments for the Shepp-Logan phantom in Table 5.4. On the other hand, for the Thorax phantom the results are similar. Hence, we conclude that the successful application of the proposed algorithm to spectral measurements acquired in bins with larger spectral ranges depends heavily on the complexity of the problem (*i.e.* separability of the materials in the chosen energy range based on spectral responses in the bins).

5.6.5 Noisy measurement patterns

We consider three different noise levels to check the robustness of ADJUST against noise. In particular, we corrupt the spectral tomographic measurements with additive Gaussian noise of strength $\{1, 10, 20\}\%$ followed by Poisson noise with intensity corresponding to the source spectrum. In all three cases, 180 angular projections in the $[0, \pi)$ range are acquired in 100 spectral bins. The effect of noise on the spectral tomographic projections is demonstrated in Figure 5.10. In Table 5.7, we list the performance measures of ADJUST on three different phantoms. We observe a steady decrease in PSNR and SSIM values for the Shepp-Logan and Thorax phantoms. However, the PSNR and SSIM do not suffer from high noise levels (*i.e.* 20% noise) for the Disks phantom. Since the Disks phantom consists of low-rank shapes of small sizes, they can be retrieved well from noisy measurements. However, the performance of ADJUST on noisy datasets may not be extended for complex shapes, as seen from the numerical studies on the Thorax and Shepp-Logan phantom. Hence, these numerical experiments demonstrate that ADJUST is stable against a moderate amount of noise, but may not be reliable against high noise levels.



(a) Clean (b) 1% noise (c) 10% noise (d) 20% noise

Figure 5.10: Clean and noisy sinograms (*i.e.* tomographic projections) of Shepp-Logan phantom for energy of 5.6 keV (top-row) and 7.7 keV (bottom-row).

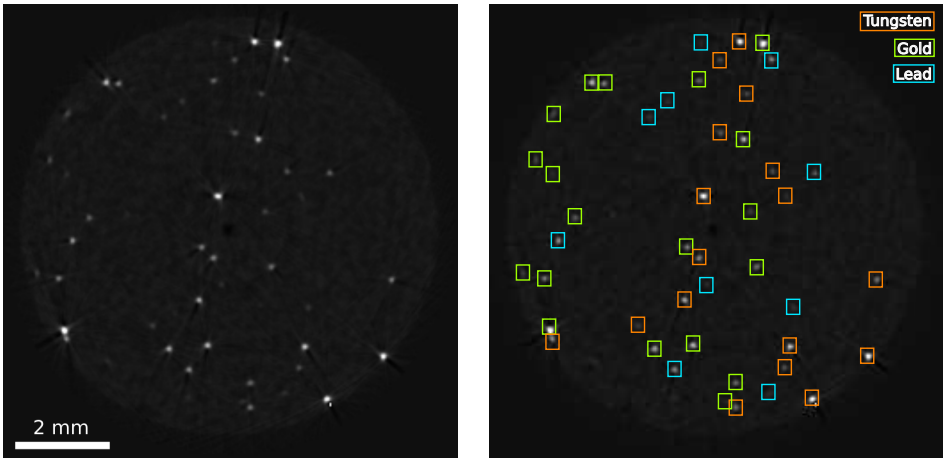
Phantom		1% noise	10% noise	20% noise
Shepp-Logan	MSE	0.0032	0.0067	0.0187
	PSNR	25.68	22.53	18.68
	SSIM	0.9738	0.9012	0.7076
Disks	MSE	0.0001	0.0034	0.0003
	PSNR	39.20	32.44	35.53
	SSIM	0.9989	0.9779	0.9785
Thorax	MSE	0.0061	0.0129	0.0123
	PSNR	29.70	25.93	23.33
	SSIM	0.8716	0.8474	0.8464

Table 5.7: Reconstruction error in terms of MSE, PSNR, SSIM for various phantoms with three different noise levels.

5.6.6 Experiment on micro-CT data

In this subsection, we test the performance of ADJUST on a publicly available X-ray microtomography dataset [262] generated by a conventional laboratory-based CT scanner equipped with a photon-counting line detector (TESCAN Polydet) containing a semiconductor crystal (CdTe) [263]. The particle mixture sample contains pure gold, tungsten and lead, along with quartz. For a total of 128 energy bins with a spectral range of ca. 20 to 160 keV, tomographic projections are acquired on 255 equally-spaced detector pixels for 600 angles in $[0, 2\pi)$. We perform a reconstruction using the simultaneous iterative reconstruction algorithm (SIRT) [140] for each channel and determine the location of each material manually (refer to Figure 5.11). We plot the gold, lead, and tungsten spectrum obtained through channel-wise reconstructions along with the corresponding NIST spectra in Figure 5.12.

For a fair comparison, we apply RU, UR, and cJoint along with ADJUST on this spectral microtomography dataset. We use the same settings as mentioned in Section 5.6.1. For ADJUST, we use the spectral dictionary with four materials (*i.e.* gold, lead, tungsten and quartz) with their spectrum obtained from NIST (the spectrum of quartz is obtained from the reconstructed channels). Since it is evident from Figure 5.12 that the NIST spectra with appropriate scaling match closely to the spectra in the sample in the range of [53.35,127.05] keV, we reduce the spectral range of the dataset from 53.35 to 127.05 keV (amounting to 64 spectral bins in total). The reconstruction results are demonstrated in Figure 5.13. From these results, it is clear that RU, UR, and cJoint cannot precisely reconstruct the spectral signatures of materials. These results, furthermore, suggest the need of stronger spectral regularization to separate the materials. Out of all the spatial maps reconstructed with these classical methods, only the spatial map of tungsten (red) produced by the RU method approximately matches the expected material map (shown in Figure 5.11). On the other hand, the spatial maps recovered from



(a) (b)

Figure 5.11: CT reconstruction of the 7th energy bin (**a**) of the spectral micro-CT dataset, in which the locations of the tungsten (orange), gold (green) and lead (blue) particles are highlighted (**b**).

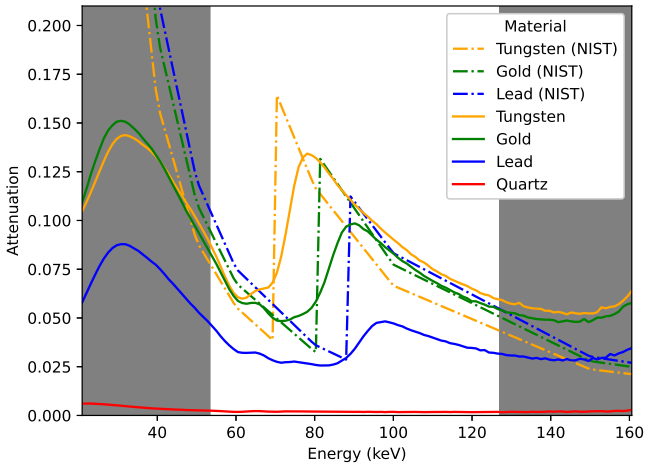


Figure 5.12: Attenuation spectra of the materials in the sample, showing both the spectra from NIST (scaled by 0.015) and the spectra extracted from the channel-wise CT reconstructions, including quartz. The spectral range is from 20.35 keV to 161.15 keV (white region), with the bins having a spectral resolution of 1.1 keV.

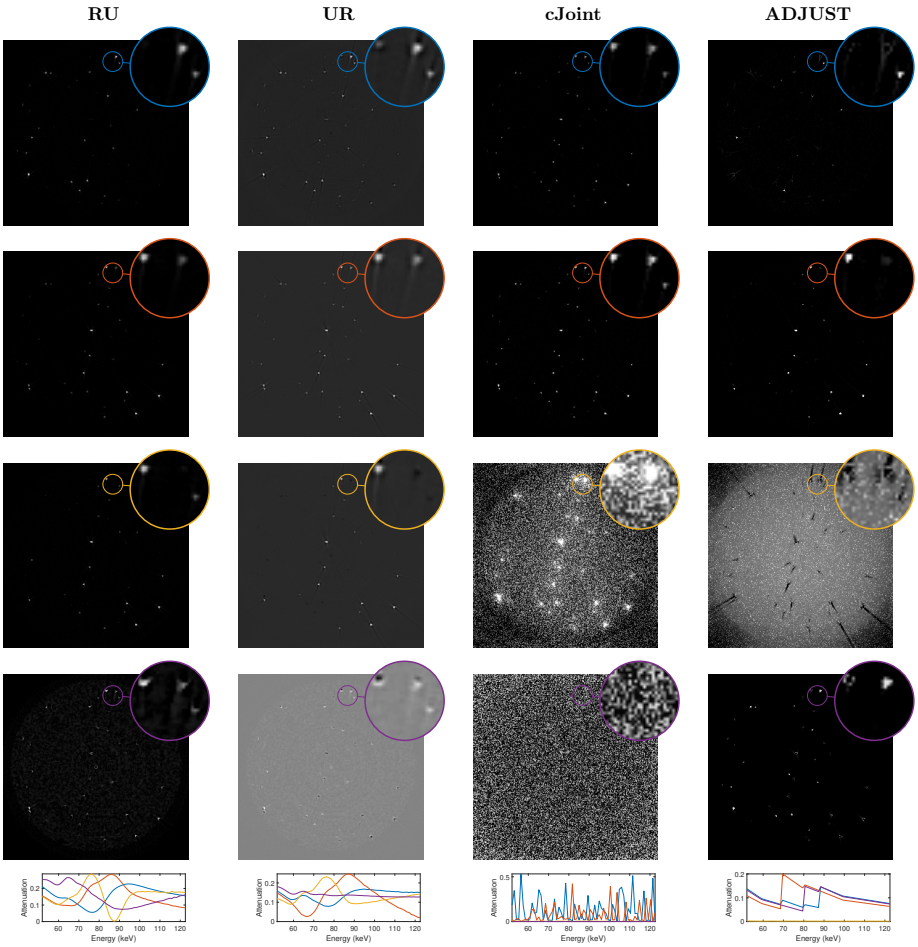


Figure 5.13: Reconstruction results of methods RU, UR, cJoint, and ADJUST on the microtomography dataset of particle mixtures. The zoomed sections demonstrate the separation capabilities of ADJUST compared to RU, UR and cJoint on particles of three different materials (tungsten, gold and lead). We match the color of zoomed sections with the spectral plots for improved readability.

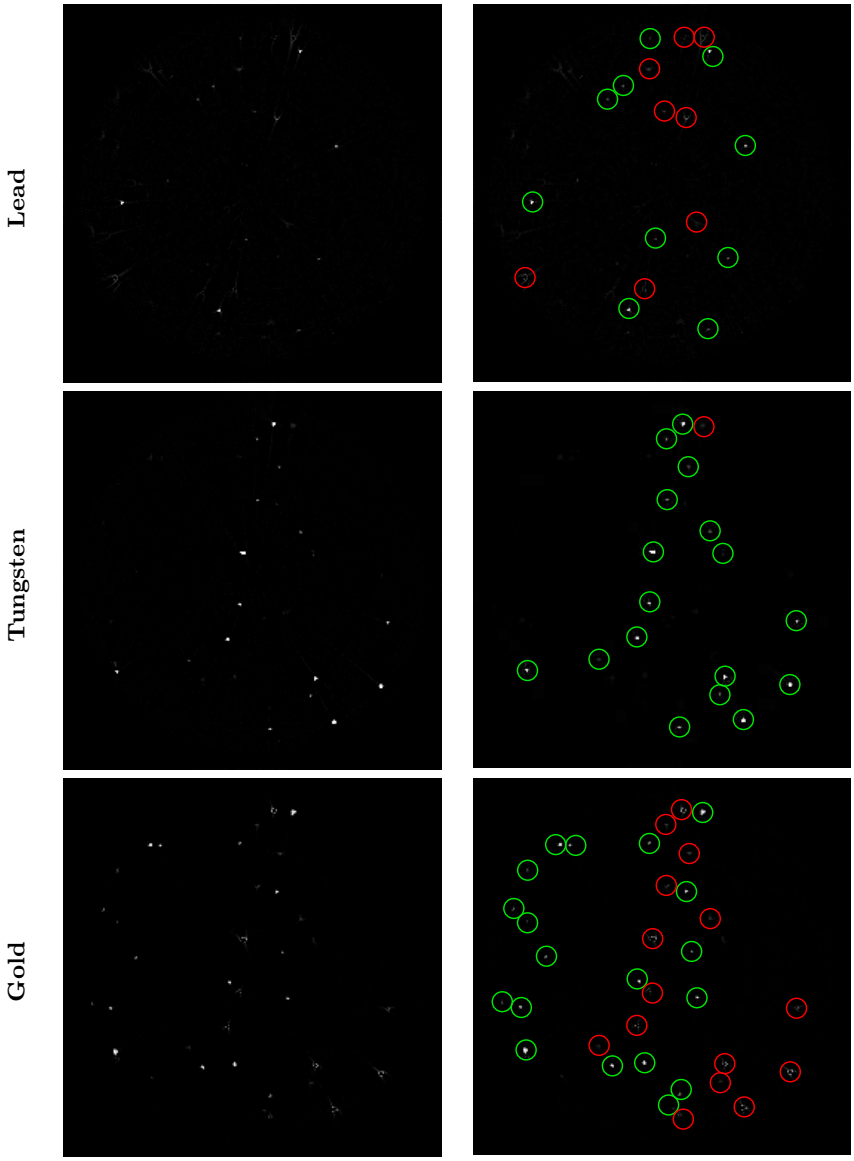


Figure 5.14: Detailed reconstruction results of ADJUST on X-ray microtomography dataset of particle mixtures. Reconstructed material maps for lead, tungsten and gold are shown in high resolution (left column). In addition, material maps with indications of identified particles are shown (right column). Green circles indicate correctly identified particles. None of the maps contain false negatives, and the remaining identified particles are false positives (denoted by red circles).

ADJUST match almost precisely with the expected material maps. Figure 5.14 shows that all particles are identified and that there are no false negatives. False positives are mostly small and faint, and many of these occur at the borders of other particles, mainly between gold and tungsten. As reference, Figure 5.14 also shows the reconstruction images for each material in high resolution. Finally, we note that the NIST spectra do not overlap entirely with the spectra recovered through channel-wise reconstruction. Therefore, the spectral dictionary must be calibrated for the spectral detector setup for various materials. We expect improvement in reconstruction results of ADJUST with a calibrated spectral dictionary.

5.7 Conclusions and discussion

Spectral imaging is an emerging topic in X-ray tomography since it adds an additional dimension to the measurements, which can be exploited to retrieve the material composition of the object of interest. Recently, joint approaches (also known as one-step methods) have emerged as a promising technique for solving the spectral imaging problem by incorporating all the prior information in a single step. These joint approaches reduce the ill-posedness of the spectral imaging problem. However, the spectral signatures of many materials are very similar, making the joint approaches likely to fail when many materials are involved. To tackle the problems with the joint approach, we propose the ADJUST framework that integrates the structure of spectral signatures by creating a dictionary of all the known materials and uses this to jointly reconstruct and carry out material decomposition in a single step. Since the resulting formulation is a bi-convex optimization problem, we propose an Alternating Accelerated Proximal Minimization (AAPM) scheme to find a solution. Through numerical experiments, we show that ADJUST performs better over practical methods as well as state-of-the-art joint approaches on various simulated phantoms.

Obtaining projections from all directions requires high experimental time. Moreover, X-ray machines may not allow for sampling in all directions. Hence, practical methods do not help in determining the material composition of structurally more complicated objects (either because of intricate structures or a wider variety of materials) with most X-ray configurations. However, ADJUST is robust against limited tomography measurement patterns on phantoms that are more complicated.

A natural question is to check if the utilization of a spectral dictionary (*i.e.* representing $\boldsymbol{F}$ as $\boldsymbol{RT}$) in two-step methods can produce optimal results. For example, in UR method, we can first decompose spectral measurements $\boldsymbol{Y} = \boldsymbol{ZRT}$, where $\boldsymbol{Z}$ incorporates the projections per material. Later, the spatial maps can be extracted by solving $\boldsymbol{WA} = \boldsymbol{Z}$. However, in the decomposition step, we allow for unrealistic projections due to lack of knowledge of tomography operator $\boldsymbol{W}$ and spatial properties of $\boldsymbol{A}$, especially when dealing with limited measurements. Moreover, allowing unrealistic projections leads to unrealistic materials since simplex

constraints on $\boldsymbol{R}$ give rise to convex combination of (many) dictionary elements, and not of only a single element. A similar argument holds for RU method combined with the spectral dictionary. These unrealistic solutions are, however, penalized by proper *spatio-spectral regularization* in the ADJUST framework. Hence, ADJUST will always perform better than spectral dictionary versions of two-step methods even when complete measurements are available.

There are some limitations to the ADJUST framework. It can only separate hard materials from each other and separate hard materials from soft materials. Moreover, if neither the spectral signature of the material is present in the dictionary, nor it can be composed as a linear combination of the elements from the dictionary, ADJUST will fail in recovering that material. Although we tested ADJUST against moderate Gaussian and Poisson noise, it is not straightforward to assume that ADJUST will behave stably against real (extremely) noisy datasets that are common in energy-dispersive X-ray tomography.

The experiments conducted on the micro-CT dataset suggest the potential of ADJUST for application to experimental spectral CT data. Apart from the noise level on experimental data, the successful application of ADJUST depends on several factors. As shown in the experiments with the limited spectral resolution, the number and range of the energy bins, combined with the problem's complexity, determines the degree of possible material decomposition. If the spectral resolution is too small such that materials cannot be distinguished anymore, ADJUST will not yield proper material decompositions (and neither will other methods). Moreover, for successful reconstruction and decomposition, there is a limit to the number of (differentiable) materials that can be included in the dictionary for a given spectral resolution.

For an uncalibrated spectral detector, the spectral dictionary can be measured, if it is not available through the manufacturer. Additionally, when working with real data, an interesting consideration for future work would be to estimate the source spectrum along with the spatial material maps and their signatures. Although our framework is based on the assumption that the spectral tomographic measurements consist of additive white noise, we can extend it to tackle Poisson noise by replacing the least-squares loss by Kullback-Leibler function. However, the bi-convexity can no longer be guaranteed and the solution obtained through AAPM may not be partially optimal. We leave this extension for future work.

Code and data availability

The source code of ADJUST, along with the RU, UR, and cJoint algorithms, are available on `https://github.com/mzeegers/ADJUST` [138]. These MATLAB codes make use of open-source toolboxes, in particular the ASTRA toolbox [1, 2], Spot Operator toolbox [37], MinConf optimization package [250]. The scripts for

testing the mentioned algorithms on the phantoms and the data are also made open-source in the ADJUST Github repository. The micro-CT spectral dataset is available at `https://rodare.hzdr.de/record/1627` [262].

6 Conclusions and outlook

6.1 Conclusions

6.1.1 Overview

Quality control is a challenging but essential procedure in industrial imaging. With X-ray imaging, the inner structure of an object can be visualized and hence, unwanted and potentially hazardous elements can be identified by detection systems. Since single X-ray radiography results in superimposed images, extracting and analyzing features from these images may lead to suboptimal decision-making. Recent years have seen a significant increase in tomographic imaging quality, especially resulting from spectral imaging and machine learning technologies. These techniques have the potential to significantly increase the effectiveness of industrial imaging without sacrificing too much on processing speeds.

As one of the most important branches of industry, food processing is the recurring theme in this dissertation. In particular, foreign object detection is an essential component that should attain both high accuracy and high throughput. In order to improve this trade-off, X-ray imaging can be used in conjunction with machine learning. **Chapter 2** addresses a common issue with machine learning regarding the need for large volumes of suitable training data. By incorporating computed tomography into a workflow, high-quality pairings of X-ray projections and ground truth locations of foreign objects can be generated with minimal manual labor. The resulting datasets enable the training of neural networks for high-speed foreign object detection.

By enhancing the features for analysis of industrial products, spectral X-ray imaging offers an improvement to standard X-ray inspection. The additional information spectral X-ray imaging offers also results in much larger data volumes, of which a significant fraction may be redundant. In **Chapter 3**, a network ar-

chitecture named Data Reduction Convolutional Neural Network (DRCNN) is proposed to mitigate this problem. With this architecture, a neural network learns to combine the image features that are needed for a specific task (foreign object detection, for example) and compresses these into a much smaller data volume. This approach increases the quality of feature extraction and reduces data sizes and possibly processing times.

Spectral X-ray imaging also offers improvement in the three-dimensional reconstruction of products of interest. In spectral X-ray CT, reconstructions can be computed with projection data probed at different X-ray photon energies. These different energy levels are treated in a more generalized manner as channels in **Chapter 4**. There, a class of algorithms named Multi-Channel DART (MCDART) is proposed that generalizes the Discrete Algebraic Reconstruction Technique (DART) – for objects consisting of a limited number of materials with known attenuation values – to multi-channel data. From a series of experiments, it can be concluded that this class of algorithms can improve reconstructions using multi-channel data.

When an industrial spectral X-ray setup is available, attenuation values of a material can be obtained by directly measuring these with the spectral detector. By doing this for many common materials, a spectral dictionary can be constructed, which can be used as prior information to steer the often ill-posed spectral reconstruction problem to a desirable solution. **Chapter 5** proposes a spectral reconstruction and material decomposition framework, named ADJUST, by posing the problem in such a way that the spectral matrix is a multiplication of an indicator matrix and a spectral dictionary. Contrary to most other spectral material decomposition methods, ADJUST is a method that performs the reconstruction and material decomposition in one step. Moreover, ADJUST can take on objects with more materials and produce more accurate reconstructions than other methods.

6.2 Contributions

The methods presented in this dissertation utilize spectral imaging and deep learning to improve aspects such as the workload, practicality, accuracies, ill-posedness and compression possibilities in problems found in X-ray imaging and CT reconstruction applications. The earlier two sections are concerned with machine learning methods, which are shown to yield improved results for industrial foreign object detection. The latter two sections are concerned with improving the accuracy of CT reconstructions by utilizing spectral X-ray imaging, both with respect to single-channel CT reconstruction as well as to other spectral CT reconstruction methods.

Deep learning in conjunction with X-ray imaging methods has rarely been applied in certain industrial areas, such as food processing. The workflow presented in **Chapter 2** for efficiently generating training data for deep learning on radiographs has application potential in an industrial setting, as evidenced by the accuracy results on the real-world dataset. By making use of computed tomography, foreign objects can be represented in a 3D volumetric space rather than a 2D radiograph. This makes separation of foreign objects from the remaining object substantially easier. As a result, the annotation of the X-ray projections becomes radically less time-intensive and less prone to interpretation. On top of this, much more training data can be generated from each scanned product, implying that the number of products that need to be scanned is limited. Regarding improving quality control by object inspection at a production line of a factory, this contributes significantly to the workload of setting up a deep learning driven analysis and decision-making machinery in an industrial context. Additionally, Chapter 2 also provides an open X-ray dataset, which is generally not easy to come by, to test methods for object detection.

The DRCNN architecture covered in **Chapter 3** applies machine learning to spectral data reduction. As suggested by the results in that chapter, when hyperspectral imaging is used for industrial X-ray imaging, training with DRCNN can be used to optimize the compression and throughput speed for a specific task, such as detection tasks. In addition, training with this architecture achieves better understanding of essential features in the data and can possibly speed up the (hyper)spectral X-ray data acquisition. This approach is helpful for all applications where hyperspectral imaging – not necessarily with X-rays – is concerned. For instance, in optical hyperspectral imaging applications such as remote sensing or surface-based hyperspectral inspection, the method contributes to better compression, transmission, speed and accuracy.

Similar to how the workflow in Chapter 2 exposes the advantages of using 3D CT with respect to separating foreign objects, the MC-DART method in **Chapter 4** exploits the improved separability of voxels in a reconstruction through multi-channel imaging. By incorporating a high-dimensional segmentation step, the method presents a framework for reconstruction of multi-channel data in a more effective manner than independent channel reconstructions. The method can be scaled to any number of channels, and therefore allows for a higher number of materials in an object. The usage and properties of the MC-DART method are demonstrated using simulated spectral X-ray imaging mechanisms. However, MC-DART can also be used for other multi-channel modalities, given that the object of reconstruction can be represented in a discrete manner.

When prior information on spectral material signatures is available in the form of a spectral dictionary, the ADJUST method proposed in **Chapter 5** can produce material maps of an object from its spectral X-ray CT data, given that the

corresponding materials are sufficiently distinct in terms of their signatures. The method allows for higher accuracies and for reconstructions of more materials than in most existing one- or two-step methods without additional hyperparameters. The results on a laboratory spectral micro-CT dataset imply that applications are possible in both spectral and hyperspectral X-ray imaging.

6.3 Future work and outlook

Spectral imaging and deep learning hold much potential for solving advanced imaging problems. To utilize these advances effectively, algorithms that can adapt to a wide range of problem settings should be developed. A running theme in the research of this dissertation is the modularity of the proposed methods. The DRCNN architecture is designed in such a manner that any CNN architecture can be inserted in the back-end of the network, of which the MSD and U-Net architectures are well-investigated examples in Chapter 3. Similarly, in the class of algorithms that the MC-DART method represents, the algebraic reconstruction method can be chosen at will. Finally, the workflow presented in Chapter 2 has many possibilities for changes and extensions. A number of these are schematically shown in Figure 6.1.

As highly suggested in this dissertation, the workflow from Chapter 2 can be enhanced with spectral X-ray imaging. This can be done in at least two ways: (*i*) spectral scanning can increase the quality of the CT reconstruction and the subsequent segmentations and virtual projections to obtain better ground truth (Fig. 6.1a), and (*ii*) spectral input data can be used in the generated training set for better deep learning training and performance (Fig. 6.1d). With spectral CT data, a naive approach of reconstructing each separate channel can be carried out in order to improve the foreign object segmentation. Even better is to incorporate existing algorithms to use the combined information to make a (discrete) reconstruction (Fig. 6.1f), for which integration of the MC-DART (Fig. 6.1g) and the ADJUST (Fig. 6.1h) algorithms are suitable. In the case of hyperspectral or multi-spectral data with a high number of bins, the DRCNN architecture can extract essential features and reduce the spectral dimension of the data to speed up the throughput in the neural network (Fig. 6.1i).

Other interesting additional extensions of the workflow include multi-angle data acquisition (often used, for example, for glass container inspection [68]), which, along with multi-energy acquisition, still needs to be fully used in X-ray imaging for detection [8]. Another improvement is to include a separate high-quality scan (Fig. 6.1b) or advanced CT reconstructions for improved segmentation accuracy. For instance, if needed, denoising or inpainting can be applied to the reconstructed CT volumes (Fig. 6.1e). Deep-learning driven noise reduction algorithms [119] and inpainting algorithms are readily available for these purposes. Finally, to further improve the effectivity of the detection by deep learning, more data augmentation

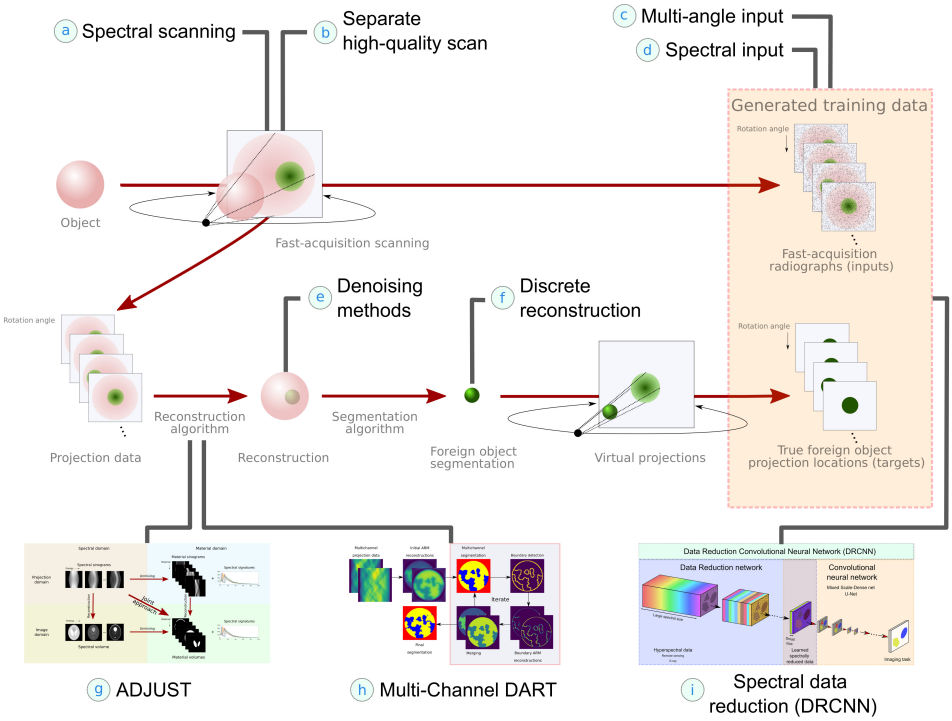


Figure 6.1: The complete workflow of data acquisition from **Chapter 2** with a number of possible extensions and enhancements: (**a**) spectral scanning, (**b**) performing an additional higher-quality scan for improved CT reconstruction, (**c**) multi-angle or (**d**) spectral input for the neural network, (**e**) application of denoising methods to improve the CT reconstruction, (**f**) direct discrete reconstruction resulting in segmented volumes, (**g**) integration of ADJUST for spectral tomography, (**h**) integration of MC-DART for multi-channel data, (**i**) integration of DRCNN in the workflow.

can be carried out in the projection domain [145]. Better still, data augmentation in the reconstruction space [17, 192], combined with realistic forward projections, yields even richer X-ray training datasets.

The evaluation of the methods presented in this dissertation focuses primarily on accuracy. While the deep learning methods are designed to be fast, the processing times are only briefly touched upon, either in terms of particular time speedups for DRCNN and time complexity classes with MC-DART. A deeper analysis of the processing speedups would be interesting, especially regarding the additional spectral dimension inherent to the data used in most of this dissertation. Nevertheless, this highly depends on the problem setting. For instance, if the X-ray imaging setup is such that the data quality deteriorates, more spectral bins may be needed for successful feature extraction with DRCNN, or more iterations are needed for proper convergence to acceptable solutions with MC-DART and ADJUST. Speed comparisons will become interesting when a practical application

and setup are considered. When the speed at a production line is high, motion blurring may become an issue, but we expect that this can be taken care of with machine learning as well. This also applies when many objects are located on the conveyor belt, and the resulting radiographs are composites of single-object radiographs. In this case, either reconstruction methods for overlapping objects [253] or data augmentation will help (although with too many objects, photon starvation could prevent any possible detection at all). Also, in our experiments, perfect (angular) alignment of the foreign object radiograph and the ground truth was not strictly needed. Some erroneous experiments revealed that a slight offset did not significantly reduce the detection rate. This is positive for industrial implementation as it may allow for some accidental misalignment. In general, the generalization of the trained neural networks with approaches in this dissertation to other data from other objects or factory line setups is a topic for further research.

With spectral imaging and spectral CT, there are at least a number of additional challenges. First, spectral detection may not be consistent. Initial laboratory experiments showed that detectors can yield different projection images when experiments are repeated under the same circumstances. Machine learning based object detection methods may overcome this problem to a certain extent, although this again depends on the separability of the materials in the object. Second, the modelling in this dissertation does not include the (not necessarily known) detector response function, which describes the distribution of measured energy for an incident beam of a specific energy. ADJUST still needs to include a detector response matrix in the modelling. In general, incorporation of a detector response function [170] makes the problem more realistic but is often left out to keep the problem convex [6].

The methods presented in this dissertation can be further enhanced. First, the workflow can be extended from segmentation to other sorts of classification, such as binary classification. Secondly, in the spectral simulations used for the workflow, DRCNN and ADJUST results, no densities are taken into account. Including this does not fundamentally change the methods, but they need to be added in the case of practical applications with ADJUST. Thirdly, an evident and straightforward extension to MC-DART is the application to three-dimensional reconstruction problems. Additionally, the requirement for the grey values to be known for MC-DART poses practical challenges. For that reason, a version with automatic grey level estimation is desirable, which can be implemented as a joint optimized problem in which both the reconstruction and the grey levels are optimized (in a similar fashion as TVR-DART [328]). Lastly, task-driven data reduction can be combined with network pruning methods (for example, LEAN [252]) to reduce training time, reduce processing time and improve feature extraction.

While the methods presented in this dissertation aim primarly for industrial imaging, their usage extends beyond this. First of all, the data reduction method and the material decomposition approach of ADJUST can be applied in remote sensing. ADJUST may be used as regularization, like, for instance, directional TV

in remote sensing [48]. Data reduction can be applied to all sorts of remote sensing problems, as access to data or equipment is getting less difficult and costly [109, 243]. The biconvex formulation and solution strategy of ADJUST can be used for other CT-related problems, for instance optimal angle selection or automatic grey value estimation in discrete tomography. In addition, ADJUST can be used for general material decomposition, for instance for biological and chemical contamination checks, or in spectroscopy.

The fields of machine learning, spectral X-ray imaging and tomography are rapidly evolving. It is expected that spectral detectors will become the standard for medical imaging in the future, and will be adopted in industrial imaging as well. Depending on the detector possibilities, spectral X-ray imaging may be used for chemical and biological contamination in addition to physical contamination. Deep learning gains momentum in industrial imaging in general and will most surely become a standard in food inspection. The combination of deep learning with spectral X-ray inspection is, therefore, a logical pathway to the future to increase accuracy and throughput. A few key limitations remain that make it how far these developments will push the possibilities. First, imaging and detection are limited to certain imaging resolution (for instance, with micro- and nanoplastics). Secondly, the difficulty of detecting of certain combinations of materials with X-rays will remain a sticking point. Therefore, additional means of noninvasive methods may be needed. For instance, phase-contrast imaging may be more interesting than absorption imaging in certain situations, as this modality yields superior results for the detection of organic materials [80].

Bibliography

[1] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers. "Fast and flexible X-ray tomography using the ASTRA toolbox". *Optics Express* 24.22 (2016), pp. 25129–25147 (cit. on pp. 47, 57, 63, 77, 93, 100, 107, 111, 131, 146).

[2] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers. "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography". *Ultramicroscopy* 157 (2015), pp. 35–47 (cit. on pp. 47, 57, 63, 77, 93, 107, 111, 131, 146).

[3] W. van Aarle, K. J. Batenburg, and J. Sijbers. "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)". *IEEE Transactions on Image Processing* 21.11 (2012), pp. 4608–4621 (cit. on p. 96).

[4] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. "TensorFlow: A system for large-scale machine learning". In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. (Savannah, GA, USA). 2016, pp. 265–283 (cit. on p. 35).

[5] J. F. P. J. Abascal, N. Ducros, V. Pronina, S. Rit, P. Rodesch, T. Broussaud, S. Bussod, P. C. Douek, A. Hauptmann, S. Arridge, et al. "Material decomposition in spectral CT using deep learning: A Sim2Real transfer approach". *IEEE Access* 9 (2021), pp. 25632–25647 (cit. on p. 119).

[6] J. F. P. J. Abascal, N. Ducros, and F. Peyrin. "Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance". *Inverse Problems* 34.12 (2018), p. 124003 (cit. on pp. 27, 154).

[7] A. Adler, M. Elad, and M. Zibulevsky. "Compressed learning: A deep neural network approach". *arXiv preprint arXiv:1610.09615* (2016) (cit. on p. 69).

[8] S. Akcay and T. Breckon. "Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging". *Pattern Recognition* 122 (2022), p. 108245 (cit. on pp. 35, 41, 152).

[9] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon. "Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery". *IEEE Transactions on Information Forensics and Security* 13.9 (2018), pp. 2203–2215 (cit. on p. 68).

[10] C. Albon. *Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning*. Sevastopol, CA, USA: O'Reilly Media, Inc., 2018 (cit. on p. 68).

[11] H. Alkadhi, A. Euler, D. Maintz, and D. Sahani. "Spectral imaging: Dual-energy, multi-energy and photon-counting CT" (2022) (cit. on p. 24).

[12] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari. "A state-of-the-art survey on deep learning theory and architectures". *Electronics* 8.3 (2019), p. 292 (cit. on p. 33).

[13] A. Alqahtani, X. Xie, and M. W. Jones. "Literature review of deep network compression". *Informatics* 8.4 (2021), p. 77 (cit. on p. 35).

[14] R. E. Alvarez and A. Macovski. "Energy-selective reconstructions in X-ray computerised tomography". *Physics in Medicine & Biology* 21.5 (1976), p. 733 (cit. on p. 26).

[15] A. H. Andersen and A. C. Kak. "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm". *Ultrasonic imaging* 6.1 (1984), pp. 81–94 (cit. on p. 17).

[16] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. "Unsupervised foreign object detection based on dual-energy absorptiometry in the food industry". *Journal of Imaging* 7.7 (2021), p. 10 (cit. on pp. 22, 40).

[17] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. "CT-based data generation for foreign object detection on a single X-ray projection". *Scientific Reports* 13.1 (2023), p. 1881 (cit. on p. 153).

[18] F. van Assche, S. Vanheule, L. van Hoorebeke, and M. N. Boone. "The spectral X-ray imaging data acquisition (SpeXIDAQ) framework". *Sensors* 21.2 (2021), p. 563 (cit. on p. 21).

[19] ASTM. "G173–03: Standard tables for reference solar spectral irradiances: Direct normal and hemispherical on 37° tilted surface". *ASTM International, West Conshohocken, PA* (2003) (cit. on p. 80).

[20] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality". *Mathematics of Operations Research* 35.2 (2010), pp. 438–457 (cit. on p. 125).

[21] N. Audebert, B. Le Saux, and S. Lefèvre. "Deep learning for classification of hyperspectral data: A comparative review". *IEEE Geoscience and Remote Sensing Magazine* 7.2 (2019), pp. 159–173 (cit. on p. 68).

[22] V. Badrinarayanan, A. Kendall, and R. Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017), pp. 2481–2495 (cit. on p. 43).

[23] E. J. Bakker, P. B. W. Schwering, and S. P. van den Broek. "From hyperspectral imaging to dedicated sensors". In: *Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process.* (Orlando, FL, USA). Vol. 4029. International Society for Optics and Photonics. 2000, pp. 312–324 (cit. on p. 66).

[24] R. Ballabriga, J. Alozy, F. Bandi, M. Campbell, N. Egidos, J. M. Fernandez-Tenllado, E. H. M. Heijne, I. Kremastiotis, X. Llopart, B. J. Madsen, et al. "Photon counting detectors for X-ray imaging with emphasis on CT". *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020) (cit. on p. 114).

[25] R. Ballabriga, J. Alozy, M. Campbell, E. Frojdh, E. Heijne, T. Koenig, X. Llopart, J. Marchal, D. Pennicard, T. Poikela, L. Tlustos, P. Valerio, W. Wong, and M. Zuber. "Review of hybrid pixel detector readout ASICs for spectroscopic X-ray imaging". *Journal of Instrumentation* 11.1 (2016), P01007 (cit. on p. 114).

[26] R. Ballabriga, M. Campbell, E. Heijne, X. Llopart, L. Tlustos, and W. Wong. "Medipix3: A 64 k pixel detector readout chip working in single photon counting mode with improved spectrometric performance". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 633 (2011), S15–S18 (cit. on p. 21).

[27] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan. "An algorithm for constrained one-step inversion of spectral CT data". *Physics in Medicine & Biology* 61.10 (2016), p. 3784 (cit. on pp. 26, 28, 119, 132).

[28] K. J. Batenburg, W. Fortes, L. Hajdu, and R. Tijdeman. "Bounds on the quality of reconstructed images in binary tomography". *Discrete Applied Mathematics* 161.15 (2013), pp. 2236–2251 (cit. on p. 102).

[29] K. J. Batenburg, W. J. Palenstijn, P. Balázs, and J. Sijbers. "Dynamic angle selection in binary tomography". *Computer Vision and Image Understanding* 117.4 (2013), pp. 306–318 (cit. on p. 18).

[30] K. J. Batenburg and J. Sijbers. "DART: A practical reconstruction algorithm for discrete tomography". *IEEE Transactions on Image Processing* 20.9 (2011), pp. 2542–2553 (cit. on pp. 18, 61, 96, 100).

[31] K. J. Batenburg and J. Sijbers. "DART: A fast heuristic algebraic reconstruction algorithm for discrete tomography". In: *IEEE International Conference on Image Processing, 2007, ICIP 2007.* (San Antonio, TX, USA). Vol. 4. IEEE. 2007, pp. IV–133 (cit. on p. 96).

[32] D. F. Bauer, C. Ulrich, T. Russ, A.-K. Golla, L. R. Schad, and F. G. Zöllner. "End-to-end deep learning CT image reconstruction for metal artifact reduction". *Applied Sciences* 12.1 (2021), p. 404 (cit. on p. 35).

[33] I. Bayram and M. E. Kamasak. "A directional total variation". In: *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO).* (Bucharest, Romania). IEEE. 2012, pp. 265–269 (cit. on p. 28).

[34] A. Beck. *First-order methods in optimization.* Philadelphia, PA, USA: SIAM, 2017 (cit. on p. 128).

[35] M. Beister, D. Kolditz, and W. A. Kalender. "Iterative reconstruction methods in X-ray CT". *Physica Medica* 28.2 (2012), pp. 94–108 (cit. on p. 14).

[36] R. Bellazzini, A. Brez, G. Spandre, M. Minuti, M. Pinchera, P. Delogu, P. L. De Ruvo, and A. Vincenzi. "PIXIE III: A very large area photon-counting CMOS pixel ASIC for sharp X-ray spectral imaging". *Journal of Instrumentation* 10.01 (2015), p. C01032 (cit. on p. 21).

[37] E. van den Berg and M. P. Friedlander. "Spot - A linear-operator toolbox" (2014) (cit. on p. 146).

[38] J. Berner, P. Grohs, G. Kutyniok, and P. Petersen. "The modern mathematics of deep learning". *arXiv preprint arXiv:2105.04026* (2021) (cit. on pp. 29, 30, 34, 35).

[39] M. W. Berry, M. Browne, A. N. L., V. P. Pauca, and R. J. Plemmons. "Algorithms and applications for approximate nonnegative matrix factorization". *Computational Statistics & Data Analysis* 52.1 (2007), pp. 155–173 (cit. on p. 132).

[40] R. Bhayana, A. Parakh, and A. Kambadakone. "Material decomposition with dual-and multi-energy computed tomography". *MRS Communications* 10.4 (2020), pp. 558–565 (cit. on pp. 24, 114).

[41] F. Bleichrodt, T. van Leeuwen, W. J. Palenstijn, W. van Aarle, J. Sijbers, and K. J. Batenburg. "Easy implementation of advanced tomography algorithms using the ASTRA toolbox with Spot operators". *Numerical Algorithms* 71.3 (2016), pp. 673–697 (cit. on p. 131).

[42] F. Bleichrodt, F. Tabak, and K. J. Batenburg. "SDART: An algorithm for discrete tomography from noisy projections". *Computer Vision and Image Understanding* 129 (2014), pp. 63–74 (cit. on p. 96).

[43] J. Bolte, S. Sabach, and M. Teboulle. "Proximal alternating linearized minimization for nonconvex and nonsmooth problems". *Mathematical Programming* 146.1 (2014), pp. 459–494 (cit. on p. 125).

[44] F. G. Bossema, M. Domínguez-Delmás, W. J. Palenstijn, A. Kostenko, J. Dorscheid, S. B. Coban, E. Hermens, and K. J. Batenburg. "A novel method for dendrochronology of large historical wooden objects using line trajectory X-ray tomography". *Scientific Reports* 11.1 (2021), pp. 1–12 (cit. on p. 13).

[45] S. P. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge, UK: Cambridge University Press, 2004 (cit. on pp. 191, 192).

[46] S. Boyd, N. Parikh, and E. Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Hanover, MA, USA: Now Publishers Inc., 2011 (cit. on p. 195).

[47] R. Bujila, A. Omar, and G. Poludniowski. "A validation of SpekPy: A software toolkit for modelling X-ray tube spectra". *Physica Medica* 75 (2020), pp. 44–54 (cit. on p. 130).

[48] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb. "Blind image fusion for hyperspectral imaging with the directional total variation". *Inverse Problems* 34.4 (2018), p. 044003 (cit. on p. 155).

[49] M. Busi, C. Kehl, J. R. Frisvad, and U. L. Olsen. "Metal artifact reduction in spectral X-ray CT using spectral deep learning". *Journal of Imaging* 8.3 (2022), p. 77 (cit. on p. 35).

[50] J.-W. Buurlage. "Real-time tomographic reconstruction". PhD thesis. Leiden University, 2020 (cit. on p. 18).

[51] T. M. Buzug. *Computed tomography: From photon statistics to modern cone-beam CT*. Berlin/Heidelberg, Germany: (1st ed.). Springer, 2008. Chap. 8, pp. 311–342 (cit. on pp. 7, 24, 61, 96).

[52] C. Cai, T. Rodet, S. Legoupil, and A. Mohammad-Djafari. "A full-spectral Bayesian reconstruction approach based on the material decomposition model applied in dual-energy computed tomography". *Medical Physics* 40.11 (2013), p. 111916 (cit. on p. 132).

[53] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy. "Deep learning for chest X-ray analysis: A survey". *Medical Image Analysis* 72 (2021), p. 102125 (cit. on p. 35).

[54] J. Cammin, J. S. Iwanczyk, and K. Taguchi. "Spectral/Photon-counting computed tomography". In: *Emerging Imaging Technologies in Medicine*. Boca Raton, FL, USA: CRC Press, 2012, pp. 42–61 (cit. on p. 23).

[55] S. Carmignato, W. Dewulf, and R. Leach. *Industrial X-ray computed tomography*. Cham, Switzerland: Springer, 2018 (cit. on pp. 5, 7, 13, 18, 24).

[56] C.-I. Chang. *Hyperspectral data processing: Algorithm design and analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2013 (cit. on p. 66).

[57] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang. "Deep learning: A primer for radiologists". *Radiographics* 37.7 (2017), pp. 2113–2131 (cit. on p. 40).

[58] L. Chen, Z. Wei, and Y. Xu. "A lightweight spectral-spatial feature extraction and fusion network for hyperspectral image classification". *Remote Sensing* 12.9 (2020), p. 1395 (cit. on p. 69).

[59] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi. "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks". *IEEE Transactions on Geoscience and Remote Sensing* 54.10 (2016), pp. 6232–6251 (cit. on p. 68).

[60] D. P. Clark and C. T. Badea. "Spectral diffusion: An algorithm for robust material decomposition of spectral CT data". *Physics in Medicine & Biology* 59.21 (2014), p. 6445 (cit. on pp. 23, 27, 118).

[61] D. P. Clark and C. T. Badea. "Hybrid spectral CT reconstruction". *PloS One* 12.7 (2017), e0180324 (cit. on p. 27).

[62] S. B. Coban, F. Lucka, W. J. Palenstijn, D. van Loo, and K. J. Batenburg. "Explorative imaging and its implementation at the FleX-ray laboratory". *Journal of Imaging* 6.4 (2020), p. 18 (cit. on pp. 45, 47).

[63] W. Cong, D. Harrison, Y. Xi, and G. Wang. "Projection decomposition for dual-energy computed tomography". *arXiv preprint arXiv:1805.05312* (2018) (cit. on p. 26).

[64] E. Cueva, A. Meaney, S. Siltanen, and M. J. Ehrhardt. "Synergistic multi-spectral CT reconstruction with directional total variation". *Philosophical Transactions of the Royal Society A* 379.2204 (2021), p. 20200198 (cit. on pp. 26–28).

[65] A. Dabravolski, K. J. Batenburg, and J. Sijbers. "A multiresolution approach to discrete tomography using DART". *PloS One* 9.9 (2014), e106090 (cit. on p. 96).

[66] Q. Dai, J.-H. Cheng, D.-W. Sun, and X.-A. Zeng. "Advances in feature selection methods for hyperspectral image processing in food industry applications: A review". *Critical Reviews in Food Science and Nutrition* 55.10 (2015), pp. 1368–1382 (cit. on p. 68).

[67] M. Danielsson, M. Persson, and M. Sjölin. "Photon-counting X-ray detectors for CT". *Physics in Medicine & Biology* 66.3 (2021), 03TR01 (cit. on p. 23).

[68] G. Demaurex and L. Sallé. "Detection of physical hazards". In: *Food Safety Management*. Elsevier, 2014, pp. 511–533 (cit. on pp. 3, 4, 152).

[69] A. M. Deshpande, A. A. Minai, and M. Kumar. "One-shot recognition of manufacturing defects in steel surfaces". *Procedia Manufacturing* 48 (2020), pp. 1064–1071 (cit. on p. 40).

[70] M. Diwakar and M. Kumar. "A review on CT image noise and its denoising". *Biomedical Signal Processing and Control* 42 (2018), pp. 73–88 (cit. on pp. 45, 61).

[71] N. Djurabekova, A. Goldberg, A. Hauptmann, D. Hawkes, G. Long, F. Lucka, and M. Betcke. "Application of Proximal Alternating Linearized Minimization (PALM) and inertial PALM to dynamic 3D CT". In: *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. (Philadelphia, PA, USA). Ed. by S. Matej and S. D. Metzler. Vol. 11072. International Society for Optics and Photonics. SPIE, 2019, pp. 30–34 (cit. on p. 61).

[72] W. Du, H. Shen, J. Fu, G. Zhang, and Q. He. "Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning". *NDT & E International* 107 (2019), p. 102144 (cit. on p. 35).

[73] Y. Dua, V. Kumar, and R. S. Singh. "Comprehensive review of hyperspectral image compression algorithms". *Optical Engineering* 59.9 (2020), p. 090902 (cit. on p. 35).

[74] J. Duchi, E. Hazan, and Y. Singer. "Adaptive subgradient methods for online learning and stochastic optimization." *Journal of Machine Learning Research* 12.7 (2011) (cit. on p. 33).

[75] N. Ducros, J. F. P. J. Abascal, B. Sixou, S. Rit, and F. Peyrin. "Regularization of nonlinear decomposition of spectral X-ray projection images". *Medical Physics* 44.9 (2017), e174–e187 (cit. on p. 119).

[76] M. C. Edwards, M. F. Stringer, et al. "Observations on patterns in foreign material investigations". *Food Control* 18.7 (2007), pp. 773–782 (cit. on p. 3).

[77] C. K. Egan, S. D. M. Jacques, T. Connolley, M. D. Wilson, M. C. Veale, P. Seller, and R. J. Cernik. "Dark-field hyperspectral X-ray imaging". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 470.2165 (2014), p. 20130629 (cit. on p. 118).

[78] C. K. Egan, S. D. M. Jacques, M. D. Wilson, M. C. Veale, P. Seller, A. M. Beale, R. A. D. Pattrick, P. J. Withers, and R. J. Cernik. "3D chemical imaging in the laboratory by hyperspectral X-ray computed tomography". *Scientific Reports* 5.1 (2015), pp. 1–9 (cit. on pp. 21, 114).

[79] L. Eger, S. Do, P. Ishwar, W. C. Karl, and H. Pien. "A learning-based approach to explosives detection using multi-energy X-ray computed tomography". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (Prague, Czech Republic). IEEE. 2011, pp. 2004–2007 (cit. on p. 66).

[80] H. Einarsdóttir, M. J. Emerson, L. H. Clemmensen, K. Scherer, K. Willer, M. Bech, R. Larsen, B. K. Ersbøll, and F. Pfeiffer. "Novelty detection of foreign objects in food using multi-modal X-ray imaging". *Food Control* 67 (2016), pp. 39–47 (cit. on pp. 40, 155).

[81] H. Einarsdóttir, B. Gudmundsson, and V. Ómarsson. "Automation in the fish industry". *Animal Frontiers* 12.2 (2022), pp. 32–39 (cit. on p. 35).

[82] G. Einarsson, J. N. Jensen, R. R. Paulsen, H. Einarsdottir, B. K. Ersbøll, A. B. Dahl, and L. B. Christensen. "Foreign object detection in multispectral X-ray images of food items using sparse discriminant analysis". In: *Scandinavian Conference on Image Analysis*. (Tromsø, Norway). Ed. by P. Sharma and F. M. Bianchi. Springer, 2017, pp. 350–361 (cit. on p. 61).

[83] G. Elmasry, M. Kamruzzaman, D.-W. Sun, and P. Allen. "Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review". *Critical Reviews in Food Science and Nutrition* 52.11 (2012), pp. 999–1023 (cit. on pp. 66, 67).

[84] C. L. Epstein. *Introduction to the mathematics of medical imaging*. Philadelphia, PA, USA: SIAM, 2007 (cit. on p. 14).

[85] A. Falkovskaya and A. Gowen. "Literature review: Spectral imaging applied to poultry products". *Poultry Science* 99.7 (2020), pp. 3709–3722 (cit. on p. 4).

[86] M. Fauvel, J. Chanussot, and J. A. Benediktsson. "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas". *EURASIP Journal on Advances in Signal Processing* 783194.1 (2009), pp. 1–14 (cit. on p. 68).

[87] L. A. Feldkamp, L. C. Davis, and J. W. Kress. "Practical cone-beam algorithm". *Journal of the Optical Society of America A* 1.6 (1984), pp. 612–619 (cit. on pp. 15, 61).

[88] F. Feng, S. Wang, C. Wang, and J. Zhang. "Learning deep hierarchical spatial-spectral features for hyperspectral image classification based on residual 3D-2D CNN". *Sensors* 19.23 (2019), p. 5276 (cit. on p. 69).

[89] Y.-Z. Feng and D.-W. Sun. "Application of hyperspectral imaging in food safety inspection and control: A review". *Critical Reviews in Food Science and Nutrition* 52.11 (2012), pp. 1039–1058 (cit. on pp. 3, 4).

[90] M. K. Ferguson, A. Ronay, Y. T. T. Lee, and K. H. Law. "Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning". *Smart and Sustainable Manufacturing Systems* 2.1 (2018) (cit. on p. 35).

[91] A. Förster, S. Brandstetter, and C. Schulze-Briese. "Transforming X-ray detection with hybrid photon counting detectors". *Philosophical Transactions of the Royal Society A* 377.2147 (2019), p. 20180241 (cit. on p. 21).

[92] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides. "Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery". *Electronic Imaging* 2017.17 (2017), pp. 185–190 (cit. on p. 66).

[93] J. Frank. *Electron tomography*. New York, NY, USA: Springer, 1992 (cit. on p. 96).

[94] E. Fredenberg. "Spectral and dual-energy X-ray imaging for medical applications". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), pp. 74–87 (cit. on pp. 19, 21).

[95] N. R. Fredette, A. Kavuri, and M. Das. "Multi-step material decomposition for spectral computed tomography". *Physics in Medicine & Biology* 64.14 (2019), p. 145001 (cit. on pp. 118, 119).

[96] C. Fröjdh, B. Norlin, and E. Fröjdh. "Spectral X-ray imaging with single photon processing detectors". *Journal of Instrumentation* 8.02 (2013), p. C02010 (cit. on p. 22).

[97] F. Fusseis, X. Xiao, C. Schrank, and F. De Carlo. "A brief guide to synchrotron radiation-based microtomography in (structural) geology and rock mechanics". *Journal of Structural Geology* 65 (2014), pp. 1–16 (cit. on p. 13).

[98] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. "A review on deep learning techniques applied to semantic segmentation". *arXiv preprint arXiv:1704.06857* (2017) (cit. on pp. 40, 41).

[99] R. Garnett. "A comprehensive review of dual-energy and multi-spectral computed tomography". *Clinical Imaging* 67 (2020), pp. 160–169 (cit. on pp. 21, 23–26, 28).

[100] *GitHub - SimonRit/OneStepSpectralCT: Matlab code for spectral CT one-step inversion. Implementation of five different methods.* `https://github.com/SimonRit/OneStepSpectralCT`. Accessed on 19 November 2021 (cit. on pp. 130, 131).

[101] L. Gjesteby, B. De Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. Wang. "Metal artifact reduction in CT: Where are we after four decades?" *IEEE Access* 4 (2016), pp. 5826–5849 (cit. on p. 59).

[102] X. Glorot, A. Bordes, and Y. Bengio. "Deep sparse rectifier neural networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. (Fort Lauderdale, FL, USA). JMLR Workshop and Conference Proceedings. 2011, pp. 315–323 (cit. on p. 31).

[103] J. Gorski, F. Pfeuffer, and K. Klamroth. "Biconvex sets and optimization with biconvex functions: A survey and extensions". *Mathematical Methods of Operations Research* 66.3 (2007), pp. 373–407 (cit. on pp. 122, 193).

[104] M. Graña, M. A. Veganzons, and B. Ayerdi. *Hyperspectral remote sensing scenes*. `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`. Accessed on 15 September 2020 (cit. on p. 77).

[105] M. Grandini, E. Bagli, and G. Visani. "Metrics for multi-class classification: An overview". *arXiv preprint arXiv:2008.05756* (2020) (cit. on p. 50).

[106] M. Graves, A. Smith, and B. Batchelor. "Approaches to foreign body detection in foods". *Trends in Food Science & Technology* 9.1 (1998), pp. 21–27 (cit. on p. 4).

[107] M. K. Griffin and H. K. Burke. "Compensation of hyperspectral data for atmospheric effects". *Lincoln Laboratory Journal* 14.1 (2003), pp. 29–54 (cit. on p. 81).

[108] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew. "A review of semantic segmentation using deep neural networks". *International Journal of Multimedia Information Retrieval* 7.2 (2018), pp. 87–93 (cit. on p. 40).

[109] R. Habel, M. Kudenov, and M. Wimmer. "Practical spectral photography". In: *Computer Graphics Forum*. Vol. 31. 2pt2. Wiley Online Library. 2012, pp. 449–458 (cit. on p. 155).

[110] M. Habermann, V. Frémont, and E. H. Shiguemori. "Feature selection for hyperspectral images using single-layer neural networks". In: *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*. (Madrid, Spain). 2017, pp. 1–6 (cit. on p. 66).

[111] R. P. Haff and N. Toyofuku. "X-ray detection of defects and contaminants in the food industry". *Sensing and Instrumentation for Food Quality and Safety* 2.4 (2008), pp. 262–273 (cit. on p. 40).

[112] P. C. Hansen, J. S. Jørgensen, and W. R. B. Lionheart. *Computed tomography: Algorithms, insight, and just enough theory*. Philadelphia, PA, USA: (1st ed.). SIAM, 2021 (cit. on pp. 14, 61).

[113] P. C. Hansen and M. Saxild-Hansen. "AIR tools - A MATLAB package of algebraic iterative reconstruction methods". *Journal of Computational and Applied Mathematics* 236.8 (2012), pp. 2167–2178 (cit. on p. 131).

[114] A. Hauptmann, O. Öktem, and C. Schönlieb. "Image reconstruction in dynamic inverse problems with temporal models". *arXiv preprint arXiv:2007.10238* (2020) (cit. on p. 61).

[115] K. He, G. Gkioxari, P. Dollár, and R. Girshick. "Mask R-CNN". In: *Proceedings of the IEEE International Conference on Computer Vision*. (Venice, Italy). IEEE, 2017, pp. 2961–2969 (cit. on p. 43).

[116] Y. He, Q. Xiao, X. Bai, L. Zhou, F. Liu, and C. Zhang. "Recent progress of nondestructive techniques for fruits damage inspection: A review". *Critical Reviews in Food Science and Nutrition* (2021), pp. 1–19 (cit. on p. 40).

[117] Y. He, X. Bai, Q. Xiao, F. Liu, L. Zhou, and C. Zhang. "Detection of adulteration in food based on nondestructive analysis techniques: A review". *Critical Reviews in Food Science and Nutrition* 61.14 (2021), pp. 2351–2371 (cit. on p. 4).

[118] B. J. Heismann, B. T. Schmidt, and T. Flohr. *Spectral computed tomography*. Bellingham, WA, USA: SPIE Press, 2012 (cit. on pp. 21, 26, 119).

[119] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg. "Noise2inverse: Self-supervised deep convolutional denoising for tomography". *IEEE Transactions on Computational Imaging* 6 (2020), pp. 1320–1335 (cit. on pp. 45, 152).

[120] A. A. Hendriksen. "Deep learning for tomographic reconstruction with limited data". PhD thesis. Leiden University, 2022 (cit. on p. 35).

[121] G. T. Herman and A. Kuba. *Discrete tomography: Foundations, algorithms, and applications*. New York, NY, USA: (1st ed.). Springer, 1999 (cit. on pp. 18, 61).

[122] G. T. Herman and A. Kuba. *Advances in discrete tomography and its applications*. New York, NY, USA: Springer Science & Business Media, 2007 (cit. on p. 18).

[123] M. R. Hestenes and E. Stiefel. "Methods of conjugate gradients for solving linear systems". *Journal of Research of the National Bureau of Standards* 49 (1952), pp. 409–436 (cit. on pp. 61, 132).

[124] T. Higaki, Y. Nakamura, F. Tatsugami, T. Nakaura, and K. Awai. "Improvement of image quality at CT and MRI using deep learning". *Japanese Journal of Radiology* 37.1 (2019), pp. 73–80 (cit. on p. 35).

[125] T. Hohweiller, N. Ducros, F. Peyrin, and B. Sixou. "Spectral CT material decomposition in the presence of poisson noise: A Kullback–Leibler approach". *IRBM* 38.4 (2017), pp. 214–218 (cit. on pp. 118, 119).

[126] K. M. Holt. "Total nuclear variation and jacobian extensions of total variation for vector fields". *IEEE Transactions on Image Processing* 23.9 (2014), pp. 3975–3989 (cit. on p. 28).

[127] J. Hsieh. *Computed tomography: Principles, design, artifacts, and recent advances*. Bellingham, WA, USA, 2009 (cit. on p. 96).

[128] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li. "Deep convolutional neural networks for hyperspectral image classification". *Journal of Sensors* 2015 (2015), p. 258619 (cit. on p. 66).

[129] H. Huang, M. Chen, and Y. Duan. "Dimensionality reduction of hyperspectral image using spatial-spectral regularized sparse hypergraph embedding". *Remote Sensing* 11.9 (2019), p. 1039 (cit. on p. 67).

[130] J. H. Hubbell and S. M. Seltzer. *Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 KeV to 20 MeV for elements Z=1 to 92 and 48 additional substances of dosimetric interest*. Tech. rep. National Institute of Standards and Technology-PL, Gaithersburg, MD, USA. Ionizing Radiation Div., 1995 (cit. on pp. 57, 78, 123, 130, 184, 197).

[131] R. Huber, G. Haberfehlner, M. Holler, G. Kothleitner, and K. Bredies. "Total generalized variation regularization for multi-modal electron tomography". *Nanoscale* 11.12 (2019), pp. 5617–5632 (cit. on p. 28).

[132] M. Imani and H. Ghassemian. "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges". *Information Fusion* 59 (2020), pp. 59–83 (cit. on p. 68).

[133] N. Jaccard, T. W. Rogers, E. J. Morton, and L. D. Griffin. "Detection of concealed cars in complex cargo X-ray imagery using deep learning". *Journal of X-ray Science and Technology* 25.3 (2017), pp. 323–339 (cit. on p. 68).

[134] S. D. M. Jacques, C. K. Egan, M. D. Wilson, M. C. Veale, P. Seller, and R. J. Cernik. "A laboratory system for element specific hyperspectral X-ray imaging". *Analyst* 138.3 (2013), pp. 755–759 (cit. on p. 114).

[135] S. Jadon. "A survey of loss functions for semantic segmentation". *arXiv preprint arXiv:2006.14822* (2020) (cit. on pp. 48, 77).

[136] F. Jolivet, J. Lesaint, C. Fournier, M. Garcin, and A. Brambilla. "An efficient one-step method for spectral CT based on an approximate linear model". *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020) (cit. on p. 119).

[137] A. A. Joy, M. A. M. Hasan, and M. A. Hossain. "A comparison of supervised and unsupervised dimension reduction methods for hyperspectral image classification". In: *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. (Cox's Bazar, Bangladesh). IEEE. 2019, pp. 1–6 (cit. on pp. 67, 68).

[138] A. Kadu and M. T. Zeegers. *mzeegers/ADJUST: ADJUST*. Zenodo, `https://doi.org/10.5281/zenodo.7821066`. Version v1.0.2. 2023 (cit. on p. 146).

[139] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Vol. 160. New York, NY, USA: Springer-Verlag, 2004 (cit. on p. 129).

[140] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Philadelphia, PA, USA: SIAM, 2001 (cit. on pp. 47, 98, 101, 141).

[141] K. V. Kale, M. M. Solankar, D. B. Nalawade, R. K. Dhumal, and H. R. Gite. "A research review on hyperspectral data processing and analysis algorithms". *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 87.4 (2017), pp. 541–555 (cit. on p. 67).

[142] M. Karcaaltincaba and A. Aktas. "Dual-energy CT revisited with multidetector CT: Review of principles and clinical applications". *Diagnostic and Interventional Radiology* 17.3 (2017) (cit. on p. 114).

[143] D. Kazantsev, J. S. Jørgensen, M. S. Andersen, W. R. B. Lionheart, P. D. Lee, and P. J. Withers. "Joint image reconstruction method with correlative multi-channel prior for X-ray spectral computed tomography". *Inverse Problems* 34.6 (2018), p. 064001 (cit. on pp. 26, 28, 61, 110, 119).

[144] D. Kern and A. Mastmeyer. "3D bounding box detection in volumetric medical image data: A systematic literature review". In: *2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA)*. (Chengdu, China). IEEE, 2021, pp. 509–516 (cit. on pp. 45, 61).

[145] K. Kim, H. Kim, J. Chun, M. Kang, M. Hong, and B. Min. "Real-time anomaly detection in packaged food X-ray images using supervised learning" (2021) (cit. on p. 153).

[146] L. Kingma D. P.; Ba. "ADAM: A method for stochastic optimization". In: *Proceedings of the International Conference on Learning Representations*. (San Diego, CA, USA). Ed. by Y. Bengio and Y. LeCun. 2015 (cit. on pp. 33, 49, 61, 75).

[147] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson. "Joint MR-PET reconstruction using a multi-channel image regularizer". *IEEE Transactions on Medical Imaging* 36.1 (2016), pp. 1–16 (cit. on p. 28).

[148] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. M. Benzel, H. A. Lowers, R. L. Driscoll, and A. J. Klein. "USGS spectral library version 7 data: US geological survey data release". *United States Geological Survey (USGS): Reston, VA, USA* (2017) (cit. on p. 80).

[149] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. M. Benzel, H. A. Lowers, R. L. Driscoll, and A. J. Klein. *USGS spectral library version 7*. Tech. rep. US Geological Survey, 2017 (cit. on p. 80).

[150] O. Kouropteva, O. Okun, and M. Pietikäinen. "Selection of the optimal parameter value for the locally linear embedding algorithm". *FSKD* 2 (2002), pp. 359–363 (cit. on p. 68).

[151] M. V. Koval'chuk, E. B. Yatsishina, A. E. Blagov, E. Y. Tereshchenko, P. A. Prosekov, and Y. A. Dyakova. "X-ray and synchrotron methods in studies of cultural heritage sites". *Crystallography Reports* 61.5 (2016), pp. 703–717 (cit. on p. 13).

[152] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh. "Feature extraction for hyperspectral image classification: A review". *International Journal of Remote Sensing* 41.16 (2020), pp. 6248–6287 (cit. on pp. 67, 68).

[153] J. Kwon, J. Lee, and W. Kim. "Real-time detection of foreign objects using X-ray imaging for dry food manufacturing line". In: *2008 IEEE International Symposium on Consumer Electronics*. (Vilamoura, Portugal). IEEE, 2008, pp. 1–4 (cit. on p. 40).

[154] M. J. Lagerwerf, D. M. Pelt, W. J. Palenstijn, and K. J. Batenburg. "A computationally efficient reconstruction algorithm for circular cone-beam computed tomography using shallow neural networks". *Journal of Imaging* 6.12 (2020), p. 135 (cit. on p. 49).

[155] J. H. Lee, Y. J. Kim, and K. G. Kim. "Bone age estimation using deep learning and hand X-ray images". *Biomedical Engineering Letters* 10.3 (2020), pp. 323–331 (cit. on p. 35).

[156] T. van Leeuwen and A. Y. Aravkin. "Variable Projection for NonSmooth Problems". *SIAM Journal on Scientific Computing* 43.5 (2021), S249–S268 (cit. on p. 125).

[157] M. M. Lell and M. Kachelrieß. "Recent and upcoming technological developments in computed tomography: High speed, low dose, deep learning, multienergy". *Investigative Radiology* 55.1 (2020), pp. 8–19 (cit. on p. 35).

[158] L. Lenchik, L. Heacock, A. A. Weaver, R. D. Boutin, T. S. Cook, J. Itri, C. G. Filippi, R. P. Gullapalli, J. Lee, and e. a. Zagurovskaya M. "Automated segmentation of tissues using CT and MRI: A systematic review". *Academic Radiology* 26.12 (2019), pp. 1695–1706 (cit. on p. 44).

[159] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough. "Photon-counting detector CT: System design and clinical applications of an emerging technology". *Radiographics* 39.3 (2019), pp. 729–743 (cit. on pp. 23, 28).

[160] J. Leuschner, M. Schmidt, P. S. Ganguly, V. Andriiashen, S. B. Coban, A. Denker, D. Bauer, A. Hadjifaradji, K. J. Batenburg, P. Maass, and M. van Eijnatten. "Quantitative comparison of deep learning-based image reconstruction methods for low-dose and sparse-angle CT applications". *Journal of Imaging* 7.3 (2021), p. 44 (cit. on p. 35).

[161] F. Li, M. K. Ng, R. Plemmons, S. Prasad, and Q. Zhang. "Hyperspectral image segmentation, deblurring, and spectral analysis for material identification". In: *Visual Information Processing XIX.* (Orlando, FL, USA). Vol. 7701. International Society for Optics and Photonics. 2010, p. 770103 (cit. on p. 66).

[162] S. Li, H. Luo, M. Hu, M. Zhang, J. Feng, Y. Liu, Q. Dong, and B. Liu. "Optical non-destructive techniques for small berry fruits: A review". *Artificial Intelligence in Agriculture* 2 (2019), pp. 85–98 (cit. on p. 40).

[163] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson. "Deep learning for hyperspectral image classification: An overview". *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019), pp. 6690–6709 (cit. on p. 68).

[164] T. Li, J. Zhang, and Y. Zhang. "Classification of hyperspectral image based on deep belief networks". In: *2014 IEEE International Conference on Image Processing (ICIP).* (Paris, France). IEEE. 2014, pp. 5132–5136 (cit. on p. 67).

[165] W. Li, F. Feng, H. Li, and Q. Du. "Discriminant analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques". *IEEE Geoscience and Remote Sensing Magazine* 6.1 (2018), pp. 15–34 (cit. on p. 67).

[166] Z. Li, S. Ravishankar, Y. Long, and J. A. Fessler. "DECT-MULTRA: Dual-energy CT image decomposition with learned mixed material models and efficient clustering". *IEEE Transactions on Medical Imaging* 39.4 (2019), pp. 1223–1234 (cit. on p. 119).

[167] G. Lin, A. Milan, C. Shen, and I. Reid. "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (Honolulu, HI, USA). IEEE, 2017, pp. 1925–1934 (cit. on p. 43).

[168] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei. "Supervised deep feature extraction for hyperspectral image classification". *IEEE Transactions on Geoscience and Remote Sensing* 56.4 (2017), pp. 1909–1921 (cit. on p. 68).

[169] B. Liu, Y. Li, G. Li, and A. Liu. "A spectral feature based convolutional neural network for classification of sea surface oil spill". *ISPRS International Journal of Geo-Information* 8.4 (2019), p. 160 (cit. on p. 69).

[170] J. Liu and H. Gao. "Material reconstruction for spectral computed tomography with detector response function". *Inverse Problems* 32.11 (2016), p. 114001 (cit. on p. 154).

[171] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (Boston, MA, USA). IEEE, 2015, pp. 3431–3440 (cit. on p. 43).

[172] Y. Long and J. A. Fessler. "Multi-material decomposition using statistical image reconstruction for spectral CT". *IEEE Transactions on Medical Imaging* 33.8 (2014), pp. 1614–1626 (cit. on pp. 26, 132).

[173] T. van de Looverbosch, E. Raeymaekers, P. Verboven, J. Sijbers, and B. Nicolaï. "Nondestructive internal disorder detection of Conference pears by semantic segmentation of X-ray CT scans using deep learning". *Expert Systems with Applications* 176 (2021), p. 114925 (cit. on p. 41).

[174] T. van de Looverbosch, M. H. R. Bhuiyan, P. Verboven, M. Dierick, D. van Loo, J. De Beenbouwer, J. Sijbers, and B. Nicolaï. "Nondestructive internal quality inspection of pear fruit by X-ray CT using machine learning". *Food Control* 113 (2020), p. 107170 (cit. on p. 13).

[175] G. Lu and B. Fei. "Medical hyperspectral imaging: A review". *Journal of Biomedical Optics* 19.1 (2014), p. 010901 (cit. on p. 66).

[176] T. Lukić and P. Balázs. "Limited-view binary tomography reconstruction assisted by shape centroid". *The Visual Computer* 38.2 (2022), pp. 695–705 (cit. on p. 18).

[177] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang. "Image and video compression with neural networks: A review". *IEEE Transactions on Circuits and Systems for Video Technology* 30.6 (2019), pp. 1683–1698 (cit. on p. 35).

[178] A. L. Maas, A. Y. Hannun, and A. Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Proceedings of the ICML Workshop on Deep Learning for Audio, Speech and Language Processing*. (Atlanta, GA, USA). Vol. 30. 1. 2013, p. 3 (cit. on pp. 31, 73).

[179] C. Maaß, M. Baer, and M. Kachelrieß. "Image-based dual energy CT using optimized precorrection functions: A practical new approach of material decomposition in image domain". *Medical Physics* 36.8 (2009), pp. 3818–3829 (cit. on p. 26).

[180] L. van der Maaten, E. Postma, and J. van den Herik. "Dimensionality reduction: A comparative review". *Journal of Machine Learning Research* 10.66-71 (2009), p. 13 (cit. on p. 68).

[181] F. J. Maestre-Deusto, G. Scavello, J. Pizarro, and P. L. Galindo. "ADART: An adaptive algebraic reconstruction algorithm for discrete tomography". *IEEE Transactions on Image Processing* 20.8 (2011), pp. 2146–2152 (cit. on p. 96).

[182] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J. Choi, C. Riess, A. Keil, and R. Fahrig. "CONRAD - A software framework for cone-beam imaging in radiology". *Medical Physics* 40.11 (2013), p. 111914 (cit. on p. 129).

[183] A. Maier, C. Syben, T. Lasser, and C. Riess. "A gentle introduction to deep learning in medical image processing". *Zeitschrift für Medizinische Physik* 29.2 (2019), pp. 86–101 (cit. on p. 35).

[184] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis. "Deep supervised learning for hyperspectral data classification through convolutional neural networks". In: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. (Milan, Italy). IEEE. 2015, pp. 4959–4962 (cit. on pp. 66, 68).

[185] H. E. Martz, C. M. Logan, D. J. Schneberk, and P. J. Shull. *X-ray imaging: Fundamentals, industrial techniques and applications*. Boca Raton, FL, USA: CRC Press, 2016 (cit. on p. 35).

[186] S. K. Mathanker, P. R. Weckler, and T. J. Bowser. "X-ray applications in food and agriculture: A review". *Transactions of the ASABE* 56.3 (2013), pp. 1227–1239 (cit. on pp. 18, 40).

[187] J. Matula, V. Polakova, J. Salplachta, M. Tesarova, T. Zikmund, M. Kaucka, I. Adameyko, and J. Kaiser. "Resolving complex cartilage structures in developmental biology via deep learning-based automatic segmentation of X-ray computed microtomography images". *Scientific Reports* 12.1 (2022), p. 8728 (cit. on p. 35).

[188] M. T. McCann, K. H. Jin, and M. Unser. "Convolutional neural networks for inverse problems in imaging: A review". *IEEE Signal Processing Magazine* 34.6 (2017), pp. 85–95 (cit. on p. 32).

[189] K. Mechlem, S. Ehn, T. Sellerer, E. Braig, D. Münzel, F. Pfeiffer, and P. B. Noël. "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model". *IEEE Transactions on Medical Imaging* 37.1 (2017), pp. 68–80 (cit. on pp. 119, 132).

[190] K. Mechlem, T. Sellerer, S. Ehn, D. Münzel, E. Braig, J. Herzen, P. Noël, and F. Pfeiffer. "Spectral angiography material decomposition using an empirical forward model and a dictionary-based regularization". *IEEE Transactions on Medical Imaging* 37.10 (2018), pp. 2298–2309 (cit. on pp. 118, 119).

[191] D. Mery, I. Lillo, H. Loebel, V. Riffo, A. Soto, A. Cipriano, and J. M. Aguilera. "Automated fish bone detection using X-ray imaging". *Journal of Food Engineering* 105.3 (2011), pp. 485–492 (cit. on pp. 11, 40).

[192] D. Mery. "Computer vision for X-ray testing". *Switzerland: Springer International Publishing* 10 (2015), pp. 978–3 (cit. on pp. 6–8, 22, 35, 153).

[193] D. Mery, D. Saavedra, and M. Prasad. "X-ray baggage inspection with computer vision: A survey". *IEEE Access* 8 (2020), pp. 145620–145633 (cit. on p. 13).

[194] H. S. El-Mesery, H. Mao, and A. E.-F. Abomohra. "Applications of non-destructive technologies for agricultural and food products quality inspection". *Sensors* 19.4 (2019), p. 846 (cit. on p. 4).

[195] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement. "Survey on deep learning for radiotherapy". *Computers in Biology and Medicine* 98 (2018), pp. 126–146 (cit. on p. 35).

[196] P. A. Midgley and M. Weyland. "3D electron microscopy in the physical sciences: The development of Z-contrast and EFTEM tomography". *Ultramicroscopy* 96.3-4 (2003), pp. 413–431 (cit. on p. 96).

[197] J. Minnema, M. van Eijnatten, H. Der Sarkissian, S. Doyle, J. Koivisto, J. Wolff, T. Forouzanfar, F. Lucka, and K. J. Batenburg. "Efficient high cone-angle artifact reduction in circular cone-beam CT using deep learning with geometry-aware dimension reduction". *Physics in Medicine & Biology* 66.13 (2021), p. 135015 (cit. on p. 35).

[198] D. Mishra, S. K. Singh, and R. K. Singh. "Deep architectures for image compression: A critical review". *Signal Processing* 191 (2022), p. 108346 (cit. on p. 35).

[199] S. Si-Mohamed, D. Bar-Ness, M. Sigovan, D. P. Cormode, P. Coulon, E. Coche, A. Vlassenbroek, G. Normand, L. Boussel, and P. Douek. "Review of an initial experience with an experimental spectral photon-counting computed tomography system". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 873 (2017), pp. 27–35 (cit. on pp. 40, 61, 114, 118).

[200] M. T. Mohd Khairi, S. Ibrahim, M. A. Md Yunus, and M. Faramarzi. "Noninvasive techniques for detection of foreign bodies in food: A review". *Journal of Food Process Engineering* 41.6 (2018), e12808 (cit. on pp. 4, 6, 40).

[201] M. P. Morigi, F. Casali, M. Bettuzzi, R. Brancaccio, and V. d'Errico. "Application of X-ray computed tomography to cultural heritage diagnostics". *Applied Physics A* 100.3 (2010), pp. 653–661 (cit. on p. 13).

[202] C. Mory, B. Sixou, S. Si-Mohamed, L. Boussel, and S. Rit. "Comparison of five one-step reconstruction algorithms for spectral CT". *Physics in Medicine & Biology* 63.23 (2018), p. 235001 (cit. on pp. 119, 129, 131).

[203] G. Motta, F. Rizzo, and J. A. Storer. *Hyperspectral data compression*. Berlin, Germany: Springer Science & Business Media, 2006 (cit. on p. 66).

[204] A. Mouton and T. P. Breckon. "A review of automated image understanding within 3D baggage computed tomography security screening". *Journal of X-ray Science and Technology* 23.5 (2015), pp. 531–555 (cit. on p. 13).

[205] K. P. Murphy. *Probabilistic machine learning: Advanced topics*. Cambridge, MA, USA: MIT Press, 2022 (cit. on pp. 29, 33).

[206] J. Nalepa, M. Myller, and M. Kawulok. "Validating hyperspectral image segmentation". *IEEE Geoscience and Remote Sensing Letters* 16.8 (2019), pp. 1264–1268 (cit. on pp. 66, 77).

[207] K. Narsaiah, A. K. Biswas, and P. K. Mandal. "Nondestructive methods for carcass and meat quality evaluation". In: *Meat Quality Analysis*. Ed. by A. K. Biswas and P. K. Mandal. Academic Press, 2020, pp. 37–49 (cit. on p. 40).

[208] R. A. Nasirudin, R. Tachibana, J. J. Näppi, K. Mei, F. K. Kopp, E. J. Rummeny, H. Yoshida, and P. B. Noël. "A comparison of material decomposition techniques for dual-energy CT colonography". In: *Medical Imaging 2015: Physics of Medical Imaging*. Vol. 9412. International Society for Optics and Photonics. 2015, 94124F (cit. on p. 26).

[209] F. Natterer. *The mathematics of computerized tomography*. Philadelphia, PA, USA: SIAM, 2001 (cit. on p. 14).

[210] B. M. Nicolaï, T. Defraeye, B. De Ketelaere, E. Herremans, M. L. A. T. M. Hertog, W. Saeys, A. Torricelli, T. Vandendriessche, and P. Verboven. "Nondestructive measurement of fruit and vegetable quality". *Annual Review of Food Science and Technology* 5 (2014), pp. 285–312 (cit. on p. 40).

[211] V. V. Nikitin, M. Carlsson, F. Andersson, and R. Mokso. "Four-dimensional tomographic reconstruction by time domain decomposition". *IEEE Transactions on Computational Imaging* 5.3 (2019), pp. 409–419 (cit. on p. 61).

[212] H. Noh, S. Hong, and B. Han. "Learning deconvolution network for semantic segmentation". In: *Proceedings of the IEEE International Conference on Computer Vision*. (Santiago, Chile). IEEE, 2015, pp. 1520–1528 (cit. on p. 43).

[213] I. Ordavo, S. Ihle, V. Arkadiev, O. Scharf, H. Soltau, A. Bjeoumikhov, S. Bjeoumikhova, G. Buzanich, R. Gubzhokov, A. Günther, et al. "A new pnCCD-based color X-ray camera for fast spatial and energy-resolved measurements". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2011), pp. 250–257 (cit. on pp. 21, 114).

[214] N. Otsu. "A threshold selection method from gray-level histograms". *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66 (cit. on pp. 48, 179).

[215] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers. "The ASTRA tomography toolbox". In: *13th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE*. (Almeria, Spain). Vol. 2013. 2013, pp. 1139–1145 (cit. on p. 100).

[216] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson. "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network". *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017), pp. 639–643 (cit. on p. 69).

[217] H. Pan, C. Zhou, Q. Zhu, and D. Zheng. "A fast registration from 3D CT images to 2D X-ray images". In: *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*. (Shanghai, China). IEEE. IEEE, 2018, pp. 351–355 (cit. on p. 41).

[218] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza. "A new deep convolutional neural network for fast hyperspectral image classification". *ISPRS Journal of Photogrammetry and Remote Sensing* 145 (2018), pp. 120–147 (cit. on pp. 68, 69).

[219] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza. "Deep learning classifiers for hyperspectral imaging: A review". *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (2019), pp. 279–317 (cit. on p. 68).

[220] N. Parikh and S. Boyd. "Proximal algorithms". *Foundations and Trends in Optimization* 1.3 (2014), pp. 127–239 (cit. on p. 125).

[221] M. S. Passmore, R. Bates, K. Mathieson, V. O'Shea, M. Rahman, P. Seller, and K. M. Smith. "Characterisation of a single photon counting pixel detector". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 466.1 (2001), pp. 202–208 (cit. on p. 120).

[222] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. "Automatic differentiation in PyTorch". In: *Proceedings of the NIPS Autodiff Workshop*. (Long Beach, CA, USA). 2017 (cit. on pp. 35, 48, 63, 77, 93).

[223] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and e. a. A. Luca. "PyTorch: An imperative style, high-performance deep learning library". In: *Advances in Neural Information Processing Systems*. (Red Hook, NY, USA). Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. 2019, pp. 8026–8037 (cit. on pp. 35, 48, 63, 77, 93).

[224] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 82).

[225] D. M. Pelt. *GitHub - dmpelt/msdnet: Python implementation of the Mixed-Scale Dense Convolutional Neural Network*. `https://github.com/dmpelt/msdnet`. Accessed on 24 November 2020. 2019 (cit. on pp. 49, 75).

[226] D. M. Pelt, K. J. Batenburg, and J. A. Sethian. "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks". *Journal of Imaging* 4.11 (2018), p. 128 (cit. on pp. 35, 49).

[227] D. M. Pelt and J. A. Sethian. "A mixed-scale dense convolutional neural network for image analysis". *Proceedings of the National Academy of Sciences* 115.2 (2018), pp. 254–259 (cit. on pp. 48, 49, 63, 67, 70, 71, 74, 75, 93).

[228] H. Petersson, D. Gustafsson, and D. Bergstrom. "Hyperspectral image analysis using deep learning - A review". In: *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. (Oulu, Finland). IEEE. 2016, pp. 1–6 (cit. on p. 68).

[229] G. Pilikos, L. Horchens, K. J. Batenburg, T. van Leeuwen, and F. Lucka. "Deep data compression for approximate ultrasonic image formation". In: *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2020, pp. 1–4 (cit. on p. 35).

[230] G. Poludniowski, A. Omar, R. Bujila, and P. Andreo. "SpekPy v2.0 - A software toolkit for modelling X-ray tube spectra". *Medical Physics* 48.7 (2021) (cit. on p. 130).

[231] T. Qiao, J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, Q. Dai, and S. Marshall. "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis". *IEEE Transactions on Geoscience and Remote Sensing* 55.1 (2016), pp. 119–133 (cit. on p. 67).

[232] J. Radon. "1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten". *Classic Papers in Modern Diagnostic Radiology* 5 (2005), p. 21 (cit. on p. 14).

[233] N. Rassouli, M. Etesami, A. Dhanantwari, and P. Rajiah. "Detector-based spectral CT with a novel dual-layer technology: Principles and applications". *Insights into Imaging* 8.6 (2017), pp. 589–598 (cit. on p. 24).

[234] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson. "Feature extraction for hyperspectral imagery: The evolution from shallow to deep". *arXiv preprint arXiv:2003.02822* (2020) (cit. on p. 67).

[235] V. Rebuffel and J. Dinten. "Dual-energy X-ray imaging: Benefits and limits". *Insight - Non-Destructive Testing and Condition Monitoring,* 49.10 (2007), pp. 589–594 (cit. on p. 61).

[236] R. Redus, A. Huber, J. Pantazis, T. Pantazis, and D. Sperry. "Design and performance of the X-123 compact X-ray and Gamma-ray spectroscopy system". In: *2006 IEEE Nuclear Science Symposium Conference Record.* (Piscataway Township, NJ, USA). Vol. 6. IEEE. 2006, pp. 3794–3797 (cit. on p. 114).

[237] E. G. Rens, M. T. Zeegers, I. Rabbers, A. Szabó, and R. M. H. Merks. "Autocrine inhibition of cell motility can drive epithelial branching morphogenesis in the absence of growth". *Philosophical Transactions of the Royal Society B* 375.1807 (2020), p. 20190386 (cit. on p. 209).

[238] D. S. Rigie and P. J. La Rivière. "Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization". *Physics in Medicine & Biology* 60.5 (2015), p. 1741 (cit. on pp. 28, 61).

[239] D. Rong, L. Xie, and Y. Ying. "Computer vision detection of foreign objects in walnuts using deep learning". *Computers and Electronics in Agriculture* 162 (2019), pp. 1001–1010 (cit. on p. 40).

[240] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-assisted Intervention.* (Munich, Germany). Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Springer, 2015, pp. 234–241 (cit. on pp. 43, 48, 67, 71, 74).

[241] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268 (cit. on p. 133).

[242] P. Russo. *Handbook of X-ray imaging: Physics and technology.* Boca Raton, FL, USA: (1st ed.). CRC Press, 2017 (cit. on pp. 7, 43).

[243] J. Salazar-Vazquez and A. Mendez-Vazquez. "A plug-and-play hyperspectral imaging sensor using low-cost equipment". *HardwareX* 7 (2020), e00087 (cit. on p. 155).

[244] M. Salehjahromi, Y. Zhang, and H. Yu. "Comparison study of regularizations in spectral computed tomography reconstruction". *Sensing and Imaging* 19.1 (2018), pp. 1–23 (cit. on pp. 27, 119).

[245] M. Al-Sarayreh, M. M. Reis, W. Q. Yan, and R. Klette. "A sequential CNN approach for foreign object detection in hyperspectral images". In: *International Conference on Computer Analysis of Images and Patterns.* (Salerno, Italy). Ed. by M. Vento and G. Percannella. Springer, 2019, pp. 271–283 (cit. on pp. 40, 41).

[246] A. Sawatzky, Q. Xu, C. O. Schirra, and M. A. Anastasio. "Proximal ADMM for multi-channel image reconstruction in spectral X-ray CT". *IEEE Transactions on Medical Imaging* 33.8 (2014), pp. 1657–1668 (cit. on pp. 61, 119).

[247] O. Scharf, S. Ihle, I. Ordavo, V. Arkadiev, A. Bjeoumikhov, S. Bjeoumikhova, G. Buzanich, R. Gubzhokov, A. Gunther, R. Hartmann, et al. "Compact pnCCD-based X-ray camera with high spatial and energy resolution: A color X-ray camera". *Analytical Chemistry* 83.7 (2011), pp. 2532–2538 (cit. on p. 114).

[248] C. O. Schirra, E. Roessl, T. Koehler, B. Brendel, A. Thran, D. Pan, M. A. Anastasio, and R. Proksa. "Statistical reconstruction of material decomposed data in spectral CT". *IEEE Transactions on Medical Imaging* 32.7 (2013), pp. 1249–1257 (cit. on p. 118).

[249] C. O. Schirra, B. Brendel, M. A. Anastasio, and E. Roessl. "Spectral CT: A technology primer for contrast agent development". *Contrast Media & Molecular Imaging* 9.1 (2014), pp. 62–70 (cit. on pp. 19, 21, 28).

[250] M. Schmidt, E. Berg, M. Friedlander, and K. Murphy. "Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm". In: *Artificial Intelligence and Statistics.* (Clearwater Beach, FL, USA). PMLR. 2009, pp. 456–463 (cit. on pp. 132, 146).

[251] T. G. Schmidt, B. A. Sammut, R. F. Barber, X. Pan, and E. Y. Sidky. "Addressing CT metal artifacts using photon-counting detectors and one-step spectral CT image reconstruction". *Medical Physics* (2022) (cit. on p. 27).

[252] R. Schoonhoven, A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg. "LEAN: Graph-based pruning for convolutional neural networks by extracting longest chains". *arXiv preprint arXiv:2011.06923* (2020) (cit. on p. 154).

[253] D. E. Schut, K. J. Batenburg, R. van Liere, and T. van Leeuwen. "TOP-CT: Trajectory with Overlapping Projections X-ray Computed Tomography". *IEEE Transactions on Computational Imaging* 8 (2022), pp. 598–608 (cit. on p. 154).

[254] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller. "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography". *IEEE Transactions on Image Processing* 23.4 (2014), pp. 1678–1693 (cit. on p. 61).

[255] D. Sero, I. Garachon, E. Hermens, R. van Liere, and K. J. Batenburg. "The study of three-dimensional fingerprint recognition in cultural heritage: Trends and challenges". *Journal on Computing and Cultural Heritage* 14.4 (2021), pp. 1–20 (cit. on p. 13).

[256] M. Sezgin and B. Sankur. "Survey over image thresholding techniques and quantitative performance evaluation". *Journal of Electronic imaging* 13.1 (2004), pp. 146–165 (cit. on p. 40).

[257] V. Sharma, A. Diba, T. Tuytelaars, and L. van Gool. "Hyperspectral CNN for image classification & band selection, with application to face recognition". *Technical Report KUL/ESAT/PSI/1604, KU Leuven, ESAT, Leuven, Belgium* (2016) (cit. on p. 66).

[258] *Siemens Healthineers, Simulation of X-ray Spectra: Online tool for the simulation of X-ray Spectra (2018).* `https://www.oem-xray-components.siemens.com/X-ray-spectra-simulation`. Accessed on 24 February 2020 (cit. on p. 185).

[259] A. Signoroni, M. Savardi, A. Baronio, and S. Benini. "Deep learning meets hyperspectral image analysis: A multidisciplinary review". *Journal of Imaging* 5.5 (2019), p. 52 (cit. on pp. 66, 69, 77).

[260] G. Silva, L. Oliveira, and M. Pithon. "Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives". *Expert Systems with Applications* 107 (2018), pp. 15–31 (cit. on pp. 40, 41).

[261] J. da Silva, F. Grönberg, B. Cederström, M. Persson, M. Sjölin, Z. Alagic, R. Bujila, and M. Danielsson. "Resolution characterization of a silicon-based, photon-counting computed tomography prototype capable of patient scanning". *Journal of Medical Imaging* 6.4 (2019), p. 043502 (cit. on p. 28).

[262] J. Sittner, J. R. A. Godinho, A. D. Renno, V. Cnudde, M. Boone, T. De Schryver, D. van Loo, M. Merkulova, A. Roine, and J. Liipo. "Data for: Spectral X-ray computed micro tomography: 3-dimensional chemical imaging [Data set]" (2022) (cit. on pp. 141, 147).

[263] J. Sittner, J. R. A. Godinho, A. D. Renno, V. Cnudde, M. Boone, T. De Schryver, D. van Loo, M. Merkulova, A. Roine, and J. Liipo. "Spectral X-ray computed micro tomography: 3-dimensional chemical imaging". *X-ray Spectrometry* 50.2 (2021), pp. 92–105 (cit. on pp. 21, 141).

[264] W. Skrzynski. "X-ray detectors in medical imaging". In: *Advanced X-ray Detector Technologies*. Springer, 2022, pp. 135–149 (cit. on pp. 22, 23).

[265] A. van der Sluis and H. A. van der Vorst. "SIRT-and CG-type methods for the iterative solution of sparse linear least-squares problems". *Linear Algebra and its Applications* 130 (1990), pp. 257–303 (cit. on p. 47).

[266] A. So and S. Nicolaou. "Spectral computed tomography: Fundamental principles and recent developments". *Korean Journal of Radiology* 22.1 (2021), p. 86 (cit. on p. 26).

[267] J. Spiegelberg and J. Rusz. "Can we use PCA to detect small signals in noisy data?" *Ultramicroscopy* 172 (2017), pp. 40–46 (cit. on p. 66).

[268] *Standard Solar Spectra | PVEducation (2019)*. `https://www.pveducation.org/pvcdrom/appendices/standard-solar-spectra`. Accessed on 8 July 2020 (cit. on p. 80).

[269] E. Ström, M. Persson, A. Eguizabal, and O. Öktem. "Photon-counting CT reconstruction with a learned forward operator". *IEEE Transactions on Computational Imaging* 8 (2022), pp. 536–550 (cit. on p. 35).

[270] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso. "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu. Springer, 2017, pp. 240–248 (cit. on pp. 48, 77).

[271] C. Sun, A. Shrivastava, S. Singh, and A. Gupta. "Revisiting unreasonable effectiveness of data in deep learning era". In: *Proceedings of the IEEE International Conference on Computer Vision*. (Venice, Italy). 2017, pp. 843–852 (cit. on p. 35).

[272] D. Sun. *Hyperspectral imaging for food quality analysis and control*. San Diego, California, USA: Elsevier, 2010 (cit. on p. 66).

[273] W. Sun and Q. Du. "Hyperspectral band selection: A review". *IEEE Geoscience and Remote Sensing Magazine* 7.2 (2019), pp. 118–139 (cit. on pp. 66, 67).

[274] K. Taguchi, I. Blevis, and K. Iniewski. *Spectral, photon counting computed tomography: Technology and applications*. Boca Raton, FL, USA: CRC Press, 2020 (cit. on pp. 22, 23, 25, 27, 28, 40, 61, 114).

[275] S. Tairi, S. Anthoine, C. Morel, and Y. Boursier. "Simultaneous reconstruction and separation in a spectral CT framework". In: *Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD), 2016*. (Strasbourg, France). IEEE. 2016, pp. 1–4 (cit. on p. 110).

[276] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding. "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation". *Medical Image Analysis* (2020), p. 101693 (cit. on p. 41).

[277] L. Tanzi, E. Vezzetti, R. Moreno, and S. Moos. "X-ray bone fracture classification using deep learning: A baseline for designing a reliable approach". *Applied Sciences* 10.4 (2020), p. 1507 (cit. on p. 35).

[278] K. Thilagavathi and A. Vasuki. "Dimension reduction methods for hyperspectral image: A survey". *International Journal of Engineering and Advanced Technology* 8 (Dec. 2018), pp. 160–167 (cit. on p. 68).

[279] R. Vaddi and M. Prabukumar. "Comparative study of feature extraction techniques for hyperspectral remote sensing image classification: A survey". In: *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*. (Madurai, India). IEEE. 2017, pp. 543–548 (cit. on p. 67).

[280] D. Valsesia and E. Magli. "High-throughput onboard hyperspectral image compression with ground-based CNN reconstruction". *IEEE Transactions on Geoscience and Remote Sensing* 57.12 (2019), pp. 9544–9553 (cit. on pp. 68, 72).

[281] J. A. T. Vasquez, R. Scapaticci, G. Turvani, M. Ricci, L. Farina, A. Litman, M. R. Casu, L. Crocco, and F. Vipiana. "Noninvasive inline food inspection via microwave imaging technology: An application example in the food industry". *IEEE Antennas and Propagation Magazine* 62.5 (2020), pp. 18–32 (cit. on p. 4).

[282] M. Veale, P. Seller, M. Wilson, and E. Liotti. "HEXITEC: A high-energy X-ray spectroscopic imaging detector for synchrotron applications". *Synchrotron Radiation News* 31.6 (2018), pp. 28–32 (cit. on pp. 21, 114).

[283] G. Vilches-Freixas, V. T. Taasti, L. P. Muren, J. B. B. Petersen, J. M. Létang, D. C. Hansen, and S. Rit. "Comparison of projection-and image-based methods for proton stopping power estimation using dual energy CT". *Physics and Imaging in Radiation Oncology* 3 (2017), pp. 28–36 (cit. on p. 25).

[284] A. S. Wang and N. J. Pelc. "Spectral photon counting CT: Imaging algorithms and performance assessment". *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020), pp. 453–464 (cit. on p. 28).

[285] G. Wang, J. C. Ye, and B. De Man. "Deep learning for tomographic image reconstruction". *Nature Machine Intelligence* 2.12 (2020), pp. 737–748 (cit. on p. 35).

[286] M. Wang. *Industrial tomography: Systems and applications*. Sawston, Cambridge, UK: Elsevier, 2015 (cit. on pp. 14, 15).

[287] X. Wang, Y. Bouzembrak, A. O. Lansink, and H. van der Fels-Klerx. "Application of machine learning to the monitoring and prediction of food safety: A review". *Comprehensive Reviews in Food Science and Food Safety* 21.1 (2022), pp. 416–434 (cit. on p. 3).

[288] R. Warr, E. Ametova, R. J. Cernik, G. Fardell, S. Handschuh, J. S. Jørgensen, E. Papoutsellis, E. Pasca, and P. J. Withers. "Enhanced hyperspectral tomography for bioimaging by spatiospectral reconstruction". *Scientific Reports* 11.1 (2021), pp. 1–13 (cit. on p. 28).

[289] T. Weidinger, T. M. Buzug, T. Flohr, S. Kappler, and K. Stierstorfer. "Polychromatic iterative statistical material image reconstruction for photon-counting computed tomography". *International Journal of Biomedical Imaging* 2016 (2016) (cit. on p. 132).

[290] R. E. Wendell and A. P. Hurter Jr. "Minimization of a non-separable objective function subject to disjoint constraints". *Operations Research* 24.4 (1976), pp. 643–657 (cit. on pp. 122, 193).

[291] *Wikipedia Commons, EM Spectrum Properties*. `https://upload.wikimedia.org/wikipedia/commons/c/cf/EM_Spectrum_Properties_edit.svg`. Accessed on 19 December 2020 (cit. on p. 5).

[292] M. J. Willemink and P. B. Noël. "The evolution of image reconstruction for CT - From filtered back projection to artificial intelligence". *European Radiology* 29.5 (2019), pp. 2185–2195 (cit. on pp. 18, 27, 28, 119).

[293] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann. "Photon-counting CT: Technical principles and clinical prospects". *Radiology* 289.2 (2018), pp. 293–312 (cit. on p. 22).

[294] K. H. Wilm. "Foreign object detection: Integration in food production". *Food Safety Magazine* 18 (2012), pp. 14–17 (cit. on p. 40).

[295] M. D. Wilson, L. Dummott, D. D. Duarte, F. H. Green, S. Pani, A. Schneider, J. W. Scuffham, P. Seller, and M. C. Veale. "A 10 cm× 10 cm CdTe spectroscopic imaging detector based on the HEXITEC ASIC". *Journal of Instrumentation* 10.10 (2015), P10011 (cit. on p. 96).

[296] P. J. Withers, C. Bouman, S. Carmignato, V. Cnudde, D. Grimaldi, C. K. Hagen, E. Maire, M. Manley, A. Du Plessis, and S. R. Stock. "X-ray computed tomography". *Nature Reviews Methods Primers* 1.1 (2021), pp. 1–21 (cit. on pp. 28, 33).

[297] H. Wu, Q. Liu, and X. Liu. "A review on deep learning approaches to image classification and object segmentation". *Computers, Materials & Continua* 60.2 (2019), pp. 575–597 (cit. on p. 40).

[298] W. Wu, P. Chen, V. V. Vardhanabhuti, W. Wu, and H. Yu. "Improved material decomposition with a two-step regularization for spectral CT". *IEEE Access* 7 (2019), pp. 158770–158781 (cit. on pp. 118, 119).

[299] W. Wu, P. Chen, S. Wang, V. Vardhanabhuti, F. Liu, and H. Yu. "Image-domain material decomposition for spectral CT using a generalized dictionary learning". *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020) (cit. on pp. 118, 119).

[300] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. "Deep learning computed tomography". In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. (Athens, Greece). Springer. 2016, pp. 432–440 (cit. on p. 35).

[301] *X-ray Mass Attenuation Coefficients, NIST*. `https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients`. Accessed on 24 February 2020 (cit. on pp. 123, 130, 184, 197).

[302] W. Xia, W. Wu, S. Niu, F. Liu, J. Zhou, H. Yu, G. Wang, and Y. Zhang. "Spectral CT reconstruction - ASSIST: Aided by Self-Similarity in Image-Spectral Tensors". *IEEE Transactions on Computational Imaging* 5.3 (2019), pp. 420–436 (cit. on p. 28).

[303] Z. Xiong, D. Sun, H. Pu, W. Gao, and Q. Dai. "Applications of emerging imaging techniques for meat quality and safety detection and evaluation: A review". *Critical Reviews in Food Science and Nutrition* 57.4 (2017), pp. 755–768 (cit. on p. 40).

[304] X. Yang, V. De Andrade, W. Scullin, E. L. Dyer, N. Kasthuri, F. De Carlo, and D. Gürsoy. "Low-dose X-ray tomography through a deep convolutional neural network". *Scientific Reports* 8.1 (2018), pp. 1–13 (cit. on p. 35).

[305] M. Yaqoob, S. Sharma, and P. Aggarwal. "Imaging techniques in agro-industry and their applications, a review". *Journal of Food Measurement and Characterization* 15.3 (2021), pp. 2329–2343 (cit. on pp. 3, 4).

[306] S. Yu, S. Jia, and C. Xu. "Convolutional neural networks for hyperspectral image classification". *Neurocomputing* 219 (2017), pp. 88–98 (cit. on p. 77).

[307] M. T. Zeegers. *A collection of 131 CT datasets of pieces of modeling clay containing stones*. Zenodo, `https://doi.org/10.5281/zenodo.5866228`. 2022 (cit. on p. 63).

[308] M. T. Zeegers. *A collection of X-ray projections of 131 pieces of modeling clay containing stones for machine learning-driven object detection*. Zenodo, `https://doi.org/10.5281/zenodo.5681008`. 2022 (cit. on p. 63).

[309] M. T. Zeegers. *mzeegers/DeepFODDataGenerator: DeepFODDataGenerator*. Zenodo, `https://doi.org/10.5281/zenodo.7825045`. Version v1.0.1. 2023 (cit. on p. 63).

[310] M. T. Zeegers. *mzeegers/DRCNN: Data Reduction CNN*. Zenodo, `https://doi.org/10.5281/zenodo.7824969`. Version v1.0.1. 2023 (cit. on p. 93).

[311] M. T. Zeegers. *mzeegers/MC-DART: MC-DART*. Zenodo, `https://doi.org/10.5281/zenodo.7824878`. Version v1.0.1. 2023 (cit. on p. 111).

[312] M. T. Zeegers, F. Lucka, and K. J. Batenburg. "A Multi-Channel DART algorithm". In: *International Workshop on Combinatorial Image Analysis*. (Porto, Portugal). Ed. by R. P. Barneva, V. E. Brimkov, and J. M. R. S. Tavares. Springer, 2018, pp. 164–178 (cit. on pp. 61, 95, 119, 209).

[313] M. T. Zeegers, D. M. Pelt, T. van Leeuwen, R. van Liere, and K. J. Batenburg. "Task-driven learned hyperspectral data reduction using end-to-end supervised deep learning". *Journal of Imaging* 6.12 (2020), p. 132 (cit. on pp. 65, 209).

[314] M. T. Zeegers, T. van Leeuwen, D. M. Pelt, S. B. Coban, R. van Liere, and K. J. Batenburg. "A tomographic workflow to enable deep learning for X-ray based foreign object detection". *Expert Systems with Applications* 206 (2022), p. 117768 (cit. on pp. 39, 209).

[315] M. T. Zeegers, A. Kadu, T. van Leeuwen, and K. J. Batenburg. "ADJUST: A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography". *Inverse Problems* 38.12 (2022), p. 125002 (cit. on pp. 113, 209).

[316] T. Zhang, S. Zhao, X. Ma, A. P. Cuadros, Q. Zhao, and G. R. Arce. "Nonlinear reconstruction of coded spectral X-ray CT based on material decomposition". *Optics Express* 29.13 (2021), pp. 19319–19339 (cit. on p. 119).

[317] X. Zhang, Y. Zheng, W. Liu, and Z. Wang. "A hyperspectral image classification algorithm based on atrous convolution". *EURASIP Journal on Wireless Communications and Networking* 2019.1 (2019), pp. 1–12 (cit. on p. 72).

[318] Y. Zhang, X. Mou, G. Wang, and H. Yu. "Tensor-based dictionary learning for spectral CT reconstruction". *IEEE Transactions on Medical Imaging* 36.1 (2016), pp. 142–154 (cit. on pp. 118, 119).

[319] Y. Zhang and H. Yu. "Convolutional neural network based metal artifact reduction in X-ray computed tomography". *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1370–1381 (cit. on p. 35).

[320] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, and G. Wang. "Spectral CT reconstruction with image sparsity and spectral mean". *IEEE Transactions on Computational Imaging* 2.4 (2016), pp. 510–523 (cit. on pp. 27, 28).

[321] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. "Pyramid scene parsing network". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (Honolulu, HI, USA). IEEE, 2017, pp. 2881–2890 (cit. on p. 43).

[322] Z. Zhao, P. Zheng, S. Xu, and X. Wu. "Object detection with deep learning: A review". *IEEE Transactions on Neural Networks and Learning Systems* 30.11 (2019), pp. 3212–3232 (cit. on p. 40).

[323] J. Zhong, F. Zhang, Z. Lu, Y. Liu, and X. Wang. "High-speed display-delayed planar X-ray inspection system for the fast detection of small fishbones". *Journal of Food Process Engineering* 42.3 (2019), e13010 (cit. on p. 40).

[324] S. Zhong. "Progress in terahertz nondestructive testing: A review". *Frontiers of Mechanical Engineering* 14.3 (2019), pp. 273–281 (cit. on p. 4).

[325] Z. Zhong, W. J. Palenstijn, J. Adler, and K. J. Batenburg. "EDS tomographic reconstruction regularized by total nuclear variation joined with HAADF-STEM tomography". *Ultramicroscopy* 191 (2018), pp. 34–43 (cit. on p. 28).

[326] T. Zhou, S. Ruan, and S. Canu. "A review: Deep learning for medical image segmentation using multi-modality fusion". *Array* 3 (2019), p. 100004 (cit. on p. 35).

[327] L. Zhu, P. Spachos, E. Pensini, and K. N. Plataniotis. "Deep learning and machine vision for food processing: A survey". *Current Research in Food Science* 4 (2021), pp. 233–249 (cit. on pp. 3, 31, 33, 35, 40).

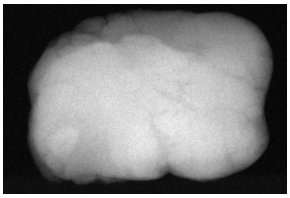
[328] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg. "TVR-DART: A more robust algorithm for discrete tomography from limited projection data with automated gray value estimation". *IEEE Transactions on Image Processing* 25.1 (2016), pp. 455–468 (cit. on pp. 96, 154).

[329] A. Ziabari, S. Venkatakrishnan, M. Kirka, P. Brackman, R. Dehoff, P. Bingham, and V. Paquit. "Beam hardening artifact reduction in X-ray CT reconstruction of 3D printed metal parts leveraging deep learning and CAD models". In: *ASME International Mechanical Engineering Congress and Exposition.* Vol. 84492. American Society of Mechanical Engineers. 2020, V02BT02A043 (cit. on p. 35).

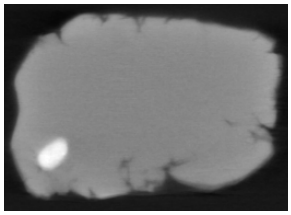
A

Appendices to Chapter 2

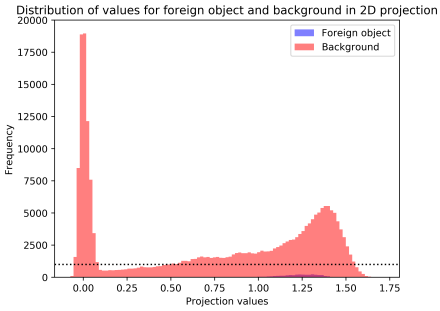
A.1 Intensity value histograms



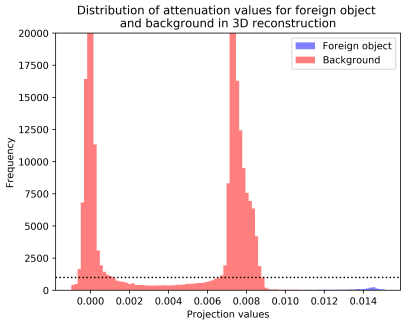
(**a**) 2D radiograph with foreign object on bottom left



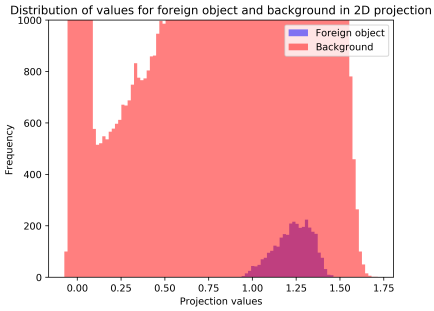
(**b**) Slice of the reconstructed 3D volume with foreign object on bottom left



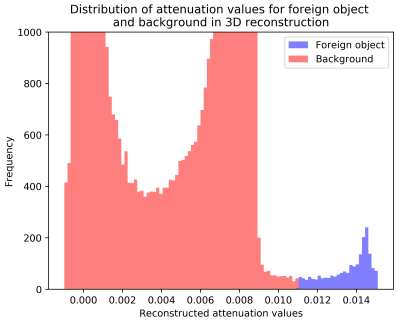
(**c**) Intensity value distribution for the 2D radiograph



(**d**) Attenuation value distribution for the slice of the reconstructed 3D object



(**e**) Intensity value distribution for the 2D radiograph (zoomed)



(**f**) Attenuation value distribution for the slice of the reconstructed 3D object (zoomed)

Figure A1: Radiograph of an object containing a foreign object (**a**) and a slice of the corresponding 3D reconstruction showing its attenuation values (**d**), indicating the difference in contrast. Additionally, histograms of intensity value distribution of the radiograph (**b**-**d**) and the attenuation value distribution of the slice of the reconstructed 3D object (**e**-**f**). In both cases, the histograms of the voxels or pixels of the foreign object are plotted separately from the other voxels or pixels. In the 3D volume, the foreign object is much easier to distinguish based on intensity values.

We compare the intensity distributions for radiographs and for a CT reconstruction of an object in Figure A1, which shows a number of statistics about the pixel and voxel intensities for object 3 (Fig. 2.6). For both approaches, the intensity value distributions are plotted and separated into values of pixel or voxels that have been marked as foreign object by the thresholding method. The 3D case has a clear separation between foreign object and the base object based on attenuation, such that a simple global threshold based on Otsu's method [214] is sufficient to segment the foreign object. On the other hand, in the 2D radiograph case, the intensity values corresponding to the foreign object locations are similar to values of the base object.

A.2 Reconstruction and ground truth similarities

In Section 2.4.8, it is verified that the direct use of generated 3D volumes results in similar ground truth projections compared to the use of the workflow, by indicating that the average Jaccard index between the ground truth pairs is 0.961. In Table A.1, the results are given in greater detail by splitting the results up for nonidentical and combined projections. In addition, we also give the MSE. We also present the similarity results for the segmentations from which the projections are generated. Lastly, results are given for the FDK and SIRT reconstruction algorithms, the latter with 200 iterations. The results indicate that by using these reconstruction algorithms the similarities between the projection pairs increase.

		Identical (%)	Jaccard remaining	Jaccard overall	MSE remaining	MSE overall
Segmentation	FDK	15	0.625	0.681	$1.34 \cdot 10^{-4}$	$1.14 \cdot 10^{-4}$
	SIRT, 100 it.	1	0.656	0.659	$1.18 \cdot 10^{-4}$	$1.17 \cdot 10^{-4}$
	SIRT, 200 it.	10	0.643	0.678	$1.28 \cdot 10^{-4}$	$1.15 \cdot 10^{-4}$
Projection	FDK	47.05	0.981	0.990	$1.57 \cdot 10^{-4}$	$0.83 \cdot 10^{-4}$
	SIRT, 100 it.	8.99	0.957	0.961	$2.97 \cdot 10^{-4}$	$3.26 \cdot 10^{-4}$
	SIRT, 200 it.	39.40	0.980	0.988	$1.59 \cdot 10^{-4}$	$0.94 \cdot 10^{-4}$

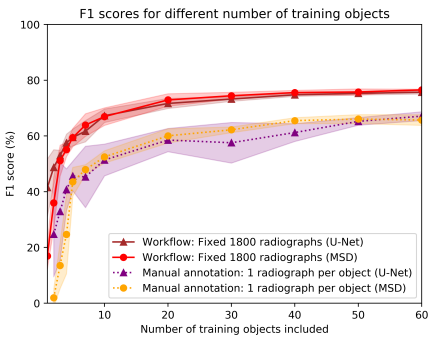
Table A.1: Similarity between ground truth volumes and the corresponding segmented volumes reconstructed from their own projections, as well as similarity between subsequent virtual projections of these volumes. The reconstruction are made over 1800 equidistant angles, and the results are averaged over these angles and 100 training objects. We measure the number of volumes that are identical to their ground truth, and the Jaccard index and the Mean Square Error (MSE) of both all examples and the nonidentical examples only.

Therefore, by adding in an even better reconstruction algorithm in terms of these similarities, an even more accurate training training set can be generated which can subsequently yield more accurate detection results than presented in Chapter 2.

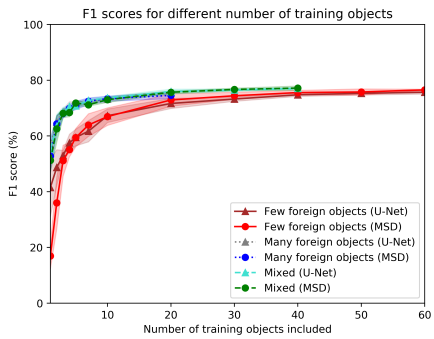
A.3 Additional quality measure

In this Appendix we show the F1 scores for all experiments in Chapter 2. The F1 score is given by

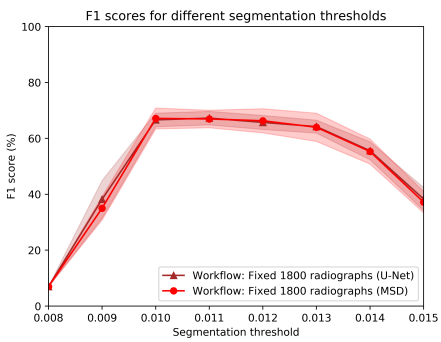
$$\left(\frac{2\mathrm{TP}_{\mathrm{FO}}}{2\mathrm{TP}_{\mathrm{FO}} + \mathrm{FN}_{\mathrm{BG}} + \mathrm{FN}_{\mathrm{FO}}}\right). \tag{A.1}$$



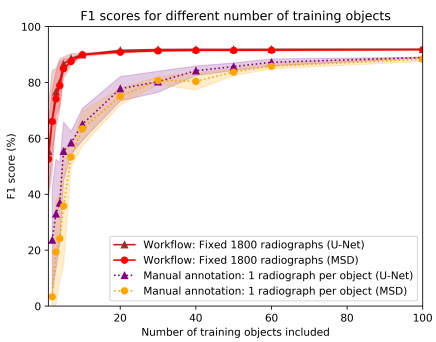
(**a**) Few foreign objects experiments



(**b**) Mixed and many foreign objects experiments



(**c**) Segmentation threshold experiments



(**d**) Simulated data experiments

Figure A2: F1 scores for the various experiments in Chapter 2: (**a**) the standard experiment with laboratory data with few foreign objects, (**b**) the experiment with many and mixed amounts of foreign objects, (**c**) the experiment with threshold variation and (**d**) the simulated experiment. The results are shown for trained U-Net and MSD networks. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

The results for this quality metric are given in Figure A2. The graphs for all experiments are consistent with the graphs for the quality measures in Section 2.4.

B

Appendices to Chapter 3

B.1 Standard data reduction methods

In this section we briefly summarize the most common data reduction methods, used for comparison in this work: PCA (unsupervised), NMF (unsupervised) and LDA (supervised). Let $\boldsymbol{X} \in \mathbb{R}^{N_b \times (N_{\text{train}} \cdot m \cdot n)}$ be a matrix representation of $\{x_i\}_{i=1}^{N_{\text{train}}}$, where the rows represent the spectral features and the columns are the data points.

B.1.1 Principal Component Analysis

Let $\widetilde{\boldsymbol{X}}$ be the centered version of data matrix $\boldsymbol{X}$ where the means of all features are shifted to zero. Principal Component Analysis (PCA) is an unsupervised method that attempts to reduce the data $\widetilde{\boldsymbol{X}}$ to $\overline{\boldsymbol{X}} \in \mathbb{R}^{N_r \times (N_{\text{train}} \cdot m \cdot n)}$, with $N_r < N_b$ the number of components, by finding an orthogonal vector $\boldsymbol{w}$ with $||\boldsymbol{w}|| = 1$ such that the projected data $\widetilde{\boldsymbol{X}}\boldsymbol{w}$ has the highest variance. The maximization $\text{Var}(\widetilde{\boldsymbol{X}}\boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{C} \boldsymbol{w}$ yields the largest eigenvalues of the covariance matrix $\boldsymbol{C} = \text{Cov}(\widetilde{\boldsymbol{X}})$. Therefore the data matrix $\widetilde{\boldsymbol{X}}$ is multiplied by the matrix $\boldsymbol{W}$, containing the N_r largest eigenvalues of $\boldsymbol{C}$, to give $\overline{\boldsymbol{X}} = \widetilde{\boldsymbol{X}}\boldsymbol{W}$. Denote the final transformation of PCA derived from data $\widetilde{\boldsymbol{X}}$ to N_r components by $T_{\boldsymbol{X}}^{\text{PCA}_{N_r}}$. If PCA is chosen to reduce the data to N_r number of bins, then the optimization problem (3.2) becomes

$$\min_F \sum_{i=0}^{N^{\text{train}}} L(F(T_{\boldsymbol{X}}^{\text{PCA}_{N_r}}(x_i^{\text{train}})), y_i^{\text{train}})$$

B.1.2 Non-Negative Matrix Factorization

Let $\boldsymbol{X}^* = \boldsymbol{X} - \min(\boldsymbol{X})$ be the nonnegative matrix version of $\boldsymbol{X}$. In Non-Negative Matrix Factorization (NMF) an attempt is made to factorize the non-negative data matrix $\boldsymbol{X}^*$ into two matrices $\boldsymbol{W} \in \mathbb{R}^{N_b \times N_r}$ and $\boldsymbol{H} \in \mathbb{R}^{N_r \times (N_{\text{train}} \cdot m \cdot n)}$ in an unsupervised manner such that $\boldsymbol{X}^* = \boldsymbol{W}\boldsymbol{H}$. The matrix $\boldsymbol{H}$ will then contain the data points compressed to N_r bins, while $\boldsymbol{W}$ describes the transformation of this matrix to recover the original data matrix $\boldsymbol{X}^*$. Since the problem is not solvable in general, the matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ are often approximated numerically by solving the minimization problem:

$$\min_{\boldsymbol{W},\boldsymbol{H}} ||\boldsymbol{X}^* - \boldsymbol{W}\boldsymbol{H}||_N^2$$

where $||\cdot||_N$ is usually the Frobenius norm. Denote the transformation of NMF derived from data $\boldsymbol{X}$ to N_r components by $T_{\boldsymbol{X}}^{\text{NMF}_{N_r}}$. Similar to PCA, if NMF is chosen to reduce the data to N_r number of bins, the optimization problem (3.2) becomes

$$\min_F \sum_{i=0}^{N^{\text{train}}} L(F(T_{\boldsymbol{X}}^{\text{NMF}_{N_r}}(x_i^{\text{train}})), y_i^{\text{train}})$$

B.1.3 Linear Discriminant Analysis

Let $\boldsymbol{Y} \in \mathcal{C}^{N_{\text{train}} \cdot m \cdot n}$ be the vector representation of $\{y_i\}_{i=1}^{N_{\text{train}}}$. Linear Discriminant Analysis (LDA) seeks to find a transformation W of the data such that ratio of the between-class scatter matrix $S_B(\boldsymbol{X}, \boldsymbol{Y})$ and within-class scatter matrix $S_w(\boldsymbol{X}, \boldsymbol{Y})$ is minimized:

$$\min_{\boldsymbol{W}} \frac{|\boldsymbol{W}^T S_B(\boldsymbol{X}, \boldsymbol{Y}) \boldsymbol{W}|}{|\boldsymbol{W}^T S_w(\boldsymbol{X}, \boldsymbol{Y}) \boldsymbol{W}|}$$

Intuitively, the data are projected on a lower-dimensional space that maximally separates the means of the projected class data points, while minimizing the variances within each class. Similar to PCA, this leads to an eigenvalue problem. Note that since the rank of between-class scatter matrix is at most $C - 1$, where $C = |\mathcal{C}|$ is the number of different classes in the target data $\boldsymbol{Y}$, the rank of $\boldsymbol{W}$ is at most $C - 1$ as well. This means that LDA can reduce the data to at most $N_r = |\mathcal{C}| - 1$ bins. Denote the transformation of NMF derived from data $\boldsymbol{X}$ to N_r components by $T_{\boldsymbol{X},\boldsymbol{Y}}^{\text{LDA}_{N_r}}$. Similar to the previous methods, if LDA is chosen to reduce the data to $N_r < C$ number of bins, then the optimization problem (3.2) becomes

$$\min_F \sum_{i=0}^{N^{\text{train}}} L(F(T_{\boldsymbol{X},\boldsymbol{Y}}^{\text{LDA}_{N_r}}(x_i^{\text{train}})), y_i^{\text{train}})$$

B.2 X-ray projection data computation

In this appendix, we provide further details on the computation of the simulated X-ray projections. The dataset consists of 100 2D images of size 512×512 with $N_b = 300$ spectral bins. These are simulated X-ray projections of 3D volumes of $1024 \times 1024 \times 1024$ voxels containing 120 cylinders with randomized lengths (uniformly distributed between 0.143 and 1.43 cm), thicknesses (uniformly distributed between 0.044 and 0.11 cm), angles and positions. For a schematic overview of the simulated X-ray setup, we refer to Figure 3.6. A virtual source and a virtual detector of size 1536×1536 are placed in front and behind the object respectively, and we use the ASTRA toolbox to compute the projections from this geometric setup. After this, we downscale the projections to 512×512 for computational efficiency, effectively rescaling the volume size as well. The detector pixel size is chosen to

be $s_{\text{pixel}} = 0.11$ mm, making the detector about 5.6 cm, while the voxel size is chosen to be $s_{\text{voxel}} = 0.11$ mm, making the object size about 3.75 cm. A cone beam geometry is used, where the source is placed 44 cm in front of the the center of the object, while the detector is placed 11 cm behind it. We use the National Institute for Standards and Technology (NIST) [130, 301] attenuation spectra for each associated material to compute for each ray an approximation of the number of photons in energy bin $I(E_i)$ hitting the detector. The computed quantity for each bin i with energy window E_i and $1 \leq i \leq N_b$ is given by the following:

$$I(E_i) = \int_{E_i^{\min}}^{E_i^{\max}} I_0(E) e^{-\int_\ell \mu(x,E)dx} dE \tag{B.1}$$

Here, $E_i^{\min}$ and $E_i^{\max}$ signify the energy range in bin i, $I_0(E)$ photon influx at energy E, ℓ is the ray trajectory and $\mu(x, E)$ is the attenuation at position x at energy E. This is approximated by inserting the assumption that $\mu(x, E)$ can be written as a linear combination of individual material attenuations:

$$\begin{aligned}\mu(x,E) &= \sum_{m\in\mathcal{M}} \mu_m(E)\ \alpha_m(x) \\ &= \sum_{m\in\mathcal{M}} \left(0.01\overline{\mu}_m(E) + 0.99\mu_{\text{polyethylene}}(E)\right)\alpha_m(x)\end{aligned}$$

where $\overline{\mu}_m(E)$ is the attenuation coefficient of material $m \in \mathcal{M}$, with $\mathcal{M}$ being the set of involved materials, and $\alpha_m(x)$ the fraction of material m at position x. Inserting this into (B.1) gives:

$$\begin{aligned}I(E_i) &= \int_{E_i^{\min}}^{E_i^{\max}} I_0(E) e^{-\int_\ell \sum_{m\in\mathcal{M}} \mu_m(E)\ \alpha_m(x)dx} dE \\ &= \int_{E_i^{\min}}^{E_i^{\max}} I_0(E) e^{-\sum_{m\in\mathcal{M}} \mu_m(E) \int_\ell \alpha_m(x)dx} dE\end{aligned}$$

The integral is numerically approximated using the midpoint rule and equally sized integration bins, which gives the following:

$$I(E_i) \approx \sum_{j=1}^{N} I_0(\widetilde{E}_j) e^{-\sum_{m\in\mathcal{M}} \mu_m(\widetilde{E}_j) \int_\ell \alpha_m(x)dx} (E_{i_{\max}} - E_{i_{\min}})$$

where N is the number of integration bins, and $\widetilde{E}_j = E_{i_{\min}} + \frac{2(j-1)+1}{2n}(E_{i_{\min}} + E_{i_{\max}})$ the average energy in the j-th integration bin. The number of integration bins is set to $N = 30$ for this computation. The integral $\int_\ell \alpha_m(x)dx$ is computed using ASTRA.

The photon influx $I_0(E)$ is a product of the source spectrum $\overline{I}_0(E)$ at energy E, the exposure time t and the detector pixel size s_{pixel}:

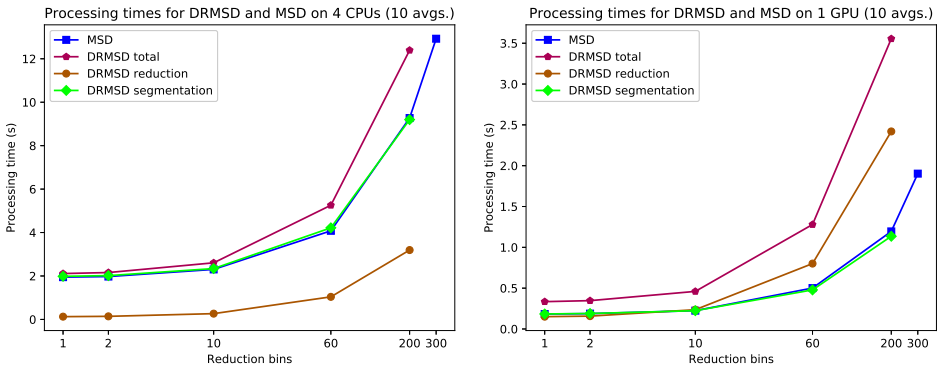
$$I_0(E) = t\overline{I}_0(E) s_{\text{pixel}}^2$$

The exposure time is chosen to be $t = 0.5$ s, and the source spectrum $\overline{I}_0$ is simulated as a radiology source spectrum for a tungsten source without filter at 70 kV, taken from Siemens Healhtineers [258]. The energy range used for this dataset is from $E_{1\min} = 14$ kV to $E_{N_{b_{\max}}} = 69$ kV, and the source spectrum including this range is given in Figure 3.8b. The final projection images in bin i are computed by dividing $I(E_i)$ by the flatfield image $I_{\text{flat}}(E_i)$ containing reference photon counts without objects

$$\frac{I(E_i)}{I_{\text{flat}}(E_i)} = \frac{I(E_i)}{\sum_{j=1}^{N} I_0(\widetilde{E}_j)(E_{i_{\max}} - E_{i_{\min}})}$$

B.3 Time comparison

In this appendix we show the measured processing time for different training setups with MSD on the generated X-ray dataset. Along with the GPU times we also include CPU times, where training is carried out on one Xeon CPU core. The processing times of the trained networks are given in Figure A1. The times for DRMSD are broken down into the data reduction part and the segmentation part. Of course, the times on the CPU cores are higher than those on the GPU core. In both cases the processing time of DRMSD reducing to 1 bin is about 7 to 8 times faster than that of MSD without any data reduction. On the CPUs the DRMSD processing time is comparable to that of MSD, with the difference increasing as the number of bins N_r increases. Note that the number of connections in both networks increase linearly with the number of reduction images N_r. When reducing



(**a**) Processing times on 4 CPUs (**b**) Processing times on 1 GPU

Figure A1: Execution time to apply a forward pass in the trained networks on both CPUs (a) and GPUs (b). The MSD times (in blue) are added as a reference, where each data point indicates the number of input channels. These values are equivalent to the processing time on data reduced by standard methods as PCA, NMF and LDA.

the data to up to 60 bins on the GPU, the DRMSD network is less than 2 times as slow as the network on PCA, NMF and LDA reduced data. For the segmentation part the speedup of the GPU versus CPU is 87%, whereas the speedup for the data reduction part is 24%. Therefore, on the CPU the differences are smaller, but for both CPU and GPU the additional data reduction processing time is acceptable. All in all, the data show that the DRMSD can offer a processing speedup compared to MSD when accomplishing hyperspectral imaging tasks, and this conclusion could hold for some other CNN architectures as well.

B.4 Robustness

Since all experiments in this work are not averaged over multiple runs due time computation time restrictions, we assess in this appendix the stability and robustness of a number of selected experiments. Included in this selection are the experiments where we witnessed the largest variation in the test results. We compute the average, standard deviation, minimum, maximum and median values of the average class accuracy over 8 different runs. The outcomes are given in Table B.1. For

Dataset	Data type	Red. type	Red. chan.	Avg.	Std.	Min.	Max.	Median
X-ray	Noisy + Many materials	DRMSD	2	99.30	0.0883	99.13	99.42	99.31
X-ray	Noisy + Many materials	DRUNet	2	98.87	0.2937	98.36	99.29	98.97
X-ray	Noisy + Many materials	LDA + MSD	2	94.09	0.7639	92.76	95.48	93.91
X-ray	Noisy + Many materials	LDA + U-Net	2	87.24	0.6757	86.10	88.21	87.36
Remote sensing	Noisy + Overlapping	DRMSD	1	95.85	0.3642	95.28	96.34	95.86
Remote sensing	Noisy + Overlapping	DRUNet	1	94.46	0.7250	93.27	95.58	94.65
Remote sensing	Noisy + Overlapping	LDA + MSD	1	59.34	1.2328	56.27	60.60	59.78
Remote sensing	Noisy + Overlapping	LDA + U-Net	1	61.11	0.7674	60.03	62.48	61.05

Table B.1: Average, standard deviation, minimum, maximum and median of the average class accuracy for various network setups, computed over 8 runs.

each experiment, the standard deviation is at most 1.24, and for DRCNN methods this is 0.73. The difference between the minimum and maximum values is at most 3.51, and for DRMSD methods this is 1.38. From these results, we conclude that all the methods presented here are sufficiently stable, and these stability properties may be expected from the other experiments in this research as well.

C
Appendices to Chapter 5

C.1 Proof of Theorem 1

Since the convex set $\mathcal{C} = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}\mathbf{1} \leq \mathbf{1}, \boldsymbol{X}^T\mathbf{1} \leq \mathbf{1}\right\}$ is composed of convex set $\mathcal{C}_1 = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}\mathbf{1} \leq \mathbf{1}\right\}$ and convex set $\mathcal{C}_2 = \left\{\boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, \boldsymbol{X} \geq 0,\, \boldsymbol{X}^T\mathbf{1} \leq \mathbf{1}\right\}$, the indicator function $\delta_{\mathcal{C}}$ (with $\delta_{\mathcal{C}}(\boldsymbol{X}) = 0$ when $\boldsymbol{X} \in \mathcal{C}$ and $\delta_{\mathcal{C}}(\boldsymbol{X}) = \infty$ otherwise) can be expressed as

$$\delta_{\mathcal{C}}(\boldsymbol{X}) = \delta_{\mathcal{C}_1}(\boldsymbol{X}) + \delta_{\mathcal{C}_2}(\boldsymbol{X}).$$

Hence, the projection onto set $\mathcal{C}$ amounts to solving the following minimization problem

$$\mathbf{proj}_{\mathcal{C}}(\boldsymbol{Z}) = \arg\min_{\boldsymbol{X}} \left\{\frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Z}\|_F^2 + \delta_{\mathcal{C}_1}(\boldsymbol{X}) + \delta_{\mathcal{C}_2}(\boldsymbol{X})\right\}.$$

Since the cost function is the composition of two indicator functions, we can redefine a minimization problem by introducing a new slack variable $\boldsymbol{Y}$:

$$\underset{\boldsymbol{X},\boldsymbol{Y}}{\text{minimize}} \left\{\frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Z}\|_F^2 + \delta_{\mathcal{C}_1}(\boldsymbol{X}) + \delta_{\mathcal{C}_2}(\boldsymbol{Y}) + \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Y}\|_F^2\right\},$$

where we have penalized the slack variable $\boldsymbol{Y}$ to stay close to the original variable $\boldsymbol{X}$ using quadratic term. The optimal point of this minimization problem must satisfy the following fixed point equation:

$$\begin{aligned} \boldsymbol{X} - \boldsymbol{Z} + \partial\delta_{\mathcal{C}_1}(\boldsymbol{X}) + \boldsymbol{X} - \boldsymbol{Y} &\in \mathbf{0}, \\ \partial\delta_{\mathcal{C}_2}(\boldsymbol{Y}) + \boldsymbol{Y} - \boldsymbol{X} &\in \mathbf{0}, \end{aligned}$$

where ∂f denotes the sub-gradient of the function f. Hence, the fixed point iteration scheme to find the optimal point leads to

$$\begin{aligned} (\boldsymbol{I} + (1/2)\partial\delta_{\mathcal{C}_1})\,\boldsymbol{X}_{t+1} &= \frac{\boldsymbol{Z} + \boldsymbol{Y}_t}{2}, \\ (\boldsymbol{I} + \partial\delta_{\mathcal{C}_2})\,\boldsymbol{Y}_{t+1} &= \boldsymbol{X}_{t+1}, \end{aligned}$$

for $t = 1, \ldots, T$ with setting $\boldsymbol{Y}_0$ to an arbitrary vector. Since the operation $(\boldsymbol{I} + \alpha\partial\delta_{\mathcal{C}})^{-1}$ with $\alpha > 0$ is equivalent to the definition of proximal operator, we can compactly rewrite the iteration scheme as

$$\begin{aligned} \boldsymbol{Y}_{t+1} &= (\boldsymbol{I} + \partial\delta_{\mathcal{C}_2})^{-1}\left((\boldsymbol{I} + (1/2)\partial\delta_{\mathcal{C}_1})^{-1}\left(\frac{\boldsymbol{Z} + \boldsymbol{Y}_t}{2}\right)\right), \\ &= \mathbf{proj}_{\mathcal{C}_2}\left(\mathbf{proj}_{\mathcal{C}_1}\left(\frac{\boldsymbol{Z} + \boldsymbol{Y}_t}{2}\right)\right). \end{aligned}$$

□

C.2 Bi-convexity of ADJUST and partial optimality

In this section, we show that the optimization problem (5.7) is bi-convex. We start with the definitions related to bi-convexity.

Definition 1 (Bi-convex set)**.** *A set $\mathcal{B} \subset \mathcal{X} \times \mathcal{Y}$ is bi-convex on $\mathcal{X} \times \mathcal{Y}$ if $\mathcal{B}_x = \{y \in \mathcal{Y} : (x, y) \in \mathcal{B}\}$ is convex for every $x \in \mathcal{X}$ and $\mathcal{B}_y = \{x \in \mathcal{X} : (x, y) \in \mathcal{B}\}$ is convex for every $y \in \mathcal{Y}$.*

Definition 2 (Bi-convex function)**.** *A function $\mathcal{F} : \mathcal{B} \to \mathbb{R}$ on a bi-convex set $\mathcal{B} \subseteq \mathcal{X} \times \mathcal{Y}$ is bi-convex if and only if for every fixed y, the function $\mathcal{F}(x, \cdot) : \mathcal{B}_x \to \mathbb{R}$ is convex on $\mathcal{B}_x$, and for every fixed x, the function $\mathcal{F}(\cdot, y) : \mathcal{B}_y \to \mathbb{R}$ is convex on $\mathcal{B}_y$.*

Definition 3 (Bi-convex optimization problem)**.** *A minimization problem of the form*

$$\underset{x,y}{\text{minimize}} \quad \mathcal{F}(x, y) \quad \text{subject to} \quad x, y \in \mathcal{B}$$

is bi-convex if the set $\mathcal{B}$ is bi-convex on $\mathcal{X} \times \mathcal{Y}$ and the objective function $\mathcal{F}$ is bi-convex on $\mathcal{B}$.

Therefore, to show bi-convexity of problem (5.7), we need to show that the constraint set $\mathcal{C}_A \times \mathcal{C}_R$ is bi-convex on $\mathbb{R}^{N\times M} \times \mathbb{R}^{M\times D}$, and the function $\mathcal{J} : \mathbb{R}^{N\times M} \times \mathbb{R}^{M\times D} \to \mathbb{R}$ is a bi-convex function.

Lemma 1. *The set $\mathcal{B} \triangleq \mathcal{C}_A \times \mathcal{C}_R$ is bi-convex on $\mathbb{R}^{N\times M} \times \mathbb{R}^{M\times D}$.*

Proof. Since the set $\mathcal{B}$ is partitioned into two independent sets $\mathcal{C}_A$ and $\mathcal{C}_R$, we only need to show that these sets are convex. The set

$$\mathcal{C}_A = \left\{ \boldsymbol{X} \in \mathbb{R}^{N\times M} \,|\, x_{ij} \geq 0, \sum_{j=1}^{M} x_{ij} \leq 1 \right\}$$

is a convex set on $\mathbb{R}^{N\times M}$ since it is an intersection of the non-negative orthant ($x_{ij} \geq 0$) with N number of hyperplanes ($\sum_{j=1}^{M} x_{ij} \leq 1$) (see 2.2.4 of [45]). Similarly, the set

$$\mathcal{C}_R = \left\{ \boldsymbol{X} \in \mathbb{R}^{M\times D} \,|\, x_{ij} \geq 0, \sum_{j=1}^{D} x_{ij} = 1, \sum_{i=1}^{M} x_{ij} \leq 1 \right\},$$

is a convex set on $\mathbb{R}^{M\times D}$ because it is an intersection of non-negative orthant ($x_{ij} \geq 0$) with M number of hyperplanes ($\sum_{j=1}^{D} x_{ij} = 1$) and D number of halfspaces ($\sum_{i=1}^{M} x_{ij} \leq 1$). Hence, from definition 1, the set $\mathcal{B} = \mathcal{C}_A \times \mathcal{C}_R$ is a bi-convex set on $\mathbb{R}^{N\times M} \times \mathbb{R}^{M\times D}$. □

Lemma 2. *The function $\mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{WART}\|_F^2$ is bi-convex.*

Proof. First, we rewrite the function in the form

$$\begin{aligned}\mathcal{J}(\boldsymbol{A},\boldsymbol{R}) &= \tfrac{1}{2}\|\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T}\|_F^2,\\ &= \tfrac{1}{2}\operatorname{Tr}\big((\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T})(\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T})^T\big) \qquad \|\boldsymbol{X}\|_F^2 = \operatorname{Tr}\big(\boldsymbol{X}\boldsymbol{X}^T\big),\\ &= \tfrac{1}{2}\underbrace{\operatorname{Tr}\big(\boldsymbol{T}^T\boldsymbol{R}^T\boldsymbol{A}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T}\big)}_{\mathcal{P}(\boldsymbol{A},\boldsymbol{R})} - \underbrace{\operatorname{Tr}\big(\boldsymbol{Y}^T\boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T}\big)}_{\mathcal{Q}(\boldsymbol{A},\boldsymbol{R})} + \tfrac{1}{2}\|\boldsymbol{Y}\|_F^2.\end{aligned}$$

Hence, to show that $\mathcal{J}(\boldsymbol{A},\boldsymbol{R})$ is bi-convex, we need to show that $\mathcal{P}(\boldsymbol{A},\boldsymbol{R})$ and $\mathcal{Q}(\boldsymbol{A},\boldsymbol{R})$ are bi-convex.

We first show the bi-convexity of $\mathcal{Q}(\boldsymbol{A},\boldsymbol{R})$. To do so, fix $\overline{\boldsymbol{A}} \in \mathcal{C}_A$. Now, let $\boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathcal{C}_R$ and $\lambda \in (0,1)$. Then we have

$$\begin{aligned}\lambda\mathcal{Q}(\overline{\boldsymbol{A}},\boldsymbol{R}_1) + (1-\lambda)\mathcal{Q}(\overline{\boldsymbol{A}},\boldsymbol{R}_2) &= \lambda\operatorname{Tr}\big(\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_1\boldsymbol{T}\big) + (1-\lambda)\operatorname{Tr}\big(\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_2\boldsymbol{T}\big)\\ &= \operatorname{Tr}\big(\lambda\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_1\boldsymbol{T}\big) + \operatorname{Tr}\big((1-\lambda)\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_2\boldsymbol{T}\big)\\ &= \operatorname{Tr}\big(\lambda\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_1\boldsymbol{T} + (1-\lambda)\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_2\boldsymbol{T}\big)\\ &= \operatorname{Tr}\big(\boldsymbol{Y}^T\boldsymbol{W}\overline{\boldsymbol{A}}(\lambda\boldsymbol{R}_1 + (1-\lambda)\boldsymbol{R}_2)\boldsymbol{T}\big)\\ &= \mathcal{Q}(\overline{\boldsymbol{A}}, \lambda\boldsymbol{R}_1 + (1-\lambda)\boldsymbol{R}_2)\end{aligned}$$

Hence, $\mathcal{Q}(\overline{\boldsymbol{A}},\boldsymbol{R})$ is a convex function over $\mathbb{R}^{M\times D}$ for every $\boldsymbol{A} \in \mathcal{C}_A$. Similarly, fixing $\overline{\boldsymbol{R}} \in \mathcal{C}_R$ and using an analogous deduction as above shows that

$$\lambda\mathcal{Q}(\boldsymbol{A}_1,\overline{\boldsymbol{R}}) + (1-\lambda)\mathcal{Q}(\boldsymbol{A}_2,\overline{\boldsymbol{R}}) = \mathcal{Q}(\lambda\boldsymbol{A}_1 + (1-\lambda)\boldsymbol{A}_2, \overline{\boldsymbol{R}})$$

for every $\boldsymbol{A}_1, \boldsymbol{A}_2 \in \mathcal{C}_A$ and $\lambda \in (0,1)$. Hence, $\mathcal{Q}(\boldsymbol{A},\overline{\boldsymbol{R}})$ is a convex function over $\mathbb{R}^{N\times M}$ for every $\boldsymbol{R} \in \mathcal{C}_R$. This shows that $\mathcal{Q}(\boldsymbol{A},\boldsymbol{R})$ is bi-convex.

Next, we show the bi-convexity of $\mathcal{P}(\boldsymbol{A},\boldsymbol{R})$. Thus, fix $\overline{\boldsymbol{A}} \in \mathcal{C}_A$. Now, to show that $\mathcal{P}(\overline{\boldsymbol{A}},\boldsymbol{R})$ is convex, we use the first-order condition (see 3.1.4 of [45]). Let $\boldsymbol{Q} = \overline{\boldsymbol{A}}^T\boldsymbol{W}^T\boldsymbol{W}\overline{\boldsymbol{A}}$ and $\boldsymbol{P} = \boldsymbol{T}\boldsymbol{T}^T$. The first-order condition states that $\forall\, \boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathbb{R}^{M\times D}$, we need

$$\begin{aligned}\mathcal{P}(\overline{\boldsymbol{A}},\boldsymbol{R}_2) &\geq \mathcal{P}(\overline{\boldsymbol{A}},\boldsymbol{R}_1) + \operatorname{Tr}\big((\boldsymbol{R}_2-\boldsymbol{R}_1)^T\nabla_{\boldsymbol{R}_1}\mathcal{P}(\overline{\boldsymbol{A}},\boldsymbol{R}_1)\big)\\ \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) &\geq \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + 2\operatorname{Tr}\Big((\boldsymbol{R}_2-\boldsymbol{R}_1)^T\overline{\boldsymbol{A}}^T\boldsymbol{W}^T\boldsymbol{W}\overline{\boldsymbol{A}}\boldsymbol{R}_1\boldsymbol{T}\boldsymbol{T}^T\Big)\\ \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) &\geq \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + 2\operatorname{Tr}\big((\boldsymbol{R}_2-\boldsymbol{R}_1)^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big)\end{aligned}$$

To arrive at this condition, let us consider

$$\begin{aligned}&\operatorname{Tr}\big((\boldsymbol{R}_1-\boldsymbol{R}_2)^T\boldsymbol{Q}(\boldsymbol{R}_1-\boldsymbol{R}_2)\boldsymbol{P}\big)\\ &= \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) - \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) - \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big)\\ &= \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) - \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) - \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}^T\boldsymbol{R}_1\boldsymbol{P}^T\big)\\ &= \operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + \operatorname{Tr}\big(\boldsymbol{R}_2^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) - 2\operatorname{Tr}\big(\boldsymbol{R}_1^T\boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big).\end{aligned}$$

Since $\boldsymbol{Q}$ and $\boldsymbol{P}$ are positive semi-definite matrices, we have

$$\mathrm{Tr}\left((\boldsymbol{R}_1 - \boldsymbol{R}_2)^T \boldsymbol{Q}(\boldsymbol{R}_1 - \boldsymbol{R}_2)\boldsymbol{P}\right) \geq 0.$$

Hence, we obtain the following relation

$$\mathrm{Tr}\big(\boldsymbol{R}_1^T \boldsymbol{Q}\boldsymbol{R}_1\boldsymbol{P}\big) + \mathrm{Tr}\big(\boldsymbol{R}_2^T \boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big) \geq 2\,\mathrm{Tr}\big(\boldsymbol{R}_1^T \boldsymbol{Q}\boldsymbol{R}_2\boldsymbol{P}\big),$$

which proves the first-order condition. Similarly, we can show that $\mathcal{P}(\boldsymbol{A}, \overline{\boldsymbol{R}})$ is a convex function over $\mathbb{R}^{N\times M}$ for fixed $\overline{\boldsymbol{R}} \in \mathcal{C}_R$. Hence, $\mathcal{P}(\boldsymbol{A}, \boldsymbol{R})$ is a bi-convex function.

Since $\mathcal{P}(\boldsymbol{A}, \boldsymbol{R})$ and $\mathcal{Q}(\boldsymbol{A}, \boldsymbol{R})$ are bi-convex functions, their linear combination is also a bi-convex function [103]. Hence, we prove that $\mathcal{J}(\boldsymbol{A}, \boldsymbol{R})$ is bi-convex. □

Corollary 1. *The optimization problem (5.7) is bi-convex.*

Proof. Since the cost function $\mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T}\|_F^2$ is bi-convex (Lemma 2) and $\mathcal{C}_A \times \mathcal{C}_R$ is a bi-convex set (Lemma 1), the optimization problem

$$\text{minimize} \quad \mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) \quad \text{subject to} \quad \boldsymbol{A} \in \mathcal{C}_A,\ \boldsymbol{R} \in \mathcal{C}_R$$

is bi-convex (from Definition 3). □

Bi-convex optimization problems may have a large number of local minima as they are global optimization problems in general [103]. Since we are interested in finding a stationary point of (5.7), we define the notion of partial optimality.

Definition 4 (Partial optimality)**.** *Let $\mathcal{F} : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ be a given function and let $(x^\star, y^\star) \in \mathcal{X} \times \mathcal{Y}$. Then, $(x^\star, y^\star)$ is called a partial optimum of $\mathcal{F}$ on $\mathcal{X} \times \mathcal{Y}$, if*

$$\mathcal{F}(x^\star, y^\star) \leq \mathcal{F}(x, y^\star) \quad \forall\, x \in \mathcal{X} \quad \textit{and} \quad \mathcal{F}(x^\star, y^\star) \leq \mathcal{F}(x^\star, y) \quad \forall\, y \in \mathcal{Y}.$$

It is easy to show that a partial optimum $z^\star = (x^\star, y^\star)$ is also a stationary point of $\mathcal{F}$ in $\mathcal{X} \times \mathcal{Y}$ if $\mathcal{F}$ is differentiable at $z^\star$. Also, the converse is true [103]. Finally, the following theorem (adapted from [290]) connects the local optimality (*i.e.* stationary points) to the partial optimality:

Theorem 2. *Let $(\boldsymbol{A}^\star, \boldsymbol{R}^\star) \in \mathcal{C}_A \times \mathcal{C}_R$ be a partial optimum of $\mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{W}\boldsymbol{A}\boldsymbol{R}\boldsymbol{T}\|_F^2$. Furthermore, let $\mathcal{U}(\boldsymbol{R}^\star)$ denote the set of all optimal solutions to (5.7) with $\boldsymbol{R} = \boldsymbol{R}^\star$ and let $\mathcal{V}(\boldsymbol{A}^\star)$ be the set of optimal solutions to (5.7) with $\boldsymbol{A} = \boldsymbol{A}^\star$. If $(\boldsymbol{A}^\star, \boldsymbol{R}^\star)$ is a local optimal solution to (5.7), then it necessarily holds that*

$$\mathcal{J}(\boldsymbol{A}^\star, \boldsymbol{R}^\star) \leq \mathcal{J}(\boldsymbol{A}, \boldsymbol{R}) \quad \forall\, \boldsymbol{A} \in \mathcal{U}(\boldsymbol{R}^\star),\ \boldsymbol{R} \in \mathcal{V}(\boldsymbol{A}^\star).$$

This theorem implies that the natural solution of any alternating minimization algorithm will lead to a partial optimal solution. The proof of the theorem can be found in [290].

C.3 Derivation of AAPM

First, we rephrase the original ADJUST problem in the following form

$$\begin{aligned} \underset{\boldsymbol{A},\boldsymbol{R}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{A},\boldsymbol{R}) + \delta_{\mathcal{C}_A}(\boldsymbol{A}) + \delta_{\mathcal{C}_R}(\boldsymbol{R}), \\ \text{subject to} \quad & \boldsymbol{WART} = \boldsymbol{Y}, \end{aligned}$$

where $\delta_{\mathcal{C}}$ is an extended value function for the constraint set $\mathcal{C}$ that is 0 when constraint is satisfied and ∞ otherwise. Here, we have introduced the constraints on the misfit between simulated and true measurements in the linear form. The Lagrangian for this optimization problem reads

$$\begin{aligned} \mathcal{L}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U}) &= \mathcal{J}(\boldsymbol{A},\boldsymbol{R}) + \delta_{\mathcal{C}_A}(\boldsymbol{A}) + \delta_{\mathcal{C}_R}(\boldsymbol{R}) + \langle \boldsymbol{U}, \boldsymbol{Y} - \boldsymbol{WART}\rangle \\ &= \underbrace{\mathcal{J}(\boldsymbol{A},\boldsymbol{R}) + \langle \boldsymbol{U}, \boldsymbol{Y} - \boldsymbol{WART}\rangle}_{\triangleq \widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U})} + \delta_{\mathcal{C}_A}(\boldsymbol{A}) + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \end{aligned} \tag{C.1}$$

where $\boldsymbol{U} \in \mathbb{R}^{J\times C}$ is a Lagrange multiplier for constraint $\boldsymbol{WART} = \boldsymbol{Y}$. The Lagrange multiplier $\boldsymbol{U}$ can also be thought of as a running-sum-of-error as it captures the misfit between the true measurements and simulated measurements. The goal is to find a saddle point of this Lagrangian, since the saddle point will give the optimal solution to (5.7). The saddle point of the Lagrangian is given by

$$(\boldsymbol{A}^\star, \boldsymbol{R}^\star, \boldsymbol{U}^\star) = \arg\max_{\boldsymbol{U}} \arg\min_{\boldsymbol{A},\boldsymbol{R}} \mathcal{L}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U}).$$

It is important to note that the Lagrangian is non-differentiable due to the presence of $\delta_{\mathcal{C}_A}$ and $\delta_{\mathcal{C}_R}$. Since the min-max problem cannot be solved using a simple gradient-based iterative scheme due to non-differentiability of the Lagrangian, we need to make use of proximal alternating iterative algorithm. To derive such scheme, we approximate the Lagrangian (C.1) near point $(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k)$ using the Taylor series for the differentiable function $\widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U})$. This approximation reads

$$\begin{aligned} \mathcal{L}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U}) &\approx \widetilde{\mathcal{L}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U}|\boldsymbol{A}_k,\boldsymbol{R}_k,\boldsymbol{U}_k) \\ &= \widetilde{\mathcal{J}}(\boldsymbol{A}_k,\boldsymbol{R}_k,\boldsymbol{U}_k) + \\ &\quad \langle \nabla_{\boldsymbol{R}}\widetilde{\mathcal{J}}(\boldsymbol{A}_k,\boldsymbol{R}_k,\boldsymbol{U}_k), \boldsymbol{R}-\boldsymbol{R}_k\rangle + 1/(2\alpha)\|\boldsymbol{R}-\boldsymbol{R}_k\|_F^2 + \\ &\quad \langle \nabla_{\boldsymbol{A}}\widetilde{\mathcal{J}}(\boldsymbol{A}_k,\boldsymbol{R}_k,\boldsymbol{U}_k), \boldsymbol{A}-\boldsymbol{A}_k\rangle + 1/(2\beta)\|\boldsymbol{A}-\boldsymbol{A}_k\|_F^2 + \\ &\quad \delta_{\mathcal{C}_A}(\boldsymbol{A}) + \delta_{\mathcal{C}_R}(\boldsymbol{R}), \end{aligned} \tag{C.2}$$

where α and β are the Lipschitz constants of the partial gradients of $\widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U})$ with respect to $\boldsymbol{A}$ and $\boldsymbol{R}$ respectively. This approximation leads to the following alternating scheme where we minimize with respect to the primal variables $\boldsymbol{A}$ and

$\boldsymbol{R}$, and maximize with respect to the dual variable $\boldsymbol{U}$:

$$\begin{aligned}
\boldsymbol{R}_{k+1} &= \arg\min_{\boldsymbol{R}} \widetilde{\mathcal{L}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U} | \boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k) \\
\boldsymbol{A}_{k+1} &= \arg\min_{\boldsymbol{A}} \widetilde{\mathcal{L}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U} | \boldsymbol{A}_k, \boldsymbol{R}_{k+1}, \boldsymbol{U}_k) \\
\boldsymbol{U}_{k+1} &= \boldsymbol{U}_k + \rho\left(\boldsymbol{W}\boldsymbol{A}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{T} - \boldsymbol{Y}\right)
\end{aligned}$$

with $k = 0, \ldots, K$, and $\rho > 0$ is the acceleration parameter. This alternating scheme requires initial values of $\boldsymbol{R}$ and $\boldsymbol{A}$, while the initial value of $\boldsymbol{U}$ can be set to $\mathbf{0}$. We update the dual variable $\boldsymbol{U}$ using the linearized ascent, a standard technique used by many alternating methods, *e.g.*, alternating direction method of multipliers [46]. Since the approximate Lagrangian (C.2) is composed of quadratic term and non-smooth terms for $\boldsymbol{A}$ and $\boldsymbol{R}$, we can express the iterates using proximal operations. To derive $\boldsymbol{R}$, we use the identity $\|\boldsymbol{X}+\boldsymbol{Y}\|_F^2 = \|\boldsymbol{X}\|_F^2 + \|\boldsymbol{Y}\|_F^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle$, or equivalently, $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle + \frac{1}{2}\|\boldsymbol{Y}\|_F^2 = \frac{1}{2}\|\boldsymbol{X}+\boldsymbol{Y}\|_F^2 - \frac{1}{2}\|\boldsymbol{X}\|_F^2$. The derivation is now as follows:

$$\begin{aligned}
\boldsymbol{R}_{k+1} &= \arg\min_{\boldsymbol{R}} \widetilde{\mathcal{L}}(\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{U} | \boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k), \\
&= \arg\min_{\boldsymbol{R}} \left\{ \langle \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k), \boldsymbol{R} - \boldsymbol{R}_k \rangle + \tfrac{1}{2\alpha}\|\boldsymbol{R} - \boldsymbol{R}_k\|_F^2 + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \right\}, \\
&= \arg\min_{\boldsymbol{R}} \left\{ \tfrac{1}{\alpha}\langle \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k), \boldsymbol{R} - \boldsymbol{R}_k \rangle + \tfrac{1}{2\alpha}\|\boldsymbol{R} - \boldsymbol{R}_k\|_F^2 + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \right\}, \\
&= \arg\min_{\boldsymbol{R}} \left\{ \tfrac{1}{\alpha} \underbrace{\left(\langle \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k), \boldsymbol{R} - \boldsymbol{R}_k \rangle + \tfrac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}_k\|_F^2 \right)}_{\text{applying the identity with } \boldsymbol{X} \triangleq \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k),\ \boldsymbol{Y} \triangleq \boldsymbol{R} - \boldsymbol{R}_k} + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \right\}, \\
&= \arg\min_{\boldsymbol{R}} \left\{ \tfrac{1}{2\alpha}\|\alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k) + \boldsymbol{R} - \boldsymbol{R}_k\|_F^2 \right. \\
&\qquad \left. \underbrace{-\tfrac{1}{2}\|\alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k)\|_F^2}_{\text{independent of } \boldsymbol{R}} + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \right\}, \\
&= \arg\min_{\boldsymbol{R}} \left\{ \tfrac{1}{2\alpha}\|\boldsymbol{R} - \boldsymbol{R}_k + \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k)\|_F^2 + \delta_{\mathcal{C}_R}(\boldsymbol{R}) \right\}, \\
&= \mathbf{prox}_{\delta_{\mathcal{C}_R}}\left(\boldsymbol{R}_k - \alpha \nabla_{\boldsymbol{R}} \widetilde{\mathcal{J}}(\boldsymbol{A}_k, \boldsymbol{R}_k, \boldsymbol{U}_k) \right),
\end{aligned} \tag{C.3}$$

where the proximal for a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ reads

$$\mathbf{prox}_{\gamma f}(\boldsymbol{z}) = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \frac{1}{2\gamma}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + f(\boldsymbol{x}) \right\}$$

with $\gamma > 0$. The proximal operator allows us to work with non-differentiable functions. Moreover, proximal operators for many functions have explicit expressions,

making it a very computationally-friendly tool. The proximal operator for $\delta_{\mathcal{C}}$ with $\mathcal{C} \subset \mathbb{R}^n$ takes the following form:

$$\mathbf{prox}_{\delta_{\mathcal{C}}}(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \delta_{\mathcal{C}}(\boldsymbol{x}) \right\}$$

Indeed, the proximal operator of a $\delta_{\mathcal{C}}$ is just an orthogonal projection of a vector onto the set $\mathcal{C}$. If the set $\mathcal{C}$ is convex, the proximal point is unique. Similar to (C.3), we can explicitly write down the update of $\boldsymbol{A}$ in terms of the proximal operator.

C.4 Gradient computations

Here we show how the gradients are computed at the final comments in Section 5.5.1. We only show the derivation of $\nabla_{\boldsymbol{A}}\widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U})$ since the derivation of $\nabla_{\boldsymbol{R}}\widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U})$ is very similar.

$$\begin{aligned}
&\nabla_{\boldsymbol{A}}\widetilde{\mathcal{J}}(\boldsymbol{A},\boldsymbol{R},\boldsymbol{U}) \\
&\quad = \nabla_{\boldsymbol{A}}\left(\tfrac{1}{2}\|\boldsymbol{Y}-\boldsymbol{WART}\|_F^2 + \langle \boldsymbol{U}, \boldsymbol{Y}-\boldsymbol{WART}\rangle\right) \\
&\quad = \nabla_{\boldsymbol{A}}\left(\tfrac{1}{2}\|\boldsymbol{Y}-\boldsymbol{WART}\|_F^2\right) + \nabla_{\boldsymbol{A}}\langle \boldsymbol{U}, \boldsymbol{Y}-\boldsymbol{WART}\rangle \\
&\quad \overset{*}{=} \tfrac{1}{2}\nabla_{\boldsymbol{A}}\left(\|\boldsymbol{Y}\|_F^2 + \|\boldsymbol{WART}\|_F^2 - 2\operatorname{Tr}\left(\boldsymbol{Y}^T\boldsymbol{WART}\right)\right) \\
&\qquad + \nabla_{\boldsymbol{A}}\operatorname{Tr}\left(\boldsymbol{U}^T(\boldsymbol{Y}-\boldsymbol{WART})\right) && \langle \boldsymbol{X},\boldsymbol{Y}\rangle = \operatorname{Tr}\left(\boldsymbol{A}^T\boldsymbol{B}\right) \\
&\quad = \tfrac{1}{2}\nabla_{\boldsymbol{A}}\left(\|\boldsymbol{WART}\|_F^2\right) - \nabla_{\boldsymbol{A}}\left(\operatorname{Tr}\left(\boldsymbol{Y}^T\boldsymbol{WART}\right)\right) \\
&\qquad + \nabla_{\boldsymbol{A}}\operatorname{Tr}\left(\boldsymbol{U}^T\boldsymbol{Y}\right) - \nabla_{\boldsymbol{A}}\operatorname{Tr}\left(\boldsymbol{U}^T\boldsymbol{WART}\right) \\
&\quad = \tfrac{1}{2}\nabla_{\boldsymbol{A}}\left(\operatorname{Tr}\left(\boldsymbol{T}^T\boldsymbol{R}^T\boldsymbol{A}^T\boldsymbol{W}^T\boldsymbol{WART}\right)\right) - \boldsymbol{W}^T\boldsymbol{YT}^T\boldsymbol{R}^T \\
&\qquad - (\boldsymbol{UTW})^T(\boldsymbol{RT})^T && \frac{\partial}{\partial \boldsymbol{X}}\operatorname{Tr}\left(\boldsymbol{AXB}\right) = \boldsymbol{A}^T\boldsymbol{B}^T \\
&\quad = \boldsymbol{W}^T\left(\boldsymbol{WART}\right)\boldsymbol{T}^T\boldsymbol{R}^T - \boldsymbol{W}^T\left(\boldsymbol{Y}\right)\boldsymbol{T}^T\boldsymbol{R}^T \\
&\qquad - \boldsymbol{W}^T\boldsymbol{UT}^T\boldsymbol{R} \\
&\quad = \boldsymbol{W}^T\left(\boldsymbol{WART}-\boldsymbol{Y}\right)\boldsymbol{T}^T\boldsymbol{R}^T - \boldsymbol{W}^T\boldsymbol{UT}^T\boldsymbol{R}.
\end{aligned}$$

In the third step (*), we use the following identity:

$$\begin{aligned}
\|\boldsymbol{X}-\boldsymbol{Y}\|_F^2 &= \operatorname{Tr}\left((\boldsymbol{X}-\boldsymbol{Y})^T(\boldsymbol{X}-\boldsymbol{Y})\right) \\
&= \operatorname{Tr}\left((\boldsymbol{X}^T-\boldsymbol{Y}^T)(\boldsymbol{X}-\boldsymbol{Y})\right) \\
&= \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{Y}^T\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{Y}^T\boldsymbol{Y}\right) \\
&= \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right) - \operatorname{Tr}\left(\boldsymbol{Y}^T\boldsymbol{X}\right) - \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{Y}\right) \\
&\quad + \operatorname{Tr}\left(\boldsymbol{Y}^T\boldsymbol{Y}\right) \\
&= \|\boldsymbol{X}\|_F^2 + \|\boldsymbol{Y}\|_F^2 - 2\operatorname{Tr}\left(\boldsymbol{Y}^T\boldsymbol{X}\right) && \left(\boldsymbol{Y}^T\boldsymbol{X}\right)^T = \boldsymbol{X}^T\boldsymbol{Y}
\end{aligned}$$

C.5 Dictionary matrix

In this section, we list the 42 materials that are used in the dictionary matrix $\boldsymbol{T}$ for the Disks and Shepp-Logan phantoms. The spectra are retrieved from the National Institute for Standards and Technology (NIST) [130, 301].

Mat. no.	Material name	At. no.
23	Vanadium	23
24	Chromium	24
25	Manganese	25
26	Iron	26
27	Cobalt	27
28	Nickel	28
29	Copper	29
30	Zinc	30
31	Gallium	31
32	Germanium	32
33	Arsenic	33
34	Selenium	34
35	Bromine	35
36	Krypton	36
37	Rubidium	37
38	Strontium	38
39	Yttrium	39
40	Zirconium	40
41	Niobium	41
42	Molybdenum	42
43	Technetium	43

Mat. no.	Material name	At. no.
44	Ruthenium	44
45	Rhodium	45
46	Palladium	46
47	Silver	47
48	Cadmium	48
49	Indium	49
50	Tin	50
51	Antimony	51
52	Tellurium	52
53	Iodine	53
54	Xenon	54
55	Cesium	55
56	Barium	56
57	Lanthanum	57
58	Cerium	58
59	Praseodymium	59
60	Neodymium	60
61	Promethium	61
62	Samarium	62
63	Terbium	63
64	Gadolinium	64

We plot the attenuation spectra for all dictionary elements for each bin within the selected range in Figure A1. Additionally, Figure A2 shows the spectra for a few selected materials. All of these materials have a K-edge in the considered spectral range.

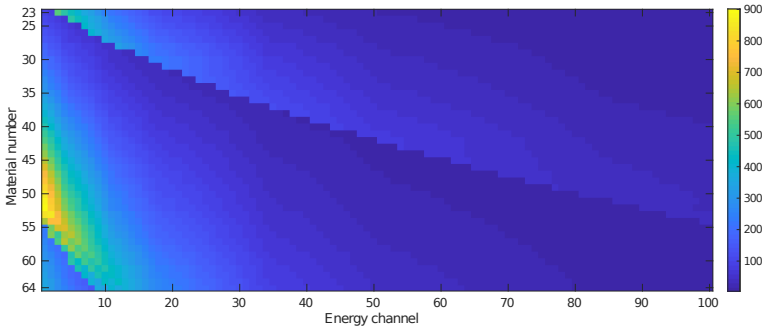


Figure A1: Dictionary matrix $\boldsymbol{T}$: Attenuation values over 100 spectral channels for 42 materials, with energies ranging from 20 keV to 119 keV.

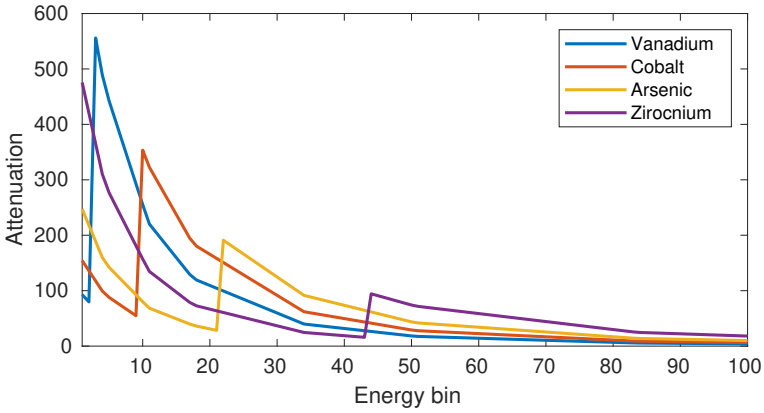


Figure A2: Attenuation values over 100 spectral channels for four materials from the dictionary matrix $\boldsymbol{T}$, with energies ranging from 5 keV to 35 keV.

C.6 Performance measures

To assess the quality of the reconstructions that ADJUST (and the comparison methods) generates, we compare the reconstructions with the ground truth. Since for the UR, RU, cJoint and ADJUST methods the best matching reconstruction of a certain channel in the ground truth may be located in a different channel in the material map matrix, a matching that minimizes the total error over the channels needs to be carried out. Let $\boldsymbol{A}^{\mathrm{GT}} \in \mathbb{R}^{N\times M}$ be the matrix containing the ground truth material maps and $\boldsymbol{A}^{\mathrm{rec}} \in \mathbb{R}^{N\times M}$ be the reconstructed material map. We compute a matrix $\boldsymbol{A}^{\mathrm{error}}$ containing the mutual errors between channels of $\boldsymbol{A}^{\mathrm{GT}}$ and $\boldsymbol{A}^{\mathrm{rec}}$, defined by

$$A_{ij}^{\mathrm{error}} = \left\| (\boldsymbol{A}_{ki}^{\mathrm{rec}})_{i\leq k\leq N} - (\boldsymbol{A}_{kj}^{\mathrm{GT}})_{1\leq k\leq N} \right\|_2$$

Given this error matrix, we use an iterative greedy approach to match the channels of the $\boldsymbol{A}^{\mathrm{GT}}$ and $\boldsymbol{A}^{\mathrm{rec}}$ matrices based on their mutual channel errors. We repeatedly compute the minimum of the error matrix and remove the possibility to match the corresponding channels. To do so, let $\mathcal{M}_0^{\mathrm{GT}} = \mathcal{M}$, $\mathcal{M}_0^{\mathrm{rec}} = \mathcal{M}$ and $\mathcal{M}_0^{\mathrm{match}} = \emptyset$. In each iteration $1 \leq l \leq M$, we compute

$$(i_l, j_l) = \underset{\substack{i \in \mathcal{M}_l^{\mathrm{rec}} \\ j \in \mathcal{M}_l^{\mathrm{GT}}}}{\arg\min} A_{ij}$$

and define $\mathcal{M}_{l+1}^{\mathrm{rec}} = \mathcal{M}_l^{\mathrm{rec}} \backslash \{i_l\}$, $\mathcal{M}_{l+1}^{\mathrm{GT}} = \mathcal{M}_l^{\mathrm{GT}} \backslash \{j_l\}$ and $\mathcal{M}_{l+1}^{\mathrm{match}} = \mathcal{M}_l^{\mathrm{match}} \cup \{(i_l, j_l\}$. Given the final channel-matching represented by $\mathcal{M}_M^{\mathrm{match}}$, we compute the following three error metrics for each $(i,j) \in \mathcal{M}_M^{\mathrm{match}}$:

- *Mean square error* (MSE) for each matched material pair:

$$\mathrm{MSE}(i,j) = \left\| (\boldsymbol{A}_{ki}^{\mathrm{rec}})_{i \leq k \leq N} - (\boldsymbol{A}_{kj}^{\mathrm{GT}})_{1 \leq k \leq N} \right\|_2^2$$

- *Peak signal-to-noise ratio* (PSNR) for each matched material pair:

$$\mathrm{PSNR}(i,j) = 10 \log_{10} \left(\left(\max_k (\boldsymbol{A}_{kj}^{\mathrm{GT}})_{1 \leq k \leq N} \right)^2 / \left\| (\boldsymbol{A}_{ki}^{\mathrm{rec}})_{i \leq k \leq N} - (\boldsymbol{A}_{kj}^{\mathrm{GT}})_{1 \leq k \leq N} \right\|_2^2 \right)$$

- *Structural similarity index* (SSIM) for each matched material pair:

$$\mathrm{SSIM}(i,j) = \big((2\mu_i \mu_j + C_1)(2\sigma_{ij} + C_2) \,/ \big(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2) \big) \big)$$

  with μ_i, μ_j and σ_i, σ_j being the means and the standard deviations of the matrices $(\boldsymbol{A}_{ki}^{\mathrm{rec}})_{i \leq k \leq N}$ and $(\boldsymbol{A}_{kj}^{\mathrm{GT}})_{1 \leq k \leq N}$ respectively, with σ_{ij} being the cross-correlation between these two matrices, and with $C_1 = (0.01L)^2$, $C_2 = (0.03L)^2$ and $L = 1$.

The averages of the MSE, PSNR and SSIM over all materials are then given by:

$$\begin{aligned}
\mathrm{MSE}_{\mathrm{avg}} &= \sum_{(i,j) \in \mathcal{M}_M^{\mathrm{match}}} \mathrm{MSE}(i,j)/M, \\
\mathrm{PSNR}_{\mathrm{avg}} &= \sum_{(i,j) \in \mathcal{M}_M^{\mathrm{match}}} \mathrm{PSNR}(i,j)/M, \\
\mathrm{SSIM}_{\mathrm{avg}} &= \sum_{(i,j) \in \mathcal{M}_M^{\mathrm{match}}} \mathrm{SSIM}(i,j)/M.
\end{aligned}$$

C.7 Numerical studies: Comparison of methods

As stated in the main chapter, we have compared ADJUST with RU, UR, and cJoint on three numerical phantoms, mainly the Shepp-Logan phantom, the Disks phantom, and the Thorax phantom. Figure A3 shows the reconstruction results (*i.e.* reconstructed spatial maps and the spectra of materials) of these methods on Disks phantom. Moreover, we also plot the performance measures of these methods per material in Figure A4.

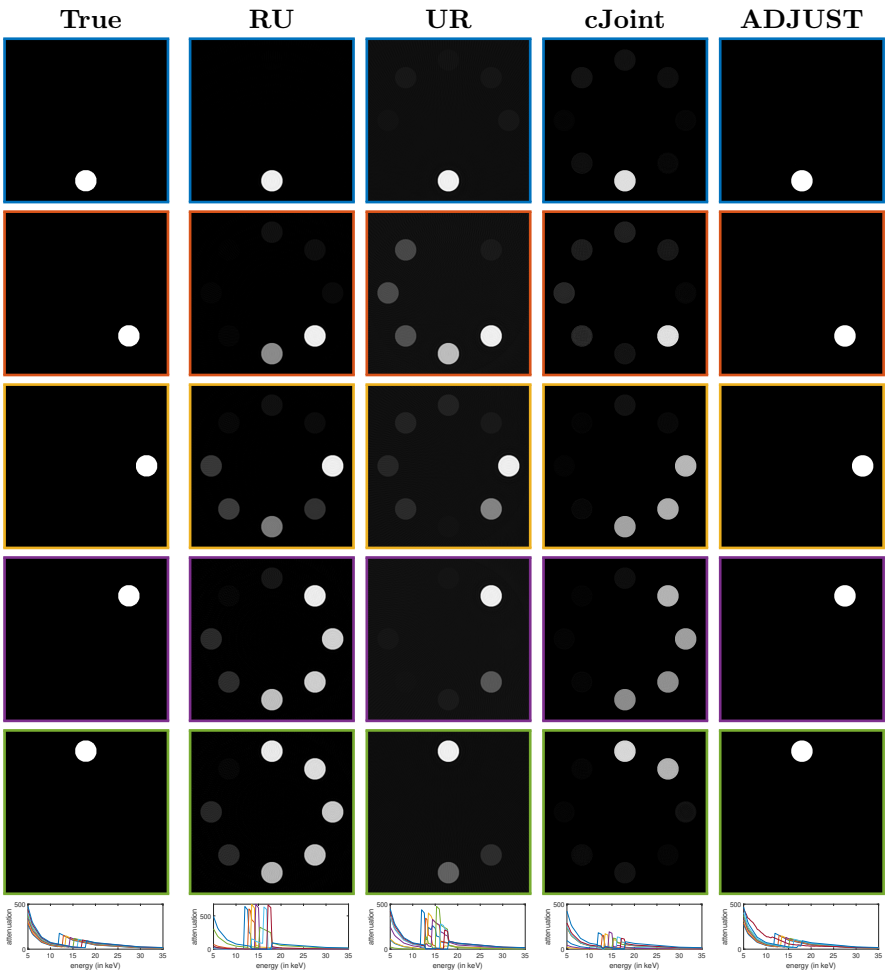


Figure A3: Visual comparison of ADJUST with RU, UR, and cJoint method on the Disks phantom. We only show the reconstructions of all disks here for the comparison. Moreover, we match the colors of the bounding box for material maps with the (recovered) spectral signatures of the materials (shown in the bottom row).

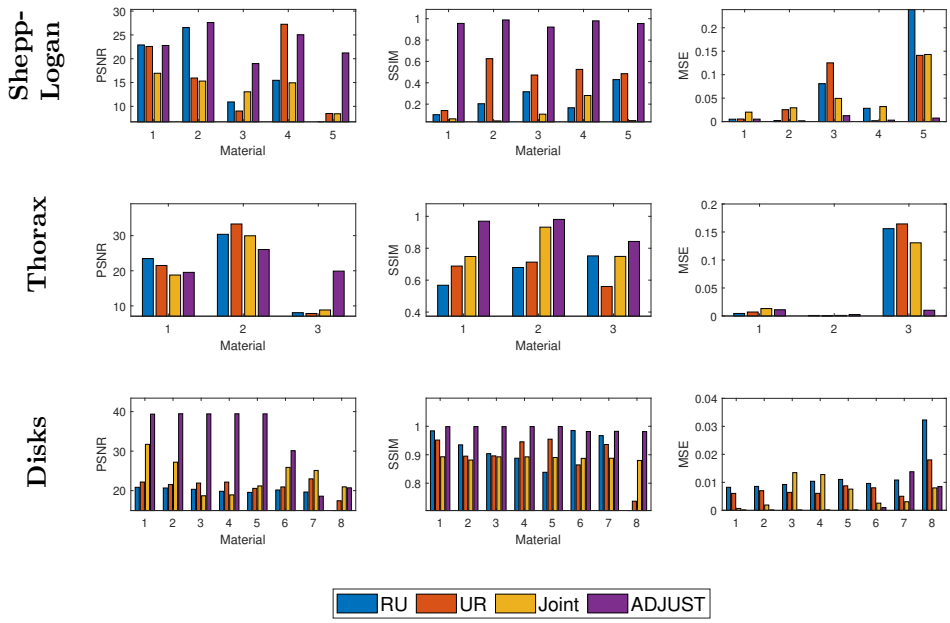


Figure A4: Performance plots showing PSNR (left column), SSIM (middle column), MSE (right column) of the reconstructed materials against the ground truth for various numerical algorithms on the phantoms.

C.8 Numerical studies: Limited measurement patterns

We consider three different types of limited measurement patterns: (i) *Sparse-angle*: tomographic projections from 10 equidistant angles in the range of 0 to π for 100 spectral channels, (ii) *Limited-view*: 60 equidistant projection angles in the limited range of $[0, 2\pi/3]$ for 100 spectral channels, (iii) *Sparse channels*: 60 equidistant angles between 0 and π, but with only 30 spectral channels. We test ADJUST on the two numerical spectral phantoms, *i.e.* the Shepp-Logan phantom and the Disks phantom. Figures A5 and A6 demonstrate the reconstructions of ADJUST for all three limited measurement patterns on these two phantoms.

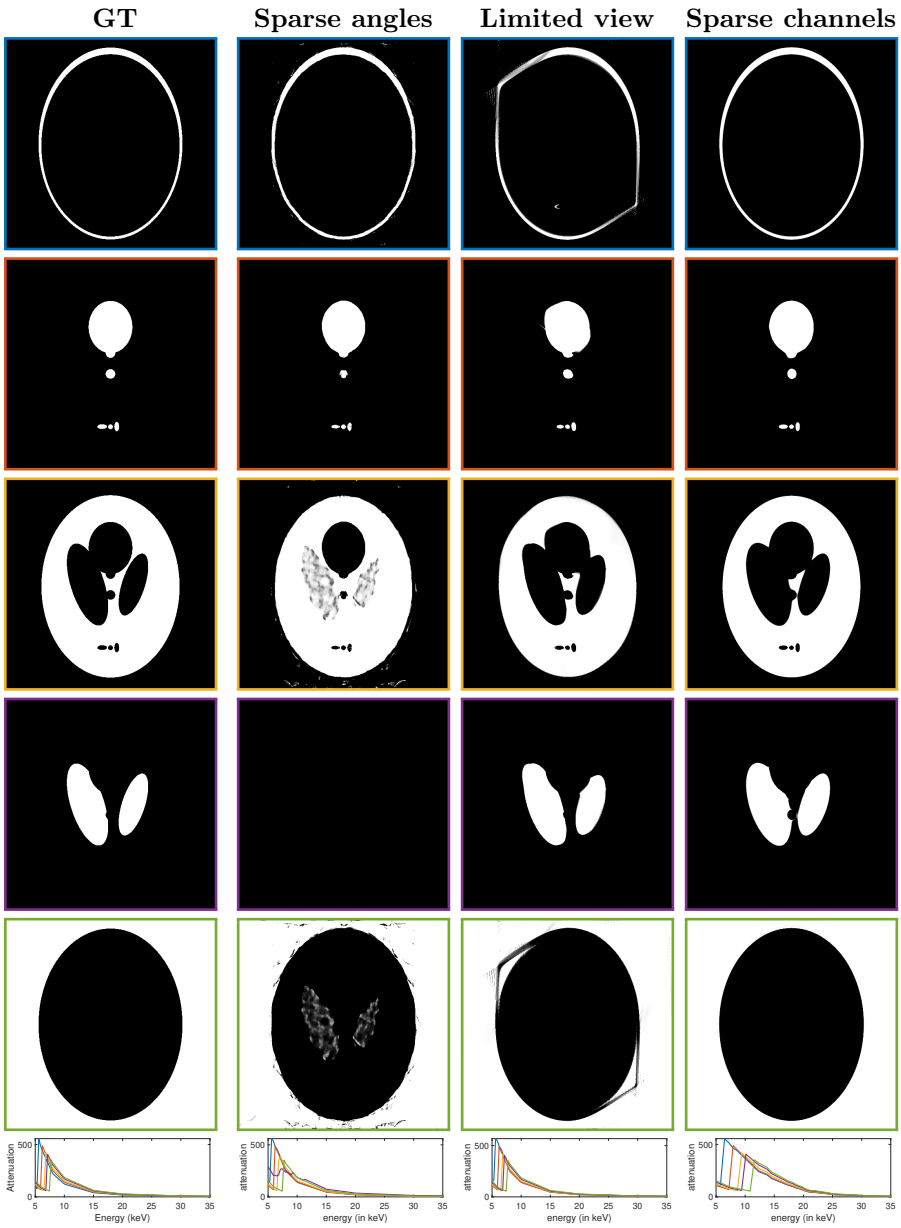


Figure A5: Results of ADJUST with sparse-angle data, limited view data and sparse spectral channels on Shepp-Logan phantom. The colors of the bounding box of material maps are matched with the spectral signatures of the materials (shown in the bottom row).

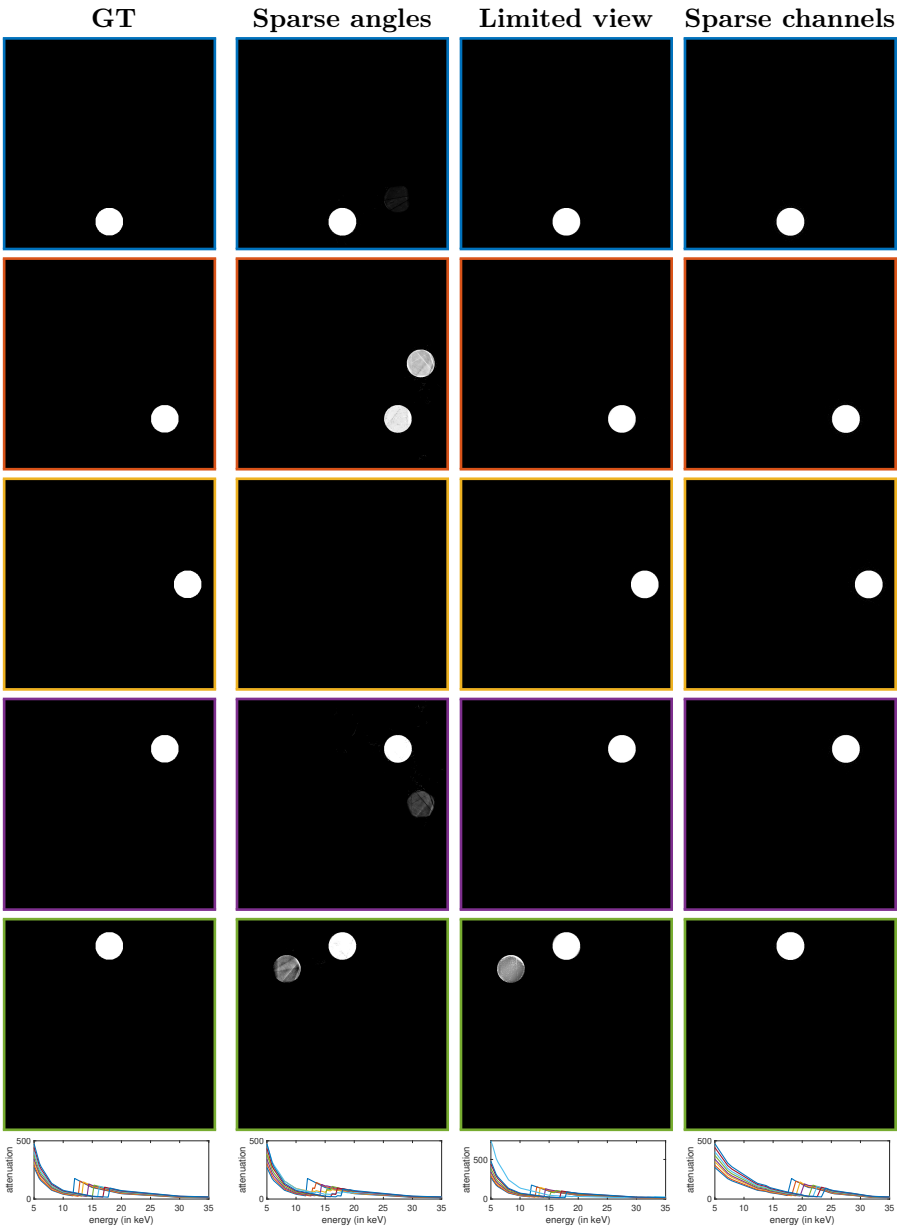


Figure A6: Results of ADJUST with sparse-angle data, limited view data and sparse spectral channels on Disks phantom. We only show material maps of first five materials. The colors of the bounding box of material maps are matched with the spectral signatures of the materials (shown in the bottom row).

C.9 Numerical studies: Mixed material phantom

We consider the Mixed Disks phantom, which consists of solid disks in an inner circle and mixed disks on an outer circle. All material mixtures are present on the outer circle. With $M = 5$ disks on the inner circle, this amounts to 10 mixed disks on the outer circle. The materials are the same as the first 5 selected materials in the Disks phantom. The ADJUST method with 2000 iterations is compared with RU, UR, and cJoint. The results of this experiment are shown in Figure A7, with the results for each material on a separate row.

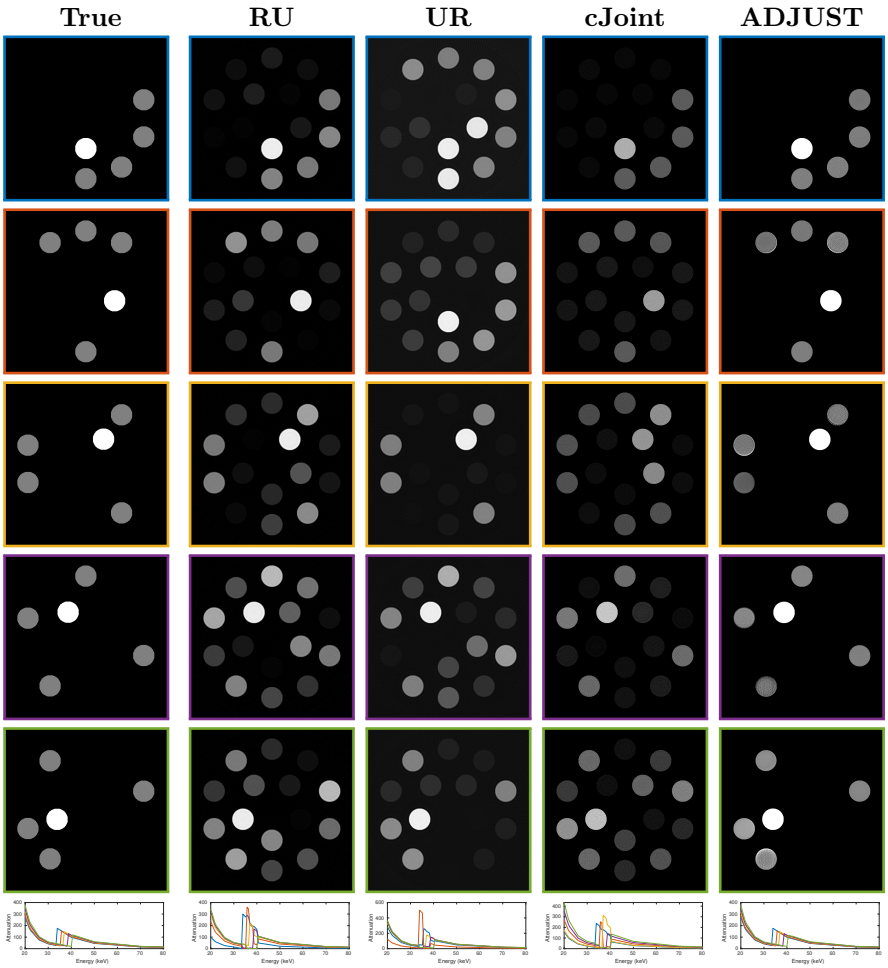


Figure A7: Comparison of various methods for spectral CT for a mixed-material Disks phantom. The materials contained in this phantom are arsenic (top row), selenium, bromine, krypton and rubidium (second-to-last row).

We see that the RU, UR, and cJoint methods are not capable of fully separating the mixtures and retrieving the disks on the inner circle. On the other hand, ADJUST nearly perfectly reconstructs the disks on the inner circle and the mixture disks on the outer circle.

C.10 Numerical studies: 3D phantom

We also apply the ADJUST algorithm to the 3D Shepp-Logan phantom to show the ability to reconstruct a 3D phantom. This 3D phantom is four times as large as the 2D Shepp-Logan phantom. The phantom is discretized on a grid of $128 \times 128 \times 128$ voxels. The phantom is shown in Figure A8. We consider 60 equidistant projection angles in the range of $[0, \pi]$ with a parallel-beam acquisition geometry. We show the visual results of the 3D material decomposition in Figure A9. The average MSE is 0.0029, the average PSNR is 26.67 and the average SSIM is 0.9763, indicating that the 3D reconstructions are almost accurate.

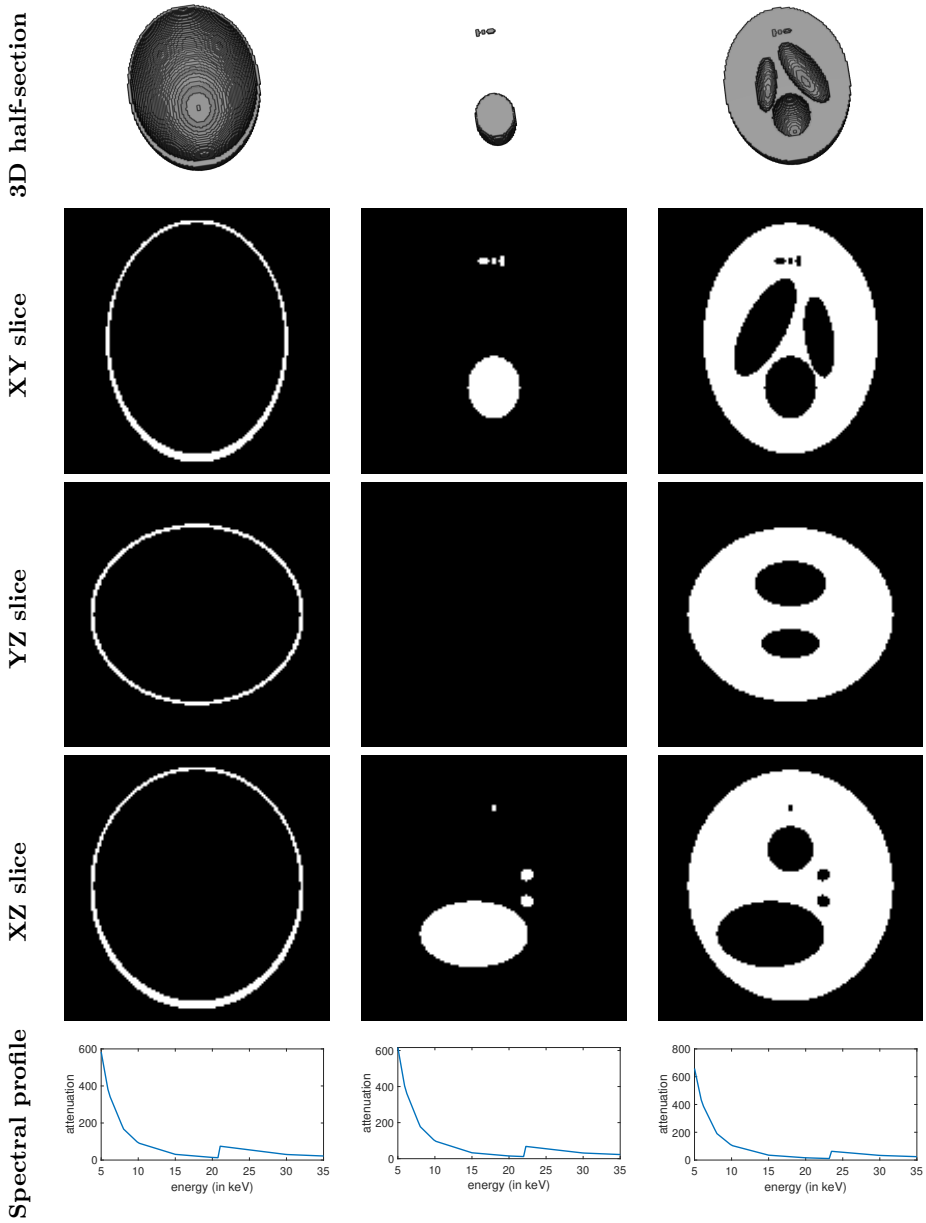


Figure A8: The true material compositions of the 3D Shepp-Logan phantom. The materials contained in this phantom are vanadium (left column), chromium (middle column), and manganese (right column).

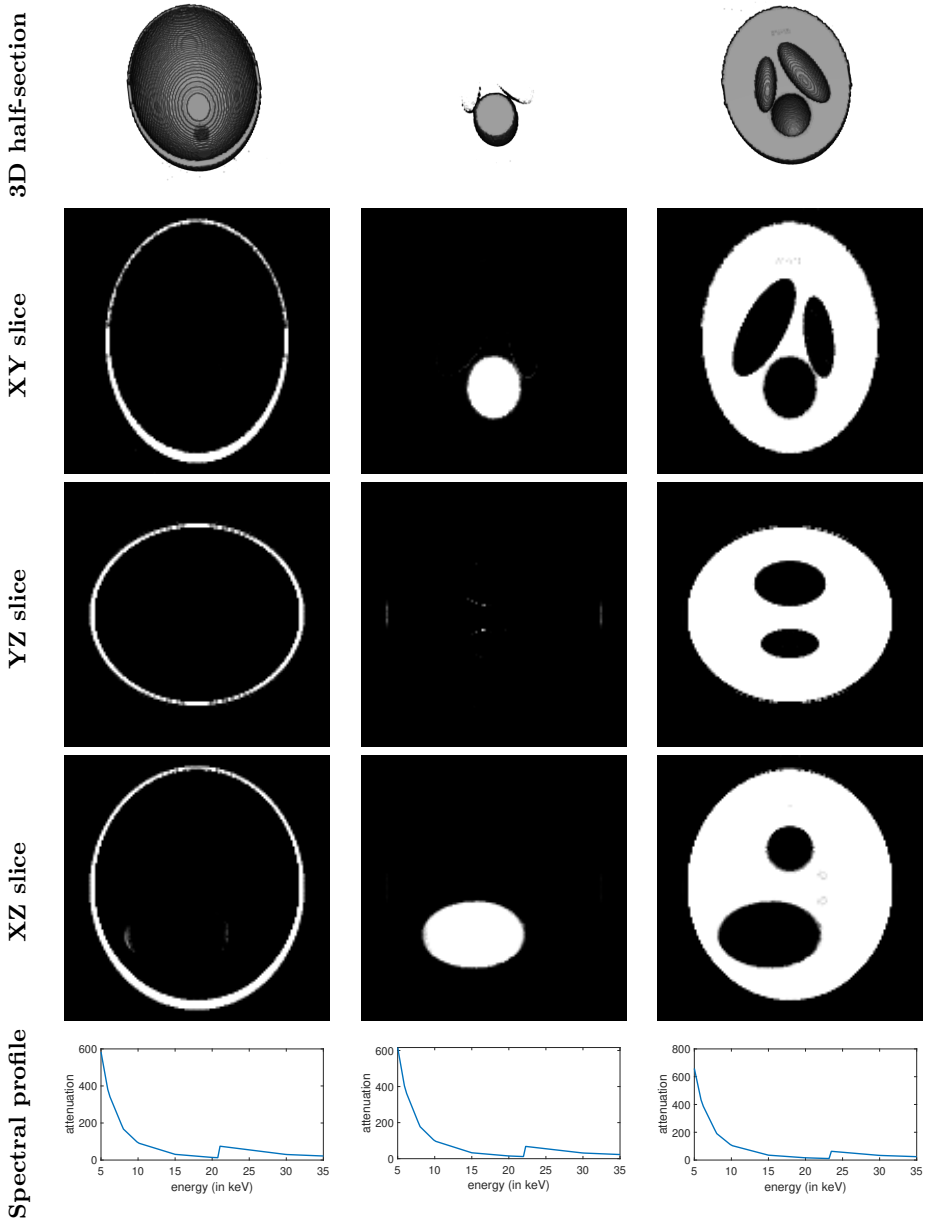


Figure A9: The reconstructed material composition of the 3D Shepp-Logan phantom from ADJUST algorithm. The materials contained in this phantom are vanadium (left column), chromium (middle column), and manganese (right column).

List of publications

Publications that are part of this dissertation:

1. M. T. Zeegers, F. Lucka, and K. J. Batenburg. "A Multi-Channel DART algorithm". In: *International Workshop on Combinatorial Image Analysis*. (Porto, Portugal). Ed. by R. P. Barneva, V. E. Brimkov, and J. M. R. S. Tavares. Springer, 2018, pp. 164–178.

2. M. T. Zeegers, D. M. Pelt, T. van Leeuwen, R. van Liere, and K. J. Batenburg. "Task-driven learned hyperspectral data reduction using end-to-end supervised deep learning". *Journal of Imaging* 6.12 (2020), p. 132.

3. M. T. Zeegers, T. van Leeuwen, D. M. Pelt, S. B. Coban, R. van Liere, and K. J. Batenburg. "A tomographic workflow to enable deep learning for X-ray based foreign object detection". *Expert Systems with Applications* 206 (2022), p. 117768.

4. M. T. Zeegers, A. Kadu, T. van Leeuwen, and K. J. Batenburg. "ADJUST: A Dictionary-based Joint reconstruction and Unmixing method for Spectral Tomography". *Inverse Problems* 38.12 (2022), p. 125002.

Publications that are not part of this dissertation:

1. E. G. Rens, M. T. Zeegers, I. Rabbers, A. Szabó, and R. M. H. Merks. "Autocrine inhibition of cell motility can drive epithelial branching morphogenesis in the absence of growth". *Philosophical Transactions of the Royal Society B* 375.1807 (2020), p. 20190386.

Summary in English

With globalization, growing world population, and other increases in scale taking place, humanity is becoming increasingly dependent on efficient industrial production. An essential component of industrial production is the execution of quality controls. These controls ensure that the product does what it is intended for and that no harm can occur when a product is used or consumed.

An important part of the industrial sector is food production. Various processes during food processing in and around factories can lead to contamination of the food product. This can include the presence of organisms (such as unwanted bacteria like E. coli in cheese) or the presence of chemicals (such as pesticides and cleaning agents on fruit) in or on the product. An important category is physical contamination, where small objects remain in the product. This can manifest as small pieces of glass, metal, or plastic that are left behind in food products. These small objects are referred to as *foreign objects* in this context and can be harmful when accidentally consumed. This is one of the reasons why supermarkets often have to organize large-scale recalls. It goes without saying that physical contamination leads to potential health risks, loss of time, financial damage, and loss of trust in the manufacturer. Because of the possibility of food contamination at various stages, quality control methods are constantly in development.

There are many methods for food contamination detection. Many methods are based on the interaction with *X-rays*, which can (partially) penetrate objects. For example, a foreign object can absorb more X-rays than the product it is hidden in. If the product is exposed to X-rays, an image can be formed by using an X-ray detector, on which the foreign object becomes visible. Although the developed methods are often applicable in a broader sense, their use for detection of foreign objects with X-rays is the common thread in this dissertation.

Although X-ray images reveal many properties, it can happen that projections of certain materials look similar to each other and are difficult to impossible to distinguish. X-ray *computed tomography* (CT) can offer a solution to this. In this method, multiple X-ray images of an object are taken at different angles. Afterwards, the object can be visualized in a three-dimensional volume using a reconstruction algorithm. Because CT generates a 3D volume instead of a 2D projection, the object can be analyzed more accurately and, for example, more thoroughly checked for the presence of a foreign object.

Even if X-ray imaging and computed tomography can provide a lot of insight into the composition of objects, there are many challenges to applying these methods for industrial purposes. In industry, it is important to make a decision based on the obtained information that is both *fast* an *accurate*. These goals are in conflict: X-ray images can be analyzed relatively quickly but do not always provide a definitive answer, while analysis of CT volumes is more time-consuming.

Additionally, in CT volumes, different materials may still appear as too similar and difficult to distinguish.

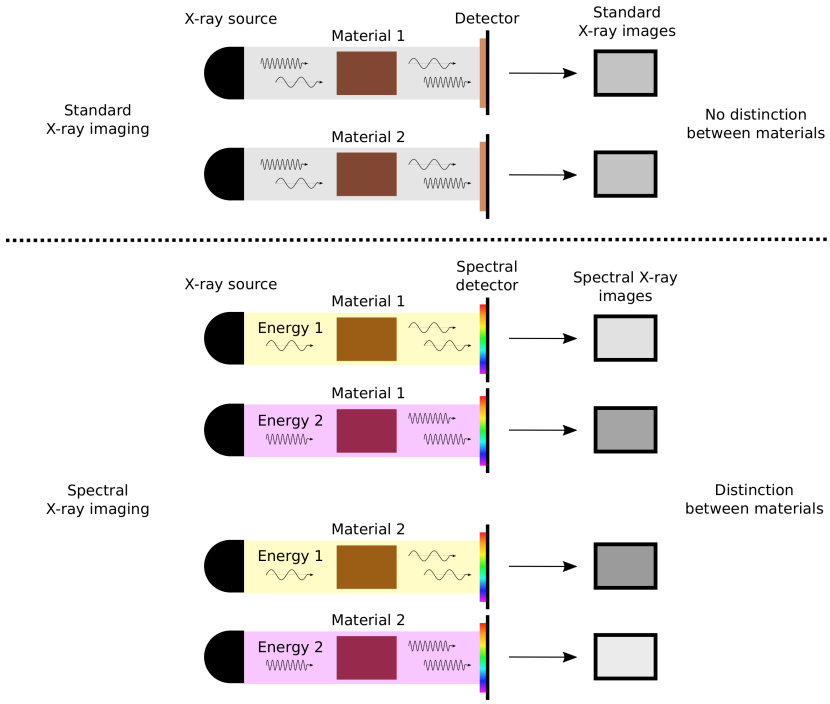
Two relatively new techniques offer possibilities to make object inspection both faster and more accurate. The field of artificial intelligence has recently experienced significant growth, in part due to the development of *deep learning* methods. In these approaches, a deep neural network (which is based on the networks of neurons in the human brain) can be trained with examples and their corresponding solutions to continuously improve its performance on a specific class of problems. A special category of neural networks is that of convolutional neural networks, which are capable of quickly detecting objects in X-ray images without requiring human intervention once they are trained.

Nevertheless, even the best X-ray images may not reveal a foreign object. To solve this, the technique of *spectral X-ray imaging* can be used. Spectral X-ray detectors can not only detect the photons that make up the X-rays, but can also determine the energy carried by each photon. The absorption of photons depending on energy is called an *absorption spectrum*. Because the absorption of photons depends not only on the material but also on the energy, this yields a unique absorption spectrum for each material. With spectral X-ray imaging, improved discrimination of materials present in an object is possible, where this is often difficult to measure with standard X-ray imaging (see Figure S1 for an example).

Similarly, for spectral X-ray imaging, the composition of an object can be better analyzed by using *spectral CT*. This can be done, for example, by reconstructing the measured CT images separately for each energy. By comparing multiple CT volumes - each corresponding to a certain energy - materials that are present in the object can be determined even better. However, as with standard CT, the reconstruction process takes extra time. Therefore, spectral CT reconstruction algorithms are constantly being developed to improve both accuracy and speed.

Machine learning and spectral X-ray imaging are two important ingredients for the methods developed in this thesis. **Chapter 1** explains how these techniques work and gives an overview of the most important existing methods. The chapters that follow cover various aspects of possible application of these techniques.

In **Chapter 2** we focus on a fundamental problem of deep learning: a large number of *training examples* is required to make a trained network function properly. However, for inspection with X-ray imaging, these examples are usually not easily available and require a lot of manual work to create. Therefore, we develop a *workflow* in which only a limited number of objects need to be CT scanned and reconstructed, from which a large number of training examples can be efficiently extracted. Through the use of a dataset collected in the FleX-ray laboratory at CWI, we demonstrate that this method can be used for foreign object detection and other deep learning based industrial applications.



Figuur S1: Schematic representation of the advantage of spectral X-ray imaging over standard X-ray imaging. With standard X-ray imaging, two different materials can result in the same intensity on the X-ray image. With spectral X-ray imaging, the energies in the X-ray beam can be separated from each other (by adjusting the source, or by measuring both energies in a spectral detector, or both). This gives multiple X-ray images per material. Based on the differences in these spectral X-ray images between materials, we can distinguish them.

The workflow is designed such that it can be expanded in many ways to obtain even better training data. An important extension is to use spectral X-ray imaging. In **Chapter 3**, we look at hyperspectral images, where many energies can be captured simultaneously. Because these hyperspectral images can contain a lot of data and can therefore be very large, it processing these with a deep learning trained network is relatively time-consuming. Therefore, it is necessary to *compress* these images. To preserve the important properties in these images for the tasks that need to be performed (such as detection of foreign objects), compression can be carried out in a *task-driven manner*. We achieve this by linking two neural networks and training these *simultaneously*: the first component is responsible for compression and the second part is responsible for performing the imaging task. We demonstrate with several examples that training of this composite network leads to stronger and more robust compression than conventional compression methods. This compression method not only has advantages in industrial tasks,

but also in other applications where efficient data transmission is required (such as sending data from a satellite to Earth).

In the following chapters are dedicated to improving CT reconstructions with spectral X-ray imaging. In **Chapter 4**, we condsider objects that contain *few different materials*. An existing algorithm specialized for these cases uses this property to achieve significantly better reconstructions than other methods that do not (fully) use this property. We develop an extension of this algorithm to multiple energy levels and demonstrate that with the use of spectral data objects consisting of more materials can be reconstructed accurately.

In **Chapter 5** we explore an alternative approach to spectral CT. We assume access to a spectral detector that can measure how much X-ray radiation is absorbed per energy by each material. If these spectra are known for a range of materials that may occur in the object (resulting in a *dictionary*), it can be used to steer a spectral CT algorithm to certain solutions. We develop such an algorithm and in a series of experiments, we not only demonstrate its robustness but also show that it can yield better results than current leading spectral CT algorithms.

Taken together, this dissertation proposes methods for improving industrial processes, with a specific motivation from food inspection. The methods use deep learning and spectral X-ray imaging. The proposed workflow for detecting foreign objects can be applied in factories, and can be expanded with the other methods proposed and analyzed in this thesis. Nevertheless, we expect that the individual methods are also applicable in a broader sense for medical purposes (such as CT scans in hospitals), security purposes (such as scanning luggage at airports), and a wide spectrum of other areas.

Summary in Dutch

Met de globalisering, groeiende wereldbevolking en andere schaalvergrotingen die nu plaatsvinden, wordt de mensheid steeds afhankelijker van efficiëntere industriële productie. Een essentieel component van industriële productie is het uitvoeren van kwaliteitscontroles. Deze controles zorgen ervoor dat het product doet waar het voor gemaakt is en er geen schade kan ontstaan wanneer een product gebruikt of geconsumeerd wordt.

Een belangrijk onderdeel van de industriële branche is de voedselproductie. Verschillende processen tijdens voedselverwerking in en rond fabrieken kunnen leiden tot contaminatie van het voedselproduct. Denk hierbij bijvoorbeeld aan de aanwezigheid van organismen (bijvoorbeeld ongewenste bacteriën zoals E. coli in kaas) of de aanwezigheid van chemicaliën (zoals pesticiden en schoonmaakmiddelen op fruit) in of op het product. Een belangrijke categorie is fysieke contaminatie, waarbij kleine objecten in het product achterblijven. Dit kan zich uiten in kleine stukjes glas, metaal of plastic die achter zijn gebleven in voedselproducten. Deze kleine objecten worden in deze context (lichaams)vreemde objecten (*foreign objects*) genoemd, en kunnen schadelijk zijn wanneer deze per abuis worden geconsumeerd. Mede hierdoor moeten supermarkten regelmatig grootschalige terugroepacties op touw zetten. Het is vanzelfsprekend dat fysieke contaminatie leidt tot mogelijke gezondheidsrisico's, tijdverlies, financiële schade en verlies van vertrouwen in de fabrikant. Vanwege de mogelijkheid van voedselcontaminatie op allerlei niveaus zijn de methodes voor kwaliteitscontroles voortdurend in ontwikkeling.

Er bestaan vele methodes voor het detecteren van voedselcontaminatie. Veel methodes zijn gebaseerd op de interactie met röntgenstraling (*X-rays*), die voorwerpen (gedeeltelijk) kan doordringen. Een foreign object kan bijvoorbeeld meer röntgenstraling absorberen dan het product waar het in verscholen is. Als het product dus met behulp van een bron wordt blootgesteld aan röntgenstraling kan er een beeld gevormd worden door middel van een röntgenstralingsdetector, waarop het foreign object zichtbaar wordt. Hoewel de ontwikkelde methodes vaak in bredere zin inzetbaar zijn, vormt het detecteren van foreign objects met röntgenstraling de rode draad in dit proefschrift.

Hoewel röntgenbeelden veel kunnen laten zien, kan het voorkomen dat projecties van bepaalde materialen op elkaar lijken en moeilijk tot onmogelijk te onderscheiden zijn. Als oplossing hiervoor kan X-ray *computertomografie* (CT) gebruikt worden. Hierbij worden meerdere röntgenbeelden van een object opgenomen onder verschillende hoeken. Hierna kan het object in kwestie met een reconstructiealgoritme gevisualiseerd worden in een driedimensionaal volume. Omdat CT een 3D volume genereert in plaats van een 2D projectie, kan het voorwerp beter geanalyseerd worden en bijvoorbeeld beter gecontroleerd worden op de aanwezigheid van een foreign object.

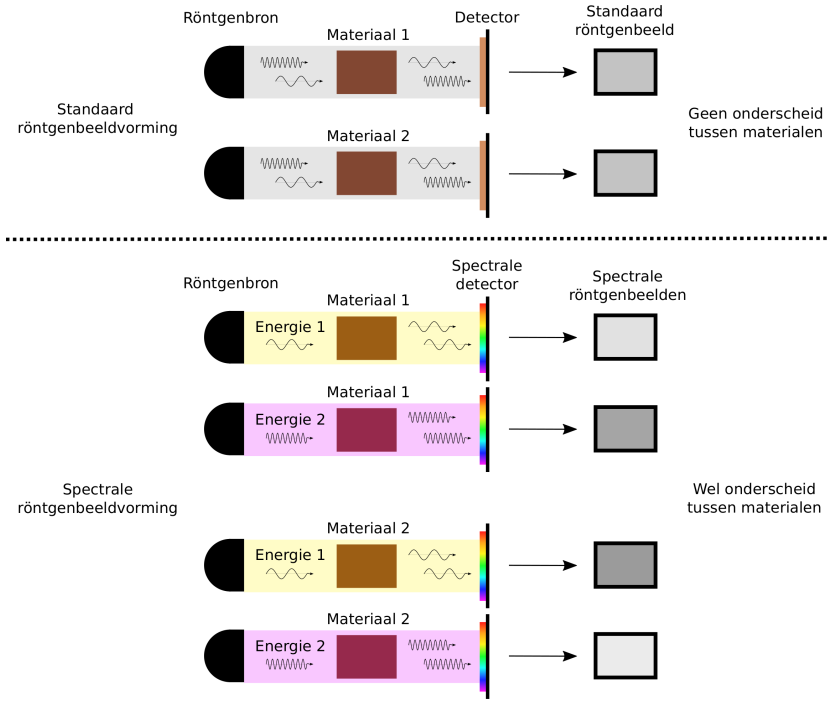
X-ray beeldvorming en computertomografie kunnen veel inzicht geven in de samenstelling van een object, maar er zijn veel uitdagingen om dit toe te passen voor industriële doeleinden. Het is belangrijk om op basis van de verkregen informatie een zowel *snelle* als een *accurate* beslissing te maken. Dit zijn conflicterende doelen: röntgenbeelden kunnen relatief snel geanalyseerd worden maar bieden niet altijd uitsluitsel, terwijl CT volumes beter te analyseren zijn maar het berekenen en analyseren ervan meer tijdrovend is. Bovendien kan het in CT volumes toch zijn dat verschillende materialen te veel op elkaar lijken en moeilijk uit elkaar te halen zijn.

Twee relatief nieuwe technieken bieden mogelijkheden om inspectie van objecten zowel sneller als accurater te maken. Het onderzoeksgebied van kunstmatige intelligentie heeft recentelijk een grote groei doorgemaakt, mede door de ontwikkeling van *deep learning* methodes. Hierbij kan een diep neuraal netwerk (die gebaseerd is op de netwerken van neuronen in de hersenen van de mens) door middel van voorbeelden met bijbehorende oplossingen getraind worden om zijn werking op een bepaalde klasse van problemen steeds te verbeteren. Een speciale categorie van neural netwerken zijn de *convolutionele neurale netwerken*. Deze zijn in staat om, eenmaal getraind, zeer snel objecten te detecteren op röntgenbeelden zonder dat daar menselijke handelingen aan te pas komen.

Desondanks kan op de beste röntgenbeelden een foreign object onzichtbaar blijven. Om dit te verbeteren is de techniek van spectrale röntgenbeeldvorming (*spectrale X-ray imaging*) bruikbaar. Spectrale röntgenstralingsdetectoren kunnen de fotonen, waaruit röntgenstralen zijn opgebouwd, niet alleen detecteren maar voor elke foton ook de energie bepalen die ze bij zich dragen. De absorptie van de fotonen afhankelijk van energie wordt een *absorptiespectrum* genoemd. Omdat de absorptie van fotonen niet alleen afhankelijk is van het materiaal maar ook van de energie, geeft dit voor ieder materiaal een uniek absorptiespectrum. Met spectrale röntgenbeeldvorming kan beter bepaald worden welke materialen er in een object zitten, waar dit bij standaard röntgenbeeldvorming vaak niet goed te meten is (zie Figuur S1 voor een voorbeeld).

Ook voor spectrale röntgenbeeldvorming geldt dat de samenstelling van een object beter geanalyseerd kan worden wanneer wordt overgegaan op *spectrale CT*. Dit kan bijvoorbeeld door de gemeten CT beelden voor elke energie apart te reconstrueren. Door meerdere CT volumes - elk behorende bij een zekere energie - met elkaar te vergelijken, kan nog beter bepaald worden welke materialen er in het object zitten. Desondanks geldt ook hier weer dat het maken van de reconstructies extra tijd kost. Daarom zijn spectrale CT reconstructiealgoritmes voortdurend in ontwikkeling om zowel de nauwkeurigheid als de snelheid te verbeteren.

Machine learning en spectrale röntgenbeeldvorming zijn twee belangrijke ingrediënten voor de methodes die in dit proefschrift ontwikkeld zijn. In **hoofdstuk 1** wordt uitgelegd hoe deze technieken werken en wordt een overzicht van de be-



Figuur S1: Schematische weergave van het voordeel van spectrale röntgenbeeldvorming ten opzichte van standaard röntgenbeeldvorming. Bij standaard röntgenbeeldvorming kunnen twee verschillende materialen resulteren in dezelfde intensiteit op het röntgenbeeld. Bij spectrale röntgenbeeldvorming kunnen de energieën in de röntgenstraal van elkaar gescheiden worden (door de bron anders in te stellen, of door beide energieën te meten in een spectrale detector, of beide). Hierdoor krijgen we per materiaal meerdere beelden. Op basis van de verschillen in deze spectrale röntgenbeelden tussen materialen kunnen we ze onderscheiden.

langrijkste bestaande methodes gegeven. De hoofdstukken daarna behandelen verschillende aspecten van de mogelijke toepassing van deze technieken.

In **hoofdstuk 2** richten we ons op een fundamenteel probleem van deep learning: er zijn veel *trainingsvoorbeelden* nodig om een getraind netwerk goed te laten functioneren. Ook voor inspectie met röntgenbeeldvorming zijn deze voorbeelden lang niet altijd gemakkelijk verkrijgbaar en vereisen veel manueel werk om te vervaardigen. Daarom ontwikkelen we een *werkstroom* waarbij enkel een gelimiteerd aantal objecten met CT gescand en gereconstrueerd hoeft te worden, waaruit op een efficiënte manier veel trainingsvoorbeelden verkregen kunnen worden. We laten aan de hand van een dataset opgenomen in het FleX-ray laboratorium van het CWI zien dat deze methode gebruikt kan worden voor foreign object detectie en andere op deep learning gebaseerde industriële toepassingen.

De werkstroom is zodanig vormgegeven dat het op vele manieren uitbreidbaar

is om nog betere trainingsdata te verkrijgen. Een belangrijke uitbreiding is om spectrale röntgenbeeldvorming te gebruiken. In **hoofdstuk 3** kijken we naar hyperspectrale beelden, waarbij zeer veel energieën tegelijkertijd kunnen worden opgenomen. Omdat deze hyperspectrale beelden zeer veel data kunnen bevatten en daardoor dus zeer groot kunnen zijn, kost het relatief veel tijd om deze te verwerken door een met deep learning getraind netwerk. Daarom is het nodig om deze beelden te *comprimeren*. Om de belangrijke eigenschappen in deze beelden te behouden voor de taken die uitgevoerd moeten worden (zoals herkenning van foreign objecten) kan de compressie op een *taakgedreven manier*. Dit doen we door twee neurale netwerken aan elkaar te koppelen en deze *tegelijk* te laten leren: het eerste deel is verantwoordelijk voor de compressie en het tweede deel is verantwoordelijk voor het uitvoeren van de taak. We laten met meerdere voorbeelden zien dat het gecombineerd trainen van dit samengestelde netwerk leidt tot sterkere en robuustere compressie dan met conventionele compressiemethodes. Deze compressiemethode heeft niet alleen voordelen bij industriële taken, maar ook in andere applicaties waarbij efficiënt data moet worden uitgewisseld (bijvoorbeeld bij het sturen van data van een satelliet naar de aarde).

In de opvolgende hoofdstukken leggen we ons toe op het verbeteren van de CT reconstructies met spectrale röntgenbeeldvorming. In **hoofdstuk 4** kijken we naar objecten die *weinig verschillende materialen* bevatten. Een bestaand algoritme dat op deze gevallen is gespecialiseerd gebruikt deze eigenschap om significant betere reconstructies te bewerkstellingen dan andere methodes die deze eigenschap niet (volledig) gebruiken. We ontwikkelen een uitbreiding van dit algoritme naar meerdere energieniveaus en tonen aan dat met het gebruik van spectrale data objecten die bestaan uit meer materialen accuraat kunnen worden gereconstrueerd.

In **hoofdstuk 5** bekijken we een alternatieve benadering van spectrale CT. We nemen aan dat er toegang is tot een spectrale detector die kan meten hoeveel röntgenstraling per energie door elk materiaal wordt opgenomen. Als deze spectra bekend zijn voor een scala aan materialen die mogelijk in het object voorkomen (resulterend in een 'referentietabel'), kan dit als hulpmiddel worden gebruikt in een spectraal CT algoritme. We ontwikkelen een dergelijk algoritme en in een reeks experimenten tonen we niet alleen aan dat het robuust is, maar ook dat het betere resultaten kan opleveren dan huidige toonaangevende spectrale CT algoritmes.

Welbeschouwd worden er in dit proefschrift methodes voorgesteld voor het verbeteren van industriële processen, met specifieke aandacht voor de toepassing van voedselinspectie. De methodes maken gebruik van deep learning en spectrale röntgenbeeldvorming. De voorgestelde werkmethode om foreign objects te detecteren is toepasbaar in fabrieken, en is uitbreidbaar met de andere methodes die in dit proefschrift worden voorgesteld en geanalyseerd. Desalniettemin verwachten we dat de individuele methodes ook in bredere zin toepasbaar zijn voor medische doeleinden (zoals bij CT-scans in ziekenhuizen), veiligheidsdoeleinden (zoals het scannen van bagage op vliegvelden) en een wijd spectrum aan andere gebieden.

Curriculum vitae

Mathé Thomas Zeegers was born in Alkmaar, the Netherlands, on the 3rd of May in 1992. He completed his secondary education (gymnasium) in 2010 at the Trinitas College locatie Han Fortmann in Heerhugowaard. In 2013, he obtained his bachelor degrees in Mathematics and Computer Science from Leiden University. After this, he obtained his master's degree (cum laude) in Computer Science in 2016 from Leiden University. The corresponding master thesis titled "Theoretical properties of 2048" was written under supervision of Dr. W.A. Kosters. After this, Mathé did an internship in the Life Science group at Centrum Wiskunde & Informatica (CWI), the Dutch national research institute for computer science and mathematics in Amsterdam. The result is the thesis titled "Cellular Potts modelling of branching morphogenesis by chemical inhibition and mechanochemical feedback", written under supervision of Dr. E. G. Rens and Prof. dr. R. M. H. Merks, after which Mathé obtained the master's degree for Mathematics from Leiden University in 2017. After this, he started his PhD research at the Computational Imaging group at CWI under the supervision of Prof. dr. K. J. Batenburg. During this time, Mathé conducted research in the field of spectral X-ray tomography, with links to discrete tomography, hyperspectral imaging and machine learning. During the PhD research, Mathé attend and gave talks at various venues, such as the International Workshop on Combinatorial Image Analysis (IWCIA) in Porto in 2018 and the Applied Inverse Problems (AIP) in Grenoble in 2019. He obtained a SIAM Student Travel Award in early 2020. In addition, from 2018 to early 2021, Mathé was a member of the PhD activity committee that organizes social events for PhD students, postdocs, trainees and other personnel at CWI.

Acknowledgements

First of all, I would like to thank my promotor Joost Batenburg for not only giving me the opportunity to pursue this PhD, but also giving me a lot of freedom in the design of the resulting work. Joost, every meeting with you was something to look forward to and catalyzed the writing process so much. I am also thankful for the tremendous help from my co-promotors. I cherish to have shared the office with Daniël Pelt, teaching me so many tips and tricks over time. His critical and precise comments are reflected all over the first three chapters. Also, I thank Tristan van Leeuwen for grasping my entire projects so quickly and providing indispensable guidance and suggestions. It goes without saying that the collaboration was exceedingly smooth and pleasant.

I would also like to thank my other co-authors. Ajinkya Kadu, it is safe to say that our collaboration is one of the most pleasant I ever had. You were always available and I learned so much from you (especially about optimization). Sophia Coban, thank you for all the scanning aid (scanning was always so much fun with you around) and all the tea. I consider the trip to (Man)chester to be the highlight of the PhD. Felix Lucka, thank you for kick-starting the PhD with your help in the Multi-Channel DART project in Chapter 4. Robert van Liere, thank you for your helpful comments on Chapters 2 and 3.

I would like to thank all other Computational Imaging group members throughout the years for the great athmosphere. Giulia Colacicco* (paranymph), thanks for all your invaluable help during the first year of the PhD and for keeping in touch afterwards. Nicola Viganò, it was always worth staying at the lunch table a little longer with you and joining the enjoyable and instructive post-lunch discussions. Džemila Šero and Georgios Pilikos, thank you for the ad hoc brainstorming and suggestions when I needed them the most. Vladyslav 'Vlad' Andriiashen*, your visits to my office were always so nice thanks to your vast knowledge, admirably positive attitude - even during difficult circumstances - and your unique sense of humour.

Other members I would like to thank for their support are Alexander Skorikov, Dirk Schut, Poulami Ganguly, Hamid Fathi, Tianyuan Wang, Roozbeh Valadian, Holger Kohr, Rien Lagerwerf, Maximillian 'Max' Kiss, Maureen van Eijnatten, Jan-Willem Buurlage, Jordi Minnema, Richard Schoonhoven, Zhichao Zhong, Henri der Sarkissian, Alexander Kostenko, Willem Jan Palenstijn, Allard Hendriksen, Rob Bisseling, Adriaan Graas (fond AIP2019 memories), Francien Bossema, Ryan Schoop and all guest researchers.

From the Nikhef collaboration I would like to thank Martin Fransen and Navrit Bal. Navrit, while unfortunately none of our many, many experiments with Medipix directly made it into this dissertation, working with you was really enjoyable and I learned a lot in the process.

CWI has been a really nice place to carry out research, thanks to all the great people that work(ed) there. In particular, I would like to thank Nada Mitrovic, Minnie Middelberg, Bikkie Aldeias, Vera Sarkol and Remco Westra for their support and their endless kindness.

I would also like to thank the members of the PhD activity committee (including those from before and after I joined). Together with Jan Czajkowski, Esteban Landerreche, Carl Shneider, Muriel Pérez, Sanne van den Berg, Léon Ouwerkerk and other members marked by a *, it was so pleasant to organize activities, get to know new people and bring everyone together!

In addition, there were many people I could share my research with and get any sort of feedback. For this, I thank Dorien Zwaneveld, Marco Virgolin, Karen Kokabisaghi, Arjan Cornelissen*, Tania Bogatsch, Christopher Esterhuyse, Nikolaj Mücke, Prashant Kumar, Thomas van Binsbergen, Yllka Velaj and Srinivasan Arunachalam. Also, thanks to Lisanne Rens, Jasmijn Baaijens, Leonie van Steijn and Marleen Balvert for particularly helpful advices.

On top of this, I would like to thank Esmée Vendel for sharing the passion for music and meeting up regularly and Hemaditya Malla* for always being available as a listening ear and to hang out with. Dean de Leo, thank you for organising CWI football and your regular cozy dinners to offer a welcome change of routine. Similarly, besides chairing the PhD activity committee in such a nice manner, I thank Mark Abspoel* for his listening ear and the many fun after-work board games.

The activity room in CWI proved to be fertile ground for making valuable friends. I thank Isabella Pozzi* and Ruben Brokkelkamp* for all the notoriously hard foosball matches, but also many other great adventures beyond that. Similarly, I thank Sven Polak for all the table tennis training. And lastly, I thank Sander Gribling for advices and Skype sessions (especially in the lockdowns), and Pieter Kleer (paranymph) for occasional proofreading and house providing, and both for being respected ping pong and foosball opponents.

Lastly, Rakesh Sarma was of great help with PyTorch related issues, which were discussed among many other things during our exciting adventures: badminton, massive bike rides and watching football after work in the office. I treasure these memories.

Finally, I would like to thank Walter Kosters and Roeland Merks for their respective contributions to leading me to this PhD.

From outside CWI, I would like to thank Liam Curtin for his interest and support, and Daniëlle van der Heide for test reading and getting my mind off of work. Also, thanks to Rajesh Mahabir, Jordi Ozir, Michiel Vos and the remaining 'LIACS crew' for keeping in touch after our studies, sharing our (working) experiences and meeting up every once in a while. Similarly, I thank Dave Bleeker for his support as well. I value our long-standing friendship, and that we still catch up to share experiences and have a lot of fun!

Lastly, I would like to express my gratitude to my family. Sascha (+Gilles), Guido and Esben, thank you for your important support. Benthen, thanks for all the mathematics we share, as well as the support and fun moments throughout life. And my parents Marianne and Siem, thank you for providing the stones used in Chapter 2 and - infinitely more importantly - for your endless and unconditional love and support.

No chickens were harmed in the making of this dissertation and its associated papers.