Faster X-ray computed tomography in real-world dynamic applications

Graas, A.B.M.

Chapter 5

Scintillator Decorrelation for Self-supervised X-ray Radiograph Denoising

This chapter is based on the article: Adriaan Graas and Felix Lucka. "Scintillator decorrelation for self-supervised X-ray radiograph denoising". In: *Measurement Science and Technology* 36.6 (2025), p. 065415. DOI: 10.1088/1361-6501/addc06

5.1 Introduction

Imaging with X-rays is widely used for direct radiography and tomographic image reconstruction (CT, Computed Tomography). Thanks to its rapid acquisition and high spatial resolution, it is found in many applications of health care, industry (e.g., foreign object detection or non-destructive testing), and scientific research (e.g., experimental physics or biomedical sciences). An important topic within X-ray imaging is to mitigate noise from the radiographs, i.e., the raw radiation intensity measurements collected by X-ray detectors [138]. Noise does not only degrade the image quality, but also leads to streaks and other artifacts in the reconstructed images. The primary source of noise is *photon noise*, which is Poisson-distributed and caused by natural random variations in X-ray generation and attenuation [3, 4, 13, 15].

Current state-of-the-art methods for image denoising rely on deep learning, i.e., on training deep neural networks to map noisy input images to noise-free outputs [139, 140]. The most common training strategy is *supervised learning* [141], for which each noisy image from the training set is paired with a corresponding low-noise or noise-free *target* – the desired network output. For the situations that supervised targets are not available, *self-supervised learning* replaces them by noisy surrogates. For instance, in the strategy suggested by Noise2Noise [39], each surrogate target must be a second, statistically-independent, realization of the image. Since, in this case, the input provides no information that can help predicting the target's noise, an optimal network is only able to predict any noise-free image features that the input and target have in common. The requirement of two noisy images, however, is limiting for many imaging applications. This is especially the case for X-ray radiography and Computed Tomography, where auxiliary paired noisy images cannot be acquired with fast dynamics or low-dose scans.

In the recent years, research has focused on *unpaired* self-supervised approaches, i.e., strategies that do not need a secondary noisy realization. One particular promising category is that of *blind-spot networks* (BSNs). BSNs do not rely on a new image as surrogate target, but instead reuse a subset of pixels from the noisy image [41, 42]. These "blind spot" pixels, denoted by their indices J, need to be masked out in the input, denoted with u, for example via replacement by zeros or random values. Similar to Noise2Noise, the input provides no information for predicting the target's noise when u_{J^c} and u_J are statistically independent, which is satisfied when noise is pixelwise independent. On the other hand, spatially-extended image features in u_{J^c} provide information for predicting the signal of u_J, which a neural network can learn thanks to repeated occurrences of image features in the data. During inference, the full noise-free images can be recovered via execution of the BSN for all sets of blind spots.

BSNs are promising methods for denoising radiographs: They do not require modified acquisition, and are able to take advantage of large sets of noisy experimental data. Their main challenge for X-ray applications, however, is that detector instruments introduce local spatial correlations via blur. As a consequence, blind-spot networks propagate noise from u_{J^c} to u_J, rendering them ineffective. Mitigating blur and improving spatial resolution has been

an important research focus to reduce the effects of, e.g., final focal spot size, scattering, and system vibrations [142, 143, 144]. However, only blurs that affect the noise, i.e., when photon fluctuations correlate between nearby detector pixels, hamper the blind-spot denoiser.

In this article, we specifically target the blur caused within scintillator detectors. We show that, when correlations are uniform over the radiograph, they can be approximately reverted using a direct deconvolution in the frequency domain, i.e., without requiring a Wiener filter or an iterative algorithm. Our approach uses a deconvolution kernel that is obtained via an empirically-estimated scintillator point-response function (PRF), which describes the photon avalanche on conversion of X-rays into visible light. We show that it is robust against modest levels of additive Gaussian noise, and that the deconvolved radiographs restore the denoising potential of BSNs. In *Results I*, we verify this workflow using numerical simulations, an experimental phantom measured with Teledyne DALSA Xineos-3131 Caesium-Iodine scintillator detectors, and training of the blind-spot denoising method called Noise2Self [41]. In *Results II*, the workflow is applied to a large real-world experimental data set that is acquired for sparse-view dynamic X-ray tomography of fluidized beds. We provide an introduction in tomography in section 5.2.1, and the reader is referred to the textbooks [4, 145] for a treatment on this topic. To motivate denoising of radiographs as a preprocessing technique before reconstruction, section 5.5.1 compares Noise2Self with the Noise2Inverse denoiser in reconstruction space [40].

The paper is organised as follows: In *Background* (section 5.2) we review topics on X-ray imaging, computed tomography, deep learning for denoising, and blind-spot denoising with correlated noise. In *Method* (section 5.3) we discuss the noise model, the direct deconvolution, and neural network training. In *Results I* (section 5.4) we study the method on numerical and experimental data. In *Results II* (section 5.5) we investigate sparse-view tomographic image reconstruction on synthetic data and show results for experimental data from an ultra-sparse-angle X-ray set-up for imaging gas-solids fluidized beds.

5.2 Background

5.2.1 Tomographic reconstruction

The goal of tomographic reconstruction is to recover a 3D image $x \in \mathbb{R}^{N_x N_y N_z}$, corresponding to the discretization of $\mu(\eta)$ on a spatially uniform grid, from a set of projection images $[y_1, \ldots, y_{N_\psi}]$ acquired at N_ψ different positions of the X-ray source and detector, e.g., under different angles of an orbital or helical trajectory around the object. The discretization of all linear equations equation (1.4) is described by the operator $A\colon x \mapsto [y_1, \ldots, y_{N_\psi}]$ called the *forward projector* (cf. chapter 9 in [4]).

Tomographic reconstruction amounts to solving the ill-conditioned and often under-determined linear system $[y_1, \ldots, y_{N_\psi}] = Ax$ in an approximate way. A wide range of methods has been developed towards this purpose, see, e.g., the textbooks [3, 4], for an in-depth treatment of two reconstruction methods that will be used in our results. The first, the *filtered-backprojection* (FBP), is a two-step reconstruction technique that first applies a

suitable filter g to the projections and subsequently executes a *backprojection* operator A^{T}, i.e., a computational procedure approximating the transpose of A:

$$x^{\mathrm{FBP}} := A^{\mathsf{T}}[g \circledast y_1, \ldots, g \circledast y_{N_\psi}], \tag{5.1}$$

where $\circledast$ denotes convolution. FBP is similar to the Feldkamp-Davis-Kress (FDK) algorithm, which is tailored to a cone beam geometry instead of a parallel beam geometry. For our experiments we selected the Ram-Lak filter, cf. chapter 6 in [4]. The second, the *simultaneous iterative reconstruction technique* (SIRT), is an example of an algebraic iterative reconstruction technique (cf. chapter 11 in [4]). In this case, x^{SIRT} is an approximate solution to the constrained weighted least-squares problem

$$\underset{x \in \mathcal{C}}{\arg\min} \ \left\| Ax - [y_1, \ldots, y_{N_\psi}] \right\|_{M_1}^2, \tag{5.2}$$

found by means of a gradient-decent-type iteration

$$x^{(k+1)} = x^{(k)} + D_1 A^{\mathsf{T}} M_1 \left([y_1, \ldots, y_{N_\psi}] - A x^{(k)} \right) \tag{5.3}$$

on equation equation (5.2) for a fixed number of iterations K, i.e., $x^{\mathrm{SIRT}} := x^{(K)}$. D_1 and M_1 are diagonal matrices that contain the row and column sums of A, and $\mathcal{C}$ is the constrained solution space.

In section 5.5, FBP and SIRT algorithms are implemented using the ASTRA Toolbox software package [80]. On a modern graphics processing unit (GPU), FBP takes about a second of runtime, whereas SIRT can take several minutes, depending on the data dimensions.

5.2.2 Supervised and self-supervised denoising

Equation (1.3) in Chapter 1 is a simplified model of an ideal X-ray radiograph. It does not account for noise and does not describe all the physical phenomena that contribute to the factual measurement data (called *model mismatch*). The noise component mainly consists of *photon noise*, which stems from the quantum nature of light. When modeled with a Poisson probability distribution, it includes X-ray photon generation in the X-ray source, interactions with atoms in the sample, and detection in the scintillator (section 5.3.1) [3, 13, 14, 15]. Another large component of noise and model mismatch is due to *scattering*. X-ray photons that scatter in a direction away from the detector are indistinguishable from attenuated photons, and therefore accounted for by Beer-Lambert's law. However, photons that arrive at the detector via different paths than l_i in (1.3) of Chapter 1 (e.g., multiple scattering, multi-source set-ups), contribute to a model mismatch. Next to photon noise, a typically smaller noise component is due to electronics within the detector instrument. Their noise characteristics depend, e.g., on the the type of photodetectors used. Examples are charge-diffusion in CMOS (complementary metal-oxide-semiconductor) pixels, defective pixels, dark currents, amplifier noise, and digitization [3, 146].

Mitigating the effects of noise in X-ray measurements and CT reconstructions has been an important research focus for already many years [138]. Before the advent of machine learning algorithms, classical denoising methods such as BM3D, non-local means (NLM), or total variation (TV) had provided state-of-the-art results for images with photon noise or mixed Poisson-Gaussian noise. Most current research focuses on deep learned denoisers, and in particular convolutional neural networks (CNNs) [139, 140, 147]. We will describe a deep learned denoiser by a function f_θ that maps noisy images into clean images. Here θ denotes the set of learnable network parameters. During *training*, θ is optimized based on examples in a data set. The hope is that by leveraging statistical properties of the training data, the trained f_θ also performs well on unseen noisy images (*generalization*). In *supervised* training, f_θ is trained on pairs of noisy and clean images $(u, \hat{u}) \sim \mathcal{D}$ from a data distribution $\mathcal{D}$. The objective is then to solve

$$\arg\min_\theta \; \mathbb{E}_{(u,\hat{u})\sim\mathcal{D}} \, \|f_\theta(u) - \hat{u}\|_2^2 \,, \tag{5.4}$$

here given using the mean-squared error (MSE), the most common loss function. We write the loss as the minimization of an expected value, taken over image pairs $(u, \hat{u}) \sim \mathcal{D}$. Practically, however, data sets of example images consist only of a finite number of image pairs, and the expectation is replaced by the empirical mean (i.e., the average over a limited number of samples from $\mathcal{D}$). For brevity, we denote $\mathbb{E} \equiv \mathbb{E}_{(u,\hat{u})\sim\mathcal{D}}$ in the forthcoming sections, while in experiments we will evaluate it with the empirical mean.

For many applications, obtaining the clean images $\hat{u}$ needed for supervised learning is impractical or impossible. *Self-supervised* learning therefore tries to modify the objective in equation (5.4) in such a way that $\hat{u}$ is not needed anymore [14]. The Noise2Noise principle [39], discussed in the introduction, does this by replacing $\hat{u}$ with a second noisy image, assuming the images are statistically independent. While its principle was used in, e.g., electron microscopy [148, 149], paired acquisition often poses challenges for real-world applications, and several methods have therefore been suggested to construct pairs from single noisy images. One example is Neighbor2Neighbor, which constructs image pairs via subsampling [150]. Other alternatives are *zero-shot methods*, such as the Deep Image Prior [147] or Zero-Shot Noise2Noise [151]. Zero-shot learning, however, is typically slower and less accurate than CNNs trained on a data set.

5.2.3 Related work

When noise is correlated, $\mathcal{J}$-invariance can in principle still be used by letting the $J \in \mathcal{J}$ in definition 1.8.1 consist of masking regions larger than pixelwise grids. MM-BSN (Multi-Mask BSN), for example, applies differently-shaped masks to suppress noise with non-uniform correlation structures [152]. Indeed, when J covers larger correlated regions, the property $f_\theta(u) - \hat{u} \perp u - \hat{u}$ may hold again, therefore removing the cross-term in equation (1.14). However, schemes that require extended masking generally cause blurring of the estimate, since information about sharp image features is generally contained in a close neighborhood of the target pixels.

Other approaches in the literature mitigate correlated noise by using pixelwise subsam-

pling in conjunction with BSNs. Example architectures are AP-BSN [153] and SS-BSN [154]. In AP-BSN, the input is subsampled to s^2 small images using a pixel-shuffle downsampling operation. With $s \in \mathbb{N}^+$ denoting the noise correlation radius, the noise in each subsampled low-resolution image becomes uncorrelated. After denoising the subimages with a BSN, the results are upscaled to a high-resolution output using a learned refinement to overcome aliasing. It has been observed that AP-BSN can be difficult to apply to scientific images [149]. Noise2SR [155] and M-Denoiser [149] extend the subsampling method by randomization and shattering, rather than pixel-shuffling.

5.3 Method

Instead of masking large regions, or employing subsampling, our method obtains $\mathcal{J}$-invariance by reverting the process that causes the spatial correlation. Section 5.3.1 first discusses the origin of correlation in X-ray radiographs obtained from scintillator detectors. Section 5.3.2 then motivates the use of a direct deconvolution, and section 5.3.3 and section 5.3.4 outline a correlation kernel estimation and neural network training procedure.

5.3.1 Scintillator noise model

Scintillator detectors have a widespread use, e.g., in medical devices and scientific laboratories including synchrotron light sources [142]. In these detectors, a crystal converts incident X-rays into visible light from the optical spectrum, which can subsequently be detected with an array of photodetectors such as CMOS active pixels [4, 156]. These pixels convert the light to electric charge, which is stored until it is read out by the detector electronics. For the X-ray wavelengths produced by cone beam sources, common choices for crystals, also called *phosphor screens*, are Caesium-Iodine (CsI) and Gadolinium Oxysulphide (GadOx). These crystals consist of high atomic-number elements that increase the probability of photoelectric interaction.

The scintillation event is responsible of the spatial correlation found in X-ray radiographs: The one-to-many X-ray conversion gives rise to a *photon shower* consisting of hundreds to thousands of optical photons. The optical photons are emitted in isotropic direction, and therefore spread out over a range of photodetectors at the scintillator exit plane. The shape of the resulting blur is quantified by a point-response function (PRF), defined as an image after illumination of the detector with an infinitely narrow X-ray pencil beam [157]. This is also commonly termed a PSF (point-spread function). The spread of the blur mainly depends on the thickness of the phosphor screen [158], but the specific blur shape depends on manufacturing details (e.g., the use of Thallium (TI) doping, transmission of charge through an amorphous silicon, or needle-grown structure of CsI). Manufacturers offer detectors with phosphor screens in a variety of thicknesses, as applications require different trade-offs between detection efficiency and spatial resolution. An analytical model of the blur can be approximated via Monte Carlo simulation [157].

When the PRF is assumed to be spatially-invariant over the radiograph, it can be modeled by a uniform convolution with a kernel h. We replace the notation for the idealized

radiograph I obtained from Beer-Lambert's law, equation equation (1.3) in Chapter 1, with the following three equations:

$$\begin{aligned} \lambda_i &:= I_i^0 \exp(-\textstyle\int_{l_i} \mu(\eta)\,\mathrm{d}\eta) && \text{(noise-free)} \\ \hat{I} &\sim \text{Poisson}\,(\lambda) && \text{(independent noise)} \\ I &:= h \circledast \hat{I} && \text{(correlated noise).} \end{aligned} \tag{5.5}$$

Here, λ is the noise-free signal, $\hat{I}$ is a multivariate random variable describing a radiograph with pixelwise-independent photon noise, and the radiograph I is now redefined via the discrete convolution of independent Poisson variables. At high photon counts, I approximates a multivariate normal distribution, i.e.,

$$I \sim \mathcal{N}(H\lambda, H\Lambda H^{\mathsf{T}}), \tag{5.6}$$

with H being the linear operator associated with h, and $\Lambda := \text{diag}(\lambda_1, \ldots, \lambda_{N_u N_v})$ a diagonal matrix. Note that the covariance matrix, $H\Lambda H^{\mathsf{T}}$, reflects the signal-dependent nature of Poisson noise.

The convolutional model equation (5.5) is a linear and uniform approximation of the detector response, allowing efficient deconvolution in tomographic pipelines (section 5.5.2). However, under certain specific imaging conditions, one or multiple of the following sources of model mismatch may become non-negligible: (i) The non-uniform nature of the PRF, due to its dependence on the incident X-ray angle and energy spectrum (the latter is known as Swank noise [159, 160]); (ii) Detector electronics, which can cause correlation through charge-sharing [161] or regional thermal effects in the photodetectors. While oftentimes negligible [146, 156], their effect is known to be more prominent at low photon counts; (iii) Scintillator responses become non-linear for high photon rates [3], and some detector instruments use heterogeneous pixel technology. In the forthcoming sections, we will investigate the correlation structures of real-world radiographs, and simulate anisotropic Gaussian noise, which is the commonly-assumed model for electronic noise.

5.3.2 Direct deconvolution

The correlation structure due to the PRF, as described in equation equation (5.5), cannot be addressed sufficiently by the blind-spot denoising strategies discussed in section 5.2.3. Convolution kernels, even with small radii, require large masking schemes or high subsampling factors, as the radius of correlation is twice the radius of convolution. This motivates our approach using *direct deconvolution*, which is available thanks to the noise model. Direct deconvolution is an approximate inverse to convolution, and an established technique in X-ray radiography for filtering noise and improving sharpness [162]. Using the Fourier transform $\mathcal{F}$, it is computed as

$$H^+(I) := \mathcal{F}^{-1}\left[\mathcal{F}[I] \cdot \frac{\mathcal{F}[h]^{\mathsf{H}}}{|\mathcal{F}[h]|^2}\right], \tag{5.7}$$

with $(\cdot)^{\mathsf{H}}$ denoting the conjugate transpose.

Equation (5.7) is efficient because convolution and deconvolution become element-wise multiplications in the frequency domain, and the Fourier transform can be computed quickly using the Fast Fourier Transform (FFT). When applied as a preprocessing step prior to a convolutional neural network, the cost of the deconvolution is typically negligible: CNNs contain already many convolutional blocks, each comprising multiple filters and working on multi-channel inputs.

Poisson case We first illustrate that, in the case of pure Poisson noise, deconvolution immediately enables blind-spot denoising. Recall that Noise2Self is based on the idea that the cross-term in the optimization objective, equation equation (1.14), should vanish. If we were to use the ordinary input, i.e., $u := I$ and $\hat{u} := H\lambda$,

$$\mathbb{E}\,\langle f_\theta(I) - H\lambda, I - H\lambda\rangle \tag{5.8}$$

would not be zero, as I_{J^c} is not statistically independent of I_J (cf. the covariance in equation equation (5.6)). However, preprocessing the image with a deconvolution gives $u := H^+ I \approx \hat{I}$. The covariance of u is then:

$$\mathbb{E}\,[u_i u_j] = (H^+ H \Lambda H^{\mathsf{T}} H^{+\mathsf{T}})_{ij} \approx \Lambda_{ij}, \tag{5.9}$$

which is a diagonal matrix, thus does not contain cross-correlations. In the theoretical case, with pure Poisson noise and a known PRF, the deconvolved images $u := H^+ I$ therefore restore the potential for blind-spot denoising.

Poisson-Gaussian case We now extend the model to real-world noise. Compared to the Poisson case, real-world data is further degraded with additive noise, which is known to produce high-oscillatory "checkerboard artifacts" after deconvolution. To see this, consider an additive anisotropic Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. By the linearity of deconvolution, we have that $H^+(I + \varepsilon) \approx \hat{I} + H^+\varepsilon$, showing that deconvolution recovers the pixelwise independent noisy image $\hat{I}$ at the cost of adding a checkerboard component $H^+\varepsilon$. Since the the checkerboard artifacts are spatially-structured, it adds a pattern that can be reproduced by blind-spot denoisers.

Instead of deconvolving with h, we will look for a $\tilde{h}$ that minimizes the overall spatial correlation better by balancing the two types of noise. Let $\widetilde{H}^+$ denote the associated deconvolution operator, and $u := \widetilde{H}^+(I + \varepsilon)$ the deconvolved radiograph. Similar to the Poisson

It is worth noting that $\hat{I}$ can be blurred more extensively than what is expected only from the scintillator PRF, due to, e.g., the finite focal spot aperture in the source, or instrument vibration [3, 142, 163]. Several methods exist for estimation and resolving such blurs, e.g., with a procedure involving a sharp-edge phantom [163]. However, using methods based on sharpness criteria for our purpose, which is decorrelation, can cause a model mismatch and lead to, e.g., ringing artifacts in iterative schemes. It is therefore that we will use a statistically-estimated PRF with direct deconvolution, and not the more sophisticated iterative Poisson-deconvolution algorithms such as Richardson-Lucy.

case equation (5.8), the goal is to maximize the statistical independence of the blind spot with the non-masked pixels. For a blind-spot u_i and a pixel u_j in the PRF's support, the covariance is

$$\mathbb{E}\left[u_i u_j\right] = (\widetilde{H}^{+}(H\Lambda H^{\mathsf{T}} + \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{N_u N_v}^2))\widetilde{H}^{+\mathsf{T}})_{ij}. \tag{5.10}$$

Since the inner term is the covariance matrix of $I + \varepsilon$, we can see that $\widetilde{H}^{+}$ should perform a diagonalization of the covariance matrix. In the case of a uniform signal λ_{const} and noise variance $\sigma^2_{\mathrm{const}}$, the closest solution to diagonalization would be given by the inverse of

$$\widetilde{H} = \left(HH^{\mathsf{T}} + \sigma^2_{\mathrm{const}}/\lambda_{\mathrm{const}}\mathrm{Id}\right)^{\frac{1}{2}}, \tag{5.11}$$

up to a factor of scaling (see Appendix 5.8 for the procedure and additional remarks). This shows that the additive noise has a regularizing effect. Note that $\widetilde{H} \to H$ as $\sigma^2_{\mathrm{const}} \to 0$, meaning that in the limit of smaller additive noise we return to the Poisson case, for which the scintillator PRF is optimal. In general, however, the solution to equation (5.10) can not be expressed analytically anymore.

5

5.3.3 PRF estimation

Algorithm 2 outlines a data-driven estimation procedure for the convolution kernel associated with $\widetilde{H}$. It relies on the availability of a set of radiographs $\{I^{(1)}, \ldots, I^{(T)}\}$ between which only the noise changes (for example, the radiographs acquired from a static object with a stationary source-detector set-up). It estimates a sample covariance matrix of the noise, and then extracts its latent convolution kernel. The presentation here excludes implementation details, such as how to sample at the image boundaries, or how to centralize and normalize the kernel. These can be found in the algorithm code in Appendix 5.7, and further details are given in Appendix 5.8.

The algorithm takes as input the radiographs, a temporal stride Δt, and a set of indices $\mathcal{J}_{\mathrm{corr}}$. Line 1 first removes the background signal via subtraction of $I^{(t+\Delta t)}$ from $I^{(t)}$. Compared to subtracting the mean, using radiograph differences can be more robust to slow source current fluctuation during an X-ray scan. Δt is a parameter that should be chosen such that noise in the frames is temporally uncorrelated. Line 2 transforms the noise such that it has uniform variance, and line 3 computes correlation coefficients for the pixels i and $i + j$, where $j \in \mathcal{J}_{\mathrm{corr}}$. Here, the index set $\mathcal{J}_{\mathrm{corr}}$, for instance a rectangular neighborhood, must be large enough to include all pixels within the correlation range of the pixel i. As the correlation distance is typically unknown, a user should increase $\mathcal{J}_{\mathrm{corr}}$ until sampling with a larger radius has no significant effect anymore on the algorithm outcome. After forming the sample covariance matrix, line 5 retrieves $\tilde{h}$ using the matrix square root [164].

For the data set of radiographs, we recommend a static object with similar experimental conditions as the denoising data set. When such data is unavailable, a static image region extracted from multiple frames of an arbitrary acquisition may already give good results. In all cases, it is necessary to avoid detector regions where the correlation is affected, e.g., via clipping of intensity values to the detection limit.

Algorithm 2:
In: Radiographs $\{I^{(1)}, \ldots, I^{(T)}\}$, temporal stride Δt, indices $\mathcal{J}_{\text{corr}} \subset \mathbb{Z}$
Out: Convolution kernel $\tilde{h}$

1: Subtracting the background signal gives the *noise differences*:

$$D^{(t)} \leftarrow I^{(t)} - I^{(t+\Delta t)}$$

2: Stabilize D using the sample variance for all pixels i:

$$M_i^{(t)} \leftarrow \frac{D_i^{(t)}}{\sqrt{s_{ii}^2}} \text{ with } s_{ii}^2 \leftarrow \frac{1}{T}\sum_{t=1}^{T} D_i^{(t)} D_i^{(t)}$$

3: Estimate correlation coefficients $\tilde{c}_j$ for all $j \in \mathcal{J}_{\text{corr}}$:

$$\tilde{c}_j \leftarrow \frac{1}{N_u N_v T} \sum_{i=1}^{N_u N_v} \sum_{t=1}^{T} M_i^{(t)} M_{i+j}^{(t)}$$

4: Construct a small Toeplitz sample covariance matrix $\widetilde{C}$ by setting the diagonals to $\tilde{c}_j$.
5: Compute the matrix square root of the positive semi-definite $\widetilde{C}$:

$$\widetilde{H} \leftarrow \sqrt{\widetilde{C}}$$

The kernel $\tilde{h}$ can be retrieved as the first row or column of $\widetilde{H}$.

5.3.4 Neural network

Deconvolved images can now be used as training data for blind-spot networks. For our experiments we will optimize the loss

$$\arg\min_{\theta} \; \mathbb{E} \; \|\log f_\theta(u) - \log u\|_2^2 \,, \tag{5.12}$$

with $u := \widetilde{H}^{+} I$. The L2-norm is well-known to overestimate high signal in radiographs due to heteroscedasticity of the noise. Several solutions are used in practice, such as a Poisson loss or Anscombe transform [165]. In the high photon count regime, an alternative is to take the log-transform, which achieves a loss that linearly relates to the reconstruction that we want to compute from the denoised data (see equation (1.4)).

For the network architecture f_θ of (5.12), we will employ a standard U-Net [38] in all experiments of the result sections. This is an image-to-image architecture with a symmetrical encoder and decoder, using convolutional downsampling operators and bilinear upsampling operators, respectively. In our implementation, the network consists of 256 feature maps in the first level, and uses skip connections between the encoder and decoder. We note that Noise2Self and Noise2Void can also be used in conjunction with other image-to-image architectures, such as DnCNN [166].

To make the U-Net $\mathcal{J}$-invariant, we make use of the default masking procedure of Noise2Self, which entails a zero-value replacement of pixels on a 3-by-3 grid. During in-

ference, the network is executed on all grids, and the resulting outputs are assembled into a single image.

5.4 Results I: Radiograph denoising

Before demonstrating our denoising approach on experimental data from a dynamic experiment without ground truths, we first validate each component in the pipeline of kernel estimation, direct deconvolution, and blind-spot denoising.

Correlation maps A useful quantity to visualize spatial correlation is the summed correlation of each individual pixel with its neighborhood. This quantity can only be computed with a sufficiently large set of static radiographs $\{I^{(1)}, \dots, I^{(T)}\}$. Denote the *correlation map* with $m \in \mathbb{R}^{N_u N_v}$. For a pixel $m_i(I)$ in the map,

$$m_i(I) = \sum_{j \neq i} \frac{\mathrm{Cov}(I_i, I_j)}{\sqrt{\mathrm{Var}(I_i)}\sqrt{\mathrm{Var}(I_j)}} \approx \frac{1}{T} \sum_{j \neq i} \sum_{t=1}^{T} M_i^{(t)} M_j^{(t)}. \tag{5.13}$$

Here, the first equality sums the Pearson correlation coefficients of pixels (I_i, I_j) and the second equality reuses the notation for variance-stabilized noise from algorithm 2.

5.4.1 Deconvolution of simulated radiographs

In this section, we will use radiographs that are simulated from a numerical foam [167], i.e., a numerical phantom that consists of a single-material cylinder with randomly-positioned spherical voids. This phantom will also be used in Results II (see figure 5.6 for a horizontal slice through the foam). Here, it is configured to contain 100 spheres per unit of cylinder, a low-dose current of $I^0 := 1000$, and a parallel beam geometry. We use the phantom software [167] to calculate the ground truth radiograph, see λ in (5.5), analytically, and then generate 5,000 128-by-128 noisy radiographs by adding simulated Poisson noise and additive Gaussian noise. This simulation yields repeated static radiographs, making them suitable for PRF estimation and visualization of local correlations with (5.13). We first test the accuracy of PRF estimation (algorithm 2), and then inspect the residual correlations after deconvolution with Gaussian noise.

From algorithm 2, errors are expected due to variance stabilization (line 2), a finite number of statistical samples (line 3), and the matrix solver (line 5). To convey an impression, the first row of figure 5.1 applies the algorithm on the 5,000 radiographs. The simulation only applies Poisson noise and blurs the result using a 5-by-5 Laplacian PRF (scale parameter $b = 0.5$). The accuracy of $\tilde{h}$, due to the variance stabilization (line 2 in algorithm 2), was 1.2% relative error after $T = 30$ static noisy images, and 1.1% after $T = 200$. Further improvement was not observed, due to nonuniform regions in the radiograph (see Appendix 5.8). The error due to the matrix square root was found to be negligibly small for the Laplacian kernel.

The three rows of figure 5.1 display three simple noise conditions for the simulated radiographs, namely: Pure Poisson noise, isotropic Gaussian noise, and anisotropic Gaussian

noise. In the pure Poisson case, the difference $\widetilde{H}^{+}I - \hat{I} \approx 0$ confirms that deconvolution with an estimated PRF still reconstructs the photon noise with very high accuracy. The histograms in the residual correlation plot furthermore confirm that deconvolution removes the correlations.

The isotropic and anisotropic cases of figure 5.1 illustrate two problems that can occur at the extremes of additive pixelwise independent Gaussian noise ε. At high levels of isotropic noise (middle row), correlations can become signal-dependent, i.e., the spatial distribution of high-intensity regions in $m(\widetilde{H}^{+}I)$ follows the radiograph intensities. Strongly anisotropic additive noise (bottom row) cannot be accounted for by a uniform kernel. When this noise is added to the lower image part, deconvolution partially decorrelates the photon noise in the upper image, while reintroducing negative-correlation checkerboard artifacts in the lower image. The positive correlation and negative correlation are visible as distinct peaks in the histogram.

The numerical experiments provide practical insights into PRF estimation. In scenarios with low photon counts or strongly anisotropic noise, uniform deconvolution is unable to fully eliminate the correlations. In the next section, we will therefore first examine the correlation maps of real-world radiographs. In the subsequent section 5.4.3, we explore how residual correlations after deconvolution are handled by blind-spot denoisers.

5.4.2 Deconvolution of an experimental phantom

In this section and section 5.4.3, the data consists of 1,000 repeated radiographs from two static "bubble phantoms". These radiographs show a PMMA (polymethyl methacrylate) cylinder that is filled with granular particles and two 23 millimeter polystyrene spheres. The set-up is described in detail in [168]: It consists of three fixed cone beam sources directed at three Xineos-3131 CsI scintillator detectors with CMOS pixel technology. The detectors are oriented as portraits and operate in a 1548-by-550 virtual region of interest to improve the read-out speed.

Using algorithm 2 with $T = 200$, $\Delta t = 3$ and an 11-by-11 correlation region $\mathcal{J}_{\text{corr}}$, we retrieve three kernels for the Xineos-3131 detectors. Since the differences between the kernels were found to be less than 0.5%, we limit the results to the first detector. Figure 5.2 shows a crop of the first detector radiograph, imaged with a tube voltage of 120 kVp and anode current of 1.5 mA. Using the 1,000 radiographs, the correlation map (equation equation (5.13)) before and after deconvolution is inspected. Before deconvolution, correlation is highly uniform, and an average pixel correlates about 356% with its surroundings, i.e., the mean of the black histogram. After deconvolution, the distribution is centered.

An important practical result is the uniform background of the correlation maps in figure 5.2. This indicates that, for this set-up, spatial correlation is not strongly dependent on the signal (i.e., the phantoms, cylinder or metal components are not visible in the correlation maps). Following the results of the previous section, we therefore expect the PRF to be approximately uniform, and the additive noise to be neither strong nor very anisotropic.

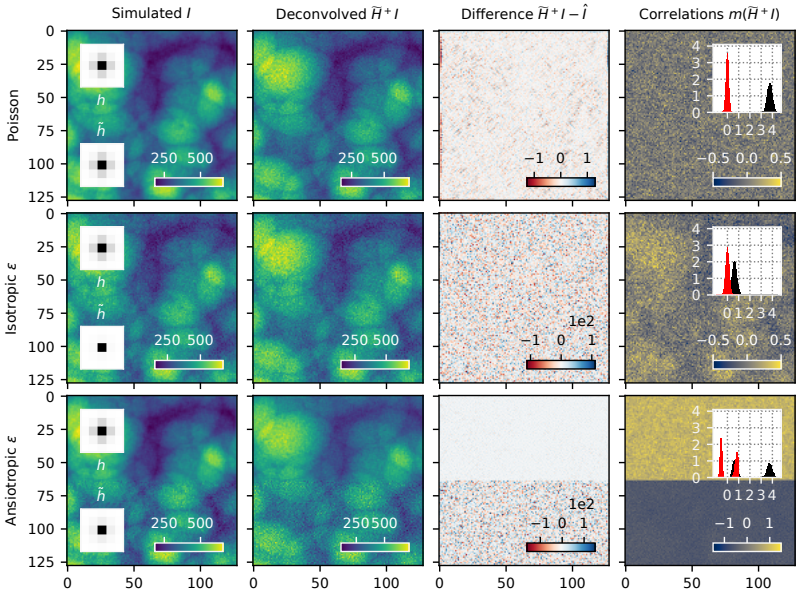

Figure 5.1: Kernel estimation and deconvolution evaluation: Simulated projection data of spheres impacted by high-noise (section 5.4.1). The **top row** contains pure Poisson noise. The **middle row** adds isotropic Gaussian noise with $\sigma^2 = 400$. The **bottom row** only adds the Gaussian noise in the lower half-plane. The insets on the residual correlation maps (right column, see equation (5.13)) show histograms of the summed correlation over the whole image before (black) and after (red) deconvolution.

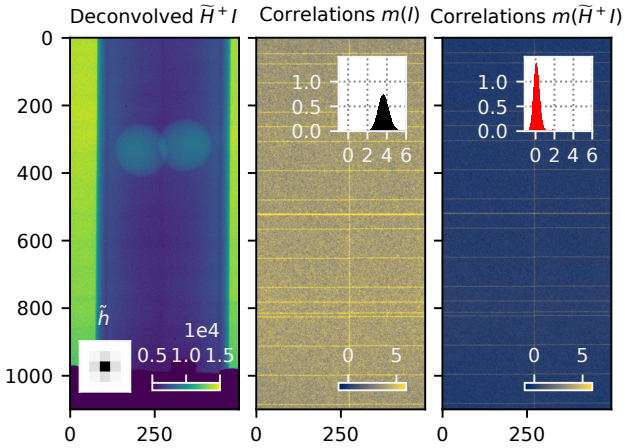

Figure 5.2: Deconvolution of 1,000 radiographs of a static PMMA cylinder containing two polystyrene balls, metal equipment, and non-uniform background radiation (cropped view, section 5.4.2). **Left** displays one radiograph and the kernel $\tilde{h}$. The **middle and right** show the correlation maps of I and $\widetilde{H}^+ I$. The lines of high correlation are due to inpainting of defective pixels during preprocessing. The probability density histograms show shifting and narrowing of the correlation distribution (black→red).

5.4.3 Blind-spot denoising with Noise2Self

We test the suitability of the deconvolved phantoms for blind-spot denoising. First, the U-Net is trained on the deconvolved radiographs $\{\widetilde{H}^+ I^{(1)}, \ldots, \widetilde{H}^+ I^{(1000)}\}$ of figure 5.2. As a ground truth to evaluate the denoising task, we use the mean (cf. equation (5.6)) over the deconvolved radiographs, i.e.,

$$\widetilde{H}^+ H\lambda \approx \frac{1}{T}\sum_{t=1}^{T} \widetilde{H}^+ I^{(t)}. \tag{5.14}$$

While training with static radiographs is not yet representative of a real-world data set, the experiment allows a better investigation of the network response to correlation structures.

Figure 5.3 shows a zoom-in on the noisy image, ground truth image, and two denoising results at different points during the training stage. The results show that the deconvolved data is denoised well and that the network even recovers some of the stacked granular particles that are difficult to discern in the raw radiograph. The ground truth equation (5.14) is not fully recovered by N2S: As we will see, this is not due to the remaining correlations, but is the usual performance of the denoiser. Noise overfitting occurs after extensive training, leading to artifacts in the image (cf. bottom right plot in figure 5.3). This is especially the case for the inpainted dead-pixel lines, which lead to a correlation structure that is more easily learned by the network (cf. figure 5.2).

To further validate our approach, we compare the resulting network against two baselines; f_{raw} using the original, i.e., correlated, noisy data, and f_{syn} using synthetic, uncorrelated, noisy data. Both are trained in exactly the same way as f. The synthetic data is generated by adding Gaussian noise to the ground truth using the sample variance extracted from the real data, and is therefore representative of a correlation-free baseline. Figure 5.4 compares the three networks and displays their evaluation on a single radiograph. For f_{raw}, the output remains noisy and displays checkerboard artifacts (see the figure inset). The denoising results, $\log f(\widetilde{H}^+ I)$, are very similar to the synthetic data, $\log f_{\text{syn}}(\widetilde{H}^+ H\lambda + \varepsilon_{\text{syn}})$. Both f and f_{syn} achieve similar denoising performance, although the synthetic network did not suffer from overfitting.

To see how the remaining correlation structure in the deconvolved data (i.e., the map $m(\widetilde{H}^+ I)$ in figure 5.2) affects f, we add different realizations of synthetic, uncorrelated noise ε to the same input $\hat{u}$ and calculate

$$p_i(f_\theta, \hat{u}) = \mathbb{E}_\epsilon \sum_{j \neq i} \frac{\operatorname{Cov}(f_\theta(\hat{u}+\epsilon)_i, (\hat{u}+\epsilon)_j)}{\sqrt{\operatorname{Var}(\epsilon_i)}\sqrt{\operatorname{Var}(\epsilon_j)}}, \tag{5.15}$$

which measures how much of the added noise is propagated from surrounding pixels to a target pixel. The result, the bottom row of figure 5.4, shows that f_{raw} relies everywhere on values from the local pixels, an indication that it learned to exploit correlations in the noise. On the other hand, f and f_{syn} only use the local neighborhoods where there are sharp image features (e.g., the dark spot and phantom edges have higher intensities in $p(f_\theta, \hat{u})$).

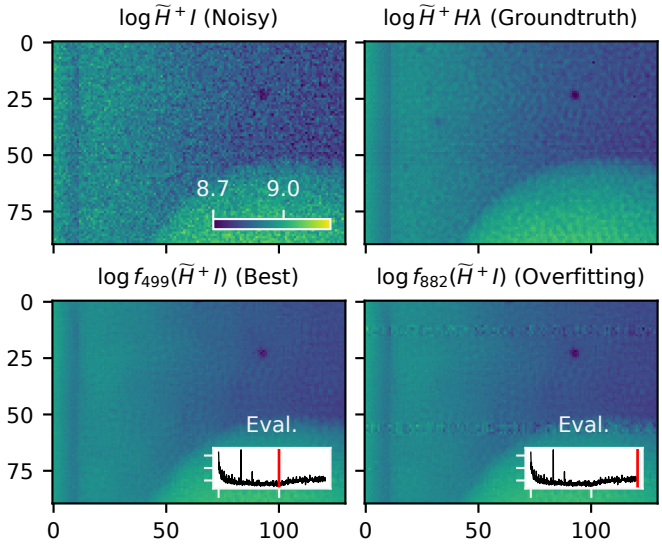

Figure 5.3: Training Noise2Self on the deconvolved data. **Top row**: Noisy and ground truth deconvolved radiographs (zoom into the upper-left part of figure 5.2). **Bottom row**: the trained Noise2Self U-Net evaluated on a single noisy image at an optimal point (epoch 499) and overfitting stage (epoch 882) during network training. *Eval.* plots the MSE evaluation of the trained network f_i at epoch i with the ground truth.

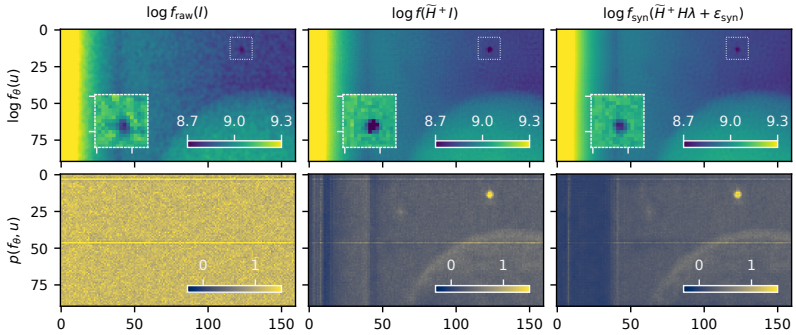

Figure 5.4: Blind-spot-denoised radiographs (figure 5.2), as well as their input-to-output correlation equation (5.15) summed over a 3-by-3 neighborhood (section 5.4.3). The **left plots** use unprocessed raw data, the **middle plots** deconvolved data, and the **right plots** synthetic independent noise. The granular background in the cylinder is a physical phenomenon due to stacking of particles. The output of f_{raw} displays checkerboard artifacts (inset in top left plot).

Moreover, both f and f_{syn} have the same average background values, an indication that f was not able to fit to remaining correlations in $m(I)$ in figure 5.2. Overall, the results in this section show that deconvolution can restore the blind-spot denoising performance of uncorrelated noise, but that care should be taken not to introduce easy-to-distinguish correlation structures such as inpainted pixels (figure 5.3).

5.5 Results II: Sparse-view Computed Tomography

For X-ray Computed Tomography, self-supervised denoising can be executed either in the projection domain [169, 170], e.g., via a Noise2Noise-like mapping between adjacent projections, or in the reconstruction domain, e.g., via reconstructions of angular subsets [40, 171] or via neighboring slices in a parallel-beam reconstruction [172]. The reconstruction domain has the advantage that the image features are compact and local, favoring denoising with CNNs, and often provides lower noise statistics due to averaging of noise from multiple projection angles [3, 4]. On the other hand, noise is spatially local in the projection domain, and can be more difficult to remove once propagated over the reconstruction domain by the CT algorithm. Self-supervised methods that work in both domains [173] are promising but are still too difficult to be applied for high-resolution volumetric tomographic data due to their high computational demands.

One promising use of BSNs is projection-domain denoising for *sparse-view CT*, i.e., reconstruction from an undersampled set of angular projections [174]. While dense samplings are typically required to recover high-resolution object details [4], sparse-view CT is advantageous when very short scanning times and/or very low radiation exposures to the sample are crucial. In section 5.5.1, we first compare self-supervised denoising in projection and reconstruction domain on simulated sparse-view CT data. Then, in section 5.5.2, we apply blind-spot denoising to real radiographs from dynamic ultra-sparse view CT using our PRF estimation and direct deconvolution pipeline.

5.5.1 Denoising of projections and CT images

In this experiment, we compare our denoising approach with the self-supervised denoiser in the reconstruction domain using different numbers of scan angles N_ψ and for two reconstruction methods, namely FBP and SIRT. While such comparison ultimately relies on the noise characteristics and capacities of the machine learning architecture, we will show that projection-domain denoising can become advantageous in the case of sparse-view geometries.

For projection-domain denoising, we will use Noise2Self, as before, and for reconstruction-domain denoising Noise2Inverse (N2I) [40]. In its simplest configuration, N2I splits projections into sets of even and odd indices, enabling even-to-odd denoising of filtered-backprojection (FBP) slices. This resembles Noise2Noise [39] on reconstructed images (section 5.1, section 5.2.2). During evaluation, even and odd network outputs are averaged.

The simulated radiographs are again of the parallel-beam numerical data used in section 5.4.1 [167], now with $N_u, N_v = 512, 128$ pixels. We use $I^0 = 10^4$ photons for generating the signal. However, as pure Poisson noise is too optimistic compared to the noise levels

in real-world data, we add 10^4 counts of zero-mean Poisson fluctuations. This follows a Gaussian distribution closely [3], and reflects a radiograph of which 50% of the signal is due to scatter. We do not model the scintillator blur, as this experiment aims to isolate the aspects of denoising in the two domains, and deconvolution can be performed with either method. The image features are therefore more complex in the projection domain (see figure 5.1) than in the reconstruction domain (figure 5.6). Both N2I and N2S are trained with the same U-Net configuration and stopped before noise fitting using a ground truth criterion.

Figure 5.5 displays the results after training and reconstruction on two reconstruction algorithms: FBP and SIRT (section 5.2.1). Focusing on FBP first, in (a), we see N2I performs remarkably well, as it even outperforms FBP reconstruction from clean projections. Figure 5.6 shows that this is possible because N2I smooths sparse-angle artifacts in image space. N2S, on the other hand, follows the lower limit of the shaded area, meaning it is close to the FBP solution with clean projections. For denser angles, its error (note the $\log_2$-scale) contributes significantly to the reconstruction. We observed that the high error of N2S is mostly due to the low number of training samples (the N_ψ projections) relative to the high image complexity.

In (b) of figure 5.5, N2I and N2S are combined with SIRT reconstruction using box constraints to restrict valid solutions to $[0, 1]$, cf. equation equation (5.3). The box constraints, which turn SIRT into a non-linear method, are chosen to illustrate that N2S can both be used with linear and non-linear methods, and will also be used in section 5.5.2. The algebraic reconstruction technique is known to be better-suited for sparse-angular geometries [4]. By our knowledge, N2I has not been demonstrated successfully with SIRT results before. Our results indeed show artifacts in the N2I-denoised images, suggesting that N2I may not be able to resolve the finer object details, possibly due to the complex noise structure after the iterative reconstruction. Blind-spot denoising with N2S, on the other hand, is independent of the reconstruction algorithm that follows. SIRT started with N2S-denoised data allows therefore for a more accurate reconstruction than N2I in the sparse-angle segment. Figure 5.6 shows that for the sparse-angle case with $N_\psi = 32$, the best denoising strategy is to combine N2S with SIRT.

5.5.2 Ultra-sparse CT of single-bubble-injected fluidized beds

We close the article by blind-spot denoising real-world projections of gas-solids fluidized beds and performing a subsequent SIRT reconstruction. A gas-solids fluidized bed is a mixture between a gas and a granulate material that behaves similarly to a fluid. Bubbles, i.e., the voids in the bed, are studied in laboratory-scale experiments, both via X-ray radiography and timeframe-by-timeframe dynamic CT. Typically, statistical quantities about the shapes and sizes of bubbles are inferred using image analysis techniques such as segmentation and tracking. In this particular example, we look at single-bubble-injected data: The injected gas is regulated at the set-up inlet such that only a single bubble travels through the granulate material at the time.

The ultra-sparse set-up at Delft University of Technology is specifically built to image

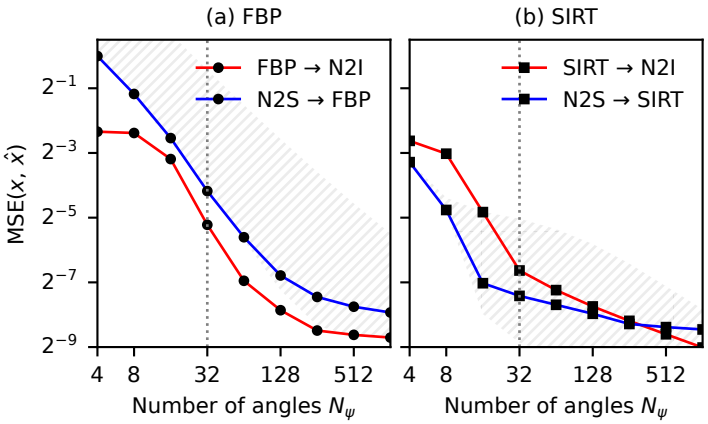

Figure 5.5: N2S and N2I combined with N_ψ sparse-view FBP and SIRT reconstructions (section 5.5.1) of the spherical phantoms [167], see figure 5.6 for a plot at the vertical indication line. Each data point is an individually trained U-Net [38] on N_ψ angles that is optimally stopped using ground truths. The limits of the shaded areas mark reconstruction from clean and noisy data.

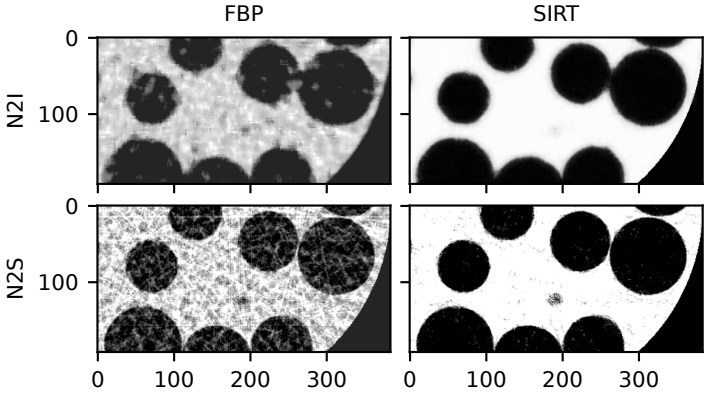

Figure 5.6: 192-by-384 region of interest taken from the $N_\psi = 32$ simulated reconstructions (see figure 5.5) at the central volume slice of the binary phantoms [167]. See figure 5.1 for an impression of the projections. A circular mask is applied to reduce the contribution of artifacts outside of the object in the MSE score of figure 5.5.

the high-velocity bubbles while they travel towards the bed's surface. It consists of three pairs of sources and Caesium-Iodine detectors (section 5.4.2), positioned into an equilateral triangle. The synchronized radiographs obtained from the set-up suffer extraordinarily from noise, on one hand due to their short exposure intervals used in the X-ray detectors, and on the other hand due to the cross-scatter of photons from the non-facing sources. This significantly deteriorates subsequent image reconstructions [168], since high levels of noise in iterative methods lead to early semi-convergence, i.e., SIRT can overfit to noise in the reconstruction [4].

In this set-up no paired noisy images are available for image-to-image denoising, and, as shown in section 5.5.1, self-supervised denoising in the reconstruction domain is challenging in the case of SIRT. Blind-spot denoising, on the other hand, is still able to take advantage of the large amount of imaged bubbles using the similarity within the data of an experiment. Moreover, for this data set, it is more computationally efficient than deep-learned denoising in reconstruction space, as the 3-tuple projections has a lower dimensionality than the volume (each timeframe has 3×1548×550 pixels and each volume has 1548×550×550 voxels, respectively).

In figure 5.7 we show a noisy and blind-spot denoised bubble of a single timeframe, after training on a subset of 1,450 timeframes of single-bubble-injection experimental data. We estimated the kernel on a static part of the experimental data, and used the same direct deconvolution and network architecture as in section 5.4.3, but preprocessed the 3×1450 projections via a tailored preprocessing procedure to remove the PMMA cylinder (figure 5.2, Chapter 2). Qualitatively, the blind-spot denoiser is able to recover both sharp and smooth features of the bubbles well. Figure 5.8 visualizes the subsequent SIRT reconstructions. The noisy bubble is visible as a denser attenuation in a highly noisy field, and we were only able to visualize it by clipping the attenuation values to a narrow range. On the other hand, the denoised bubbles provide a much better contrast in the reconstruction, have gradual interfaces, and we expect will be better suited for further statistical analysis of the fluidized beds.

5.6 Discussion

Direct deconvolution is an efficient and effective approach enabling self-supervised denoising of real-world X-ray radiographs by reverting scintillator-induced correlations. For the deconvolution kernel, an empirically-obtained estimate of the scintillator point-response function has an implicit regularizing effect, as it balances the introduction of new correlations as checkerboard-artifacts with the removal of existing correlations. In this article, we demonstrated that the denoising potential of Noise2Self is restored, and that no ground truths or paired noisy training examples are required. In the results, we highlight its importance for projection-domain denoising via an example of real-world dynamic sparse-view CT.

Our approach tested well on raw data from Caesium-Iodine detectors in the presence of high photon counts, high noise due to scatter, and low anisotropic additive noise. We expect this scenario to be representative for cone beam set-ups where the photon noise

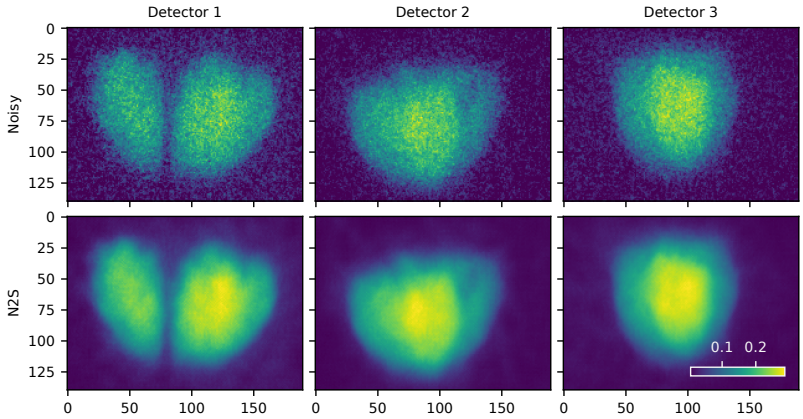

Figure 5.7: Top row: preprocessed projections (Chapter 2) of the single-bubble-injection experiment at timeframe 238/4500 (view is restricted to the bubble). Detector 1 shows that the bubble has split in two. **Bottom row**: Noise2Self network outputs.

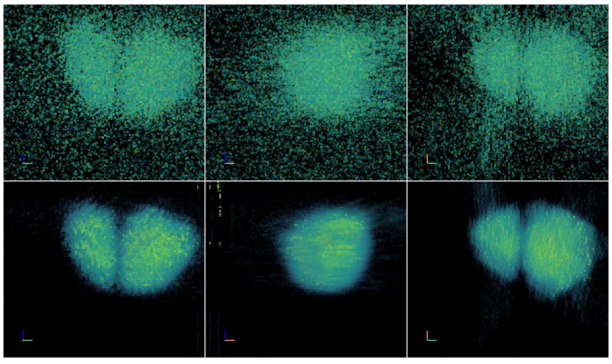

Figure 5.8: Ultra-sparse SIRT reconstructions of the noisy and denoised projections displayed in figure 5.7. **Left**: Side view along the projection axis of detector 1. **Middle**: Side view along the horizontal axis that is orthogonal to the projection axis. **Right**: Top view.

is dominant and the scintillator blur approximately uniform. It would be interesting to benchmark the method in the limits of different noise conditions [161, 170], such as the low-intensity radiation used in biomedical microscopy, monochromatic or filtered X-ray energy spectra, and different detectors, e.g., thick monolithic scintillators [142].

In algorithm 2, we sampled correlations from radiographs of a static object, and used these to find a suitable deconvolution kernel. While this approach is efficient, its estimation error may prove too large for certain situations, e.g., with limited quality sampling data, for detectors with a large blur radius, or for set-ups that utilize a varying tube current [138]. In this case, solving the ground truth λ and kernel $\tilde{h}$ simultaneously, e.g., with a blind deconvolution algorithm, could be investigated. At the same time, uniform deconvolution becomes less effective in the presence of anisotropic and nonlinear noise. In order to move to a more sophisticated model, such as non-uniform deconvolution, it is necessary to examine the correlation structures in real-world radiographs in more detail, and in particular the way that they depend on the signal.

An open question is if kernel estimation and deconvolution can be integrated together with self-supervised learning, for example as an additional loss penalty [45]. Doing so is not straightforward: When both the deconvolved and convolved data are present in the loss, e.g., as a "Deconvolved2Convolved" strategy, neural networks can collapse into a convolution, in which case the noise is still propagated. Further investigation is also needed to see if pre-removal of non-Poisson noise prior to deconvolution, or learned deconvolution [175], could prove advantageous.

Acknowledgements

We thank Evert Wagner, Sophia Podber, and Luis Portela of Delft University of Technology for the radiograph data of polystyrene phantoms and single-bubble-injected fluidized beds. This work was supported by the Dutch Research Council (NWO, project numbers 613.009.106, 613.009.116, and VI.Vidi.223.059).

5.7 Appendix: Source code availability

Software is available at `github.com/adriaangraas/scintillatordecorrelator`. The radiographs of experimental phantoms (section 5.4.2) have been published on Zenodo [176]. Data of fluidized beds is available upon reasonable request.

5.8 Appendix: Estimation of the deconvolution kernel

Statistical moments of convoluted radiographs We consider radiographs given as $I := H\hat{I} + \varepsilon$, with $\hat{I} \sim \text{Poisson}(\Lambda)$ and anisotropic $\varepsilon \sim \mathcal{N}(0, \Sigma)$ with $\Sigma := \text{diag}(\sigma_1^2, \ldots, \sigma_{N_u N_v}^2)$. Their mean estimates the clean convoluted image:

$$\mathbb{E}[I] = \mathbb{E}[h] \circledast \mathbb{E}[\hat{I}] + \mathbb{E}[\varepsilon] = h \circledast \lambda. \tag{5.16}$$

Linearity of the covariance with independence of $\hat{I}$ and ε gives

$$\operatorname{Cov}(H\hat{I}+\varepsilon) = H\operatorname{Cov}(\hat{I})H^{\mathsf{T}} + \operatorname{Cov}(\varepsilon) = H\Lambda H^{\mathsf{T}} + \Sigma, \tag{5.17}$$

and each (i,j)-th entry can be written as

$$\begin{aligned}\operatorname{Cov}(H\hat{I}+\varepsilon)_{ij} &= \textstyle\sum_k \sum_l H_{il}\Lambda_{lk}H^{\mathsf{T}}_{kj} + \Sigma_{ij} && (5.18)\\ &= \textstyle\sum_k h_{i-k}\lambda_k h_{j-k} + \Sigma_{ij} && (5.19)\\ &= \textstyle\sum_k h_k h_{j-i+k}\lambda_{i-k} + \Sigma_{ij}, && (5.20)\end{aligned}$$

where the second equality uses the definition of convolution, $H_{ij} = h_{i-j}$ and $H^{\mathsf{T}}_{ij} = h_{j-i}$, and the third equality shifts the indices.

5

Statistical moments of radiograph differences Let I and I' be independent observations. For the difference $I - I'$, the mean is

$$\mathbb{E}[I - I'] = \mathbb{E}[I] - \mathbb{E}[I'] = 0, \tag{5.21}$$

using equation (5.16), whereas the covariance is given by

$$\operatorname{Cov}(I - I') = \operatorname{Cov}(I) + \operatorname{Cov}(I') = 2\operatorname{Cov}(I). \tag{5.22}$$

Similarly, $\operatorname{Var}(I - I') = 2\operatorname{Var}(I)$.

Kernel estimation problem The goal is to find a deconvolution operator $\widetilde{H}^{+}$ to transform $H\hat{I} + \varepsilon$ into $\hat{I}$, i.e., an image with statistically independent noise. Equation equation (5.9) and equation (5.10) in section 5.3.2 explain that we would like $\widetilde{H}^{+}$ to perform a diagonalization of the covariance matrix:

$$H\Lambda H^{\mathsf{T}} + \Sigma = \widetilde{H}\Lambda\widetilde{H}^{\mathsf{T}}. \tag{5.23}$$

Rearranging the terms, and using equation (5.20), shows that the kernel $\tilde{h}$ of $\widetilde{H}$ must obtain

$$\sum_k (h_k h_{j-i+k} - \tilde{h}_k\tilde{h}_{j-i+k})\lambda_{i-k} + \Sigma_{ij} = 0 \tag{5.24}$$

for all elements (i,j). Achieving zero correlation, however, is not possible, as a solution $\tilde{h}$ cannot simultaneously cancel the signal (λ_{i-k}) and noise (Σ_{ij}) terms.

Correlation coefficients Algorithm 2 instead finds a minimizer of equation (5.24) via the use of correlation coefficients. Denote with

$$M := \frac{I - I'}{\sqrt{\mathrm{Var}(I - I')}}, \tag{5.25}$$

the variance-stabilized noise image (division is pixelwise). The covariance between M_i and M_j computes to

$$\begin{aligned}
\mathrm{Cov}(M_i, M_j) &= \frac{2\,\mathrm{Cov}(I_i, I_j)}{\sqrt{2\,\mathrm{Var}(I_i)}\sqrt{2\,\mathrm{Var}(I_j)}} && (5.26)\\
&\approx \frac{\lambda_i \sum_k h_k h_{j-i+k} + \Sigma_{ij}}{\lambda_i \sum_k h_k^2 + \sigma_i} && (5.27)\\
&= \frac{1}{\alpha_i} \sum_k h_k h_{j-i+k} + \Sigma_{ij}/\lambda_i && (5.28)\\
&=: \frac{1}{\alpha_i} \tilde{c}_j. && (5.29)
\end{aligned}$$

The first equality uses equation (5.22), the second equality applies equation (5.18) and uses $\lambda_i \approx \lambda_{i-k}$ to remove λ_{i-k} from the sum. The third and fourth equalities introduce the notation for the correlation coefficients and scaling parameters $\alpha_i = \sum_k h_k^2 + \sigma_i^2/\lambda_i$.

Under the smoothness assumption, $\lambda_i \approx \lambda_{i-k}$, it is straightforward to see that the coefficients $\tilde{c}_j$ provide a direct solution to equation equation (5.24). From here, finding a suitable deconvolution kernel entails finding a solution $\tilde{h}$ that best fits $\tilde{c}$. The accuracy of $\tilde{c}_j$ depends on the number of samples (cf. line 3 in algorithm 2), which in practice is improved by averaging $\tilde{c}_j$ over a set of pixels in the image.

Solution using the matrix square root Since the resulting kernel must be uniform, it suffices to formulate the optimization problem for a single pixel (i.e., one row of equation (5.23)). To do so, we form a small Toeplitz sample correlation matrix $\widetilde{C}$, using $\tilde{c}_j$ on the j-th diagonals. The matrix is then decomposed into convolution operators via the matrix square root, i.e., $\widetilde{H} \approx (\widetilde{H}\widetilde{H}^\intercal)^{1/2} = \widetilde{C}^{1/2}$ using a blocked Schur algorithm [164]. From here $\tilde{h}$ can be extracted as the first row or column.

Additional remarks

- *Smoothness assumption*: In equation equation (5.27) we have assumed $\lambda_i \approx \lambda_{i-k}$, which enables retrieving $\widetilde{H}$ via fast sampling (algorithm 2). The approach is justified for most radiographs, as their ground truth often consists of homogeneous or smooth regions, which bias the estimates $\tilde{c}_j$ towards a correct value. While we find our assumption to work well in practice (section 5.4.1), for situations where it would lead to an error outside of acceptable bounds, the sampling region can be restricted to a smaller or smoother region of the radiograph. We further reflect on this assumption in the discussion.

- *Scaling*: In X-ray imaging set-ups, radiograph intensities are calibrated to represent physical attenuation coefficients [3]. To ensure deconvolution does not interfere with

the calibration process, we normalize the found kernel $\tilde{h}$ such that its integral value equals one.