From pixels to patterns: AI-driven image analysis in multiple domains

Javanmardi, S.

CHAPTER 1

Introduction

1.1 Background

In the digital age of today, our interaction with the world is primarily shaped by our senses. Among these senses, the visual experience plays a pivotal role in our everyday lives. Through our ability to observe and process various scenes, ranging from street views and entertainment to advertisements and personal communication, we gain a deeper understanding of our surroundings. Notably, these scenes manifest themselves in the form of retinal images, which permeate our lives incessantly. As a result, we find ourselves constantly engaged in the processing of diverse images, showcasing our remarkable proficiency in this domain.

Digital images serve as a medium for conveying information in various aspects of our daily lives, including education, biology, entertainment, and advertisement. The proliferation of images can be attributed to technological advancements such as digital cameras, smartphones, and high-speed internet. In fact, in 2023, an estimated 6.92 billion people used smartphones, leading to a staggering 1.81 trillion photos taken worldwide each year (Adhikari and Roy, 2024). With 57,246 photos captured every second or 5.0 billion per day, these images are frequently shared on social media platforms, profoundly influencing our daily experiences (Kontogianni et al., 2024).

As humans, we possess the remarkable ability to attribute meaning to images through semantic interpretation. For example, when we encounter an image containing a dog, we label it as a scene with a dog. This semantic connection is made possible by the vast amount of experience and patterns stored in our brains. By ascribing semantic meaning to images, we understand their content and also extract value from them. The question arises: How can we make inferences and derive semantic meaning from images? In fact, our brain does this.
If we ask the question as to whether this task can be transferred to Artificial Intelligence (AI), the concept of AI comes into play. AI refers to the development of computer systems capable of performing tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language understanding. AI replicates and simulates human capabilities, often relying on training from past experiences to enable computers to comprehend scenes, much like humans.

In Figure 1.1, a description in a comprehensive form of the workflow of image analysis is portrayed by using DALL.E (Ramesh et al., 2021) in two main stages.

On the left side, a researcher is engaged in the acquisition of images, meticulously capturing various samples with precision using photographic equipment and a microscope. This represents the crucial phase of collecting raw data in the form of visual representations. On the right side, she is immersed in computer-aided image analysis. Through specialized software, the data scientist extracts distinctive features from the images, facilitating a more in-depth study and understanding of the represented images. This stage signifies the transformation of raw visual data into computational metrics, ensuring a structured research approach.

Figure 1.1: Stages of Image Analysis: From Acquisition to Feature Extraction.

AI techniques encompass a wide range of approaches, including rule-based systems, expert systems, and Machine Learning (ML). Among these, Deep Learning (DL), a subset of Machine Learning, has emerged recently as a particularly powerful method for analyzing visual content. Deep Learning algorithms are capable of processing complex, high-dimensional data, such as images, making them well-suited for addressing the challenges posed by the vast amount of visual data available today. This thesis specifically addresses the use of Deep Learning in different domains of image processing. In Figure 1.2, we illustrate the core architecture of the models utilized in this thesis, which are centered around Deep Learning techniques. This architecture encapsulates the multi-faceted approach of our Deep Learning networks utilized across various research stages in this thesis.

It functions systematically, processing input data across different models to perform

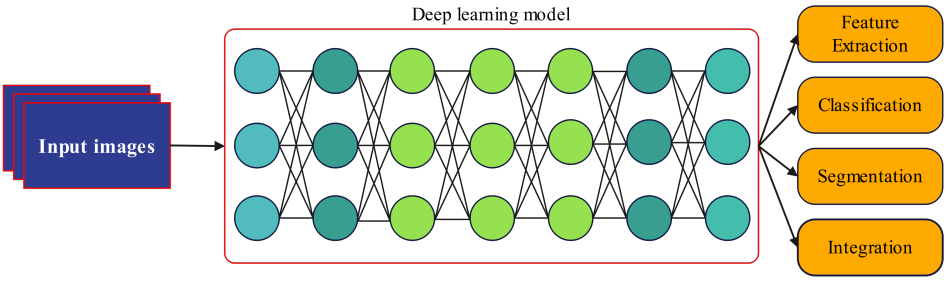

Figure 1.2: Core structure of models using Deep neural networks in our research.

tasks such as feature extraction, classification, segmentation, and integration. The strength lies in the ability of models to discern complex details from input images at various levels of abstraction, making it a cornerstone of modern image analysis techniques. The aforementioned tasks collectively address the range of research questions posed in our study. The structured, sequential design (of Figure 1.3)is reflective of established Convolutional Neural Network methodologies, ensuring a robust foundation for Deep Learning applications within our research. CNN is regarded as an excellent model in Deep Learning, characterized by its sequential layers that systematically process input data. The fundamental structure of a CNN is exemplified in Figure 1.3, which schematically represents the layered architecture crucial for various levels of image abstraction.

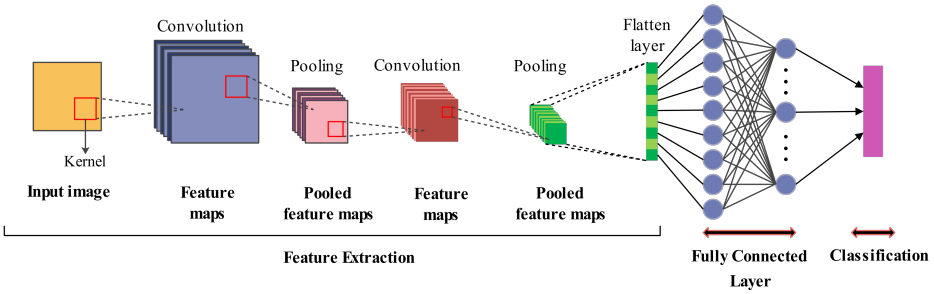

Figure 1.3: General structure of a Convolutional Neural Network.

In this architecture, each convolutional layer is typically followed by pooling layers. Convolutional layers focus on detecting patterns in the input, while pooling layers reduce the spatial size, parameters, and computation, preventing overfitting. Following pooling layers, there is a batch normalization step, which standardizes

the inputs to the subsequent layer, thereby accelerating the training process and providing some regularization. Batch normalization is crucial for numerical stability and network convergence. This layered configuration, forming the encoder, efficiently captures essential features in input images. The decoder, utilizing upsampling, enables precise localization for tasks like segmentation a foundational structure for our methodology in complex image processing tasks.

Pooling layers contribute to spatial size reduction, parameter and computation reduction, and prevention of overfitting by condensing the representation. In addition, regularization techniques, such as batch normalization, are applied to standardize inputs, expediting the training process and enhancing numerical stability. This regularization is crucial for maintaining network integrity and promoting convergence. This is crucial for maintaining numerical stability and improving the convergence of the network. Additionally, pooling layers are interspersed between convolutional layers to reduce the spatial size of the representation, diminish the number of parameters and computation in the network, and hence, prevent overfitting.

This configuration represents the encoder part of the network, which is designed to compactly capture the salient features of the input images. The decoder part then follows a reverse process, utilizing upsampling techniques to spatially expand the encoded features to allow for precise localization keys for tasks such as segmentation. This encoder-decoder structure is pivotal to our methodology, providing a robust framework for the complex image processing tasks addressed in this research. Unlike traditional Machine Learning techniques that necessitate manual feature engineering, Deep Learning models can automatically learn relevant features from raw data. This automation results in improved performance and reduces the need for human intervention.

In this thesis, we leverage Deep Learning methods for image processing to advance various fields, including biology, agriculture, and Natural Language Processing (NLP). These methods are selected for their superior ability in visual data analysis, their capacity to generalize, and their adaptability to new domains. Our objective is to study the use of Deep Learning techniques to devise innovative architectures and strategies that address complex research questions associated with NLP and computer vision in these domains.

The benefits of image processing and Deep Learning techniques are evident in various domains of science. They offer improved visualization, enabling the identification of patterns, features, and anomalies in scenes captured by images. Furthermore, image processing allows for the extraction of quantitative information, facilitating accurate decision-making and performance evaluations. Additionally, image quality can be enhanced through noise reduction, distortion correction, and contrast and sharpness adjustments.

Within the biotechnology in the agricultural domain, a significant focus of this thesis, image processing techniques hold the potential to enhance agricultural yield and refine agricultural processes. Deep Learning models can predict crop yields by analyzing historical data together with conditions, enabling informed decision-making for farmers. Moreover, these techniques can detect plant diseases from images, allowing for early interventions. Additionally, Deep Learning can optimize irrigation, fertilization, and pesticide application, reducing resource consumption and minimizing environmental impacts through precision agriculture. In parallel with our exploration of agriculture, this thesis also incorporates image-processing techniques within the realm of biology. This is driven by the recognition of the transformative potential these techniques hold for various biological studies and processes.

As a part of our research, this thesis explores the integration of image processing techniques in NLP, with a specific emphasis on the challenges and intricacies of image captioning. Image captioning stands as a compelling interdisciplinary endeavor that merges the visual with the textual, necessitating both a nuanced understanding of image content and the ability to express this understanding through coherent, human-like language. Central to this exploration is the concept of 'concepts' the semantic entities and relationships within images that imbue them with meaning beyond mere pixel matrices. These concepts are critical; they transform an image from a simple array of pixels into a canvas of inherent meaning and context.

Traditional image processing tasks, such as noise reduction and edge detection, typically focus on the visual aspect and do not require an understanding of image content. Hence, the absence of 'concepts' in these tasks is not seen as a limitation. However, the scenario shifts dramatically when we consider image captioning. Here, the objective is not just to process images visually but to comprehend and articulate the content within. Thus, concepts play an indispensable role in image

captioning, acting as the conduit through which the visual content is interpreted and described in language. They ensure that our captions are not only accurate but also richly connected to the true essence of the image.

In addressing the pivotal role of concepts, this thesis leverages Deep Learning methodologies to tap into their ability to learn and represent complex abstractions from data. This strategy enables the extraction and interpretation of latent concepts within images, facilitating the generation of insightful, human-like captions. These captions transcend mere descriptions, capturing the context and semantic depth of the visual content. Through this synthesis of image processing and NLP, we underscore the potential for more nuanced and intelligent systems capable of engaging with the visual world in a meaningful way.

The early stages of our research benefited from datasets with uniform, simple backgrounds, which simplified feature extraction and segmentation for the Deep Learning models. This controlled environment allowed us to establish baseline performance metrics and refine our CNN architectures. Building on these initial successes, we are now advancing to tackle more complex datasets. These upcoming datasets feature varied and dynamic backgrounds, closely simulating real-world scenarios where images rarely come with consistent backdrops. This complexity poses new challenges for our models, requiring them to distinguish and focus on the main subjects amidst a plethora of visual information. This progression is crucial for testing the adaptability of our models and enhancing their generalizability, preparing them for real-world applications where simple backgrounds are the exception rather than the norm.

Inspired by the pioneering applications and successes of CNNs in this field, we are motivated to explore beyond foundational principles. By delving into the semantics of image content, we aim to extend our understanding and capabilities, illustrating the rich potential of combining image processing with NLP to create more sophisticated and context-aware systems.

1.2 Research Questions and Contributions

This thesis explores the multifarious applications of Deep Learning for interpreting image data, tackling four critical tasks: feature extraction, classification, segmentation, and integration. It stands at the forefront, merging Deep Learning

with domains such as biotechnological applications in agriculture, biological sciences, and computational linguistics. This cross-disciplinary strategy highlights our dedication to pushing the boundaries of Deep Learning to solve intricate and industry-specific challenges.

Capitalizing on the innovative application of Deep Learning techniques in the advancement of agricultural sciences, this research introduces its first set of questions in the context of a case study centered on corn seed image analysis. These questions are designed to explore the potential and effectiveness of Deep Learning methodologies in enhancing the accuracy and efficiency of corn seed variety classification based on single-instance image processing. Our aims are exploring the robustness and accuracy of CNN in identifying specific features of corn seeds, crucial for maintaining seed purity and optimizing crop yield. The efficacy of CNNs as feature extractors in classifying corn seed varieties, regardless of seed orientation, is a pioneering approach in this field. This investigation aims to uncover the specific features and aspects of corn seed images that Deep Learning algorithms can most effectively utilize for precise and reliable classification. This research dives into the dynamic field of biotechnology in agriculture, with a particular focus on crop variety identification and feature extraction via sophisticated Deep Learning techniques. The study aims to make significant contributions to distinguishing between different crop varieties effectively. Therefore, our first set of research questions are raised as follows:

RQ1: Is it possible to accurately classify different corn-seed varieties by analyzing single-instance seed images using Deep Learning, focusing on specific image-derived features?

SubQ1: In the context of successful classification, which specific features extracted by Deep Learning algorithms contribute most to determining the accuracy of these classifications?

The result of this study is published in (Javanmardi et al., 2021). As demonstrated in the first research question (RQ1), we find ourselves motivated to further explore the efficacy of CNNs in the context of agriculture. The application of CNNs in the classification of corn seed varieties has provided convincing evidence of its robustness and accuracy, proving its utility beyond traditional methods. These advancements have preserved the purity of seeds and enhanced the yield of crops.

They provide a versatile strategy that accommodates differences in the orientation of seeds.

This research serves as a pioneering instance of using a CNN as a feature extractor for classifying multiple corn seed varieties, proving its efficacy regardless of seed orientation on a conveyor. Given the pressing need for accurate identification of corn seed varieties to ensure crop purity and yield, the study focuses on identifying specific features extracted from corn seed images that would lead to the most precise classification of different corn seed varieties. This is of utmost significance for seed producers and farmers, as the distinction between seed varieties impacts production quality as well as profits. Particularly for corn, the problem is exacerbated by the substantial overlap in morphological and color characteristics among seed varieties which are known as hand-crafted features.

The potential for unauthorized distribution of lower-quality seeds as premium varieties further underscores the urgency of precise classification and sorting techniques. While conventional methods such as visual inspection suffer from error rates and complexity, automated image processing offers a non-destructive, cost-effective solution. This study seeks to contribute to this area by delving into advanced computer vision techniques, specifically focusing on feature extraction by Deep Learning models, known to significantly influence classification accuracy. Thus, the research aims to provide a valuable solution to the ongoing challenge of accurately classifying visually very similar corn seed varieties, thereby safeguarding agricultural interests and market integrity while overcoming the limitations of existing methodologies.

Expanding upon the successful application of Deep Learning in the field of biotechnology in agriculture as explored in RQ1, particularly its proficiency in feature extraction, we transition to our second research inquiry (RQ2). In RQ2, we explore a critical component of food processing: utilizing Deep Learning for classifying mulberry ripeness stages. This inquiry aims to harness the capabilities of Deep Learning models for accurately classifying mulberries based on the stages of their ripeness. This would be a crucial factor in their utilization and linking directly to their nutritional and functional qualities.

The rationale behind this direction lies in the critical need to enhance postharvest management practices, where automation through Deep Learning can potentially eliminate risks like microbial contamination and human error. By extending the

application of Deep Learning techniques to the analysis of mulberries, this research aims to develop a more holistic approach that seamlessly integrates advanced technology with specific food technology requirements. Focusing on computer vision-based automation in the context of berry image processing, this endeavor strives to advance the frontiers of what can be accomplished in agricultural automation using computer vision. The goal is addressing practical agricultural needs and contributing novel insights to the field of image-based fruit analysis. To guide this exploration, we present our second set of research questions as follows:

RQ2: Can Deep Learning techniques effectively determine the ripeness stage of mulberries by analyzing single-instance images of these fruits?

SubQ2: Assuming successful ripeness stage classification, what would be the impact on the efficiency and effectiveness of post-harvest processing in mulberries?

The result of these research questions is published in (Ashtiani et al., 2021). This evolves as the transition mulberries from unripe to fully ripe states, impacting potential health benefits. Ripeness affects the taste and texture. In addition, the concentration of phytochemicals known for their antioxidative properties. These phytochemicals are of significant interest in preventive medicine and nutrition. Moreover, certain ripening stages may alter the presence of compounds beneficial for cardiovascular health and diabetes management (Martins et al., 2023). Current methods, reliant on human assessment or chemical analysis, face limitations in accuracy and cost.

The variation in ripening patterns of mulberries and its short shelf-life further complicated matters. Here, the utilization of CNN may offer a promising solution, as its classification enables the detection of ripeness with a high accuracy in a non-destructive manner. These technological advances harness the potential to optimize the use of these nutritionally rich fruits and enhance their application across industries, mitigating wastage while ensuring quality. Integrating this classification approach enables better leveraging of the medicinal properties of mulberries, particularly in fields like nutraceuticals and functional foods, where precision in ripeness correlates with optimal health benefits.

The demonstrated efficiency, accuracy, and adaptability of Deep Learning techniques in postharvest agricultural technology highlight their potential for wider

application. This encouraging trend has led us to further investigate their use in biology, particularly for the complex task of image segmentation. Focusing on Deep Learning for segmentation, this case study aims to demonstrate the versatility of CNNs for biological structures, contributing valuable insights for biomedical research. Thus, our third research question (RQ3) is raised in a case study focusing on processing images of zebrafish to understand its biology:

RQ3: How effective and swift can Deep Learning be in segmenting images of zebrafish larvae obtained from microscope setups, particularly in distinguishing larvae from their surrounding environment?

SubQ3: If such segmentation is achieved effectively, what are the potential impacts on High-Throughput Screening (HTS) pipelines and multi-dimensional imaging processes in research settings?

These inquiries are pivotal, as precise segmentation directly influences the comprehensive understanding of zebrafish biology and fosters future advancements in the field. The result of this study is published in (Javanmardi et al., 2023).

By transitioning to the domain of biological model systems, we aim to illustrate the remarkable versatility of CNNs and their potential to contribute valuable insights across a wide array of scientific inquiries. This transition reflects the adaptability of Deep Learning and opens new horizons for interdisciplinary collaboration and research innovation.

Zebrafish stands as a cornerstone in biomedical research, attributed to their significant genomic parallels with humans and their versatility in elucidating diverse biological processes, from toxicology to drug targeting. A notable shift from 2D to intricate 3D image analysis is underway to unlock deeper biological insights, particularly in terms of accurate shape measurements. The advent of advanced imaging modalities has heightened the need for precise 3D reconstruction from these images. Central to this is the segmentation process, which dictates the integrity of the subsequent 3D reconstruction.

In our subsequent research questions, we plan to explore the integration of Deep Learning with NLP for synthesizing information. This approach is anticipated to offer a sophisticated method for enhancing our understanding and application of these technologies. Hence, our fourth research questions arise:

RQ4: Can the integration of Deep Learning (focused on image analysis) and Natural Language Processing lead to significant improvements in image captioning techniques?

SubQ4: If so, what are the expected advancements in terms of performance and accuracy that could be realized across various domains, such as medical imaging, surveillance, and digital media?

The results of these research questions are presented in these papers (Javanmardi et al., 2022), (Javanmardi et al., 2023). The motivation behind these research questions stem from the growing need to advance image captioning through the integration of Deep Learning techniques and Natural Language Processing.

Image captioning has emerged as a critical task bridging the domains of computer vision and NLP, allowing machines to generate descriptive textual explanations for images. However, the existing methods often struggle to capture the nuanced relationships between visual content and natural language expressions. Combining Deep Learning and NLP is an opportunity to create more robust and contextually accurate image captions. This integration holds the promise of elevating performance and accuracy across diverse domains, from healthcare and autonomous driving to e-commerce and content creation. Successful implementation will yield improved automated image understanding and enhanced accessibility, user engagement, and information retrieval, revolutionizing the way we interact with and interpret visual content across a wide range of applications.

In essence, our research aims to evaluate the efficacy of Deep Learning for diverse image analysis tasks. We specifically focus on analyzing individual pixel content within images, particularly identifying the foreground of interest against the background. This differentiation can be challenging in real-world settings, as demonstrated by our case studies. However, by leveraging the power of DL, we propose an effective solution to this hurdle.

Our methodology commences with single-instance images containing a well-defined object. This approach proves particularly advantageous in applications where precise measurements are crucial. By scrutinizing both the single-instance images and the capture process, we can achieve an accurate description of the target object. Importantly, our analysis primarily concentrates on the object density distribution within these images, as the background holds minimal relevance to our study. While single-instance images might have some irregularities, they are

generally easier to interpret. However, describing the contents of an image requires a different approach. This challenge opens a fascinating opportunity to explore how DL can aid in developing efficient image captioning solutions. In this endeavor, we integrate DL with Natural Language Processing, leveraging the capability of AI to emulate human reasoning in understanding the conceptual content of images

1.3 Thesis Overview

The content of this thesis represents a systematic application of Deep Learning across diverse domains, including agriculture, and biology, and extends into the complex intersection of image processing and Natural Language Processing. The organization of the thesis is designed to facilitate a comprehensive grasp of these varied applications, delineated into four research-focused chapters that correspond to our central research questions.

Chapter 2 sets the stage by tackling research question 1 (RQ1) through a case study in seed quality control. Here, we explore the potential of Deep Learning to revolutionize corn seed classification (Javanmardi et al., 2021). This foundational chapter delves into identifying the most effective features for classifying seed varieties based on images. Our research introduces a groundbreaking approach by leveraging Deep Learning and Machine Learning techniques as feature extractors. This innovative method demonstrates robustness across varying seed orientations, highlighting the transformative power of Deep Learning in the post-harvest agricultural domain.

Chapter 3 dives into the realm of classification using Deep Learning for post-harvest agricultural processing, focusing on berries. Here, we present a groundbreaking application of CNNs to classify mulberries according to their ripeness. This marks a significant advancement in the field, as we propose the first state-of-the-art, efficient methodologies within the context of Deep Learning for mulberry sorting. This research paves the way for the development of intelligent mulberry sorting systems, as evidenced by our published work (Ashtiani et al., 2021).

Chapter 4 dives into the biological domain with RQ3, exploring the segmentation application of Deep Learning in microscopic image processing. Utilizing a captivating case study on zebrafish larvae segmentation and classification, we investigate the effectiveness of CNNs in tackling this intricate task (Javanmardi et al., 2023).

This research opens exciting new avenues for integrating Deep Learning into medical research, potentially unlocking insights that could spark further technological advancements.

Chapter 5 pushes the boundaries by exploring the integration of Deep Learning and NLP in RQ4. Here, we delve into the revolutionary concept of fusing these two powerful fields to elevate image captioning. Building upon the Deep Learning foundations established in prior chapters, we analyze how this synergy unlocks new levels of performance and precision in language generation. This ultimately offers a holistic perspective on the transformative potential arising from this potent combination, as explored in our publications (Javanmardi et al., 2022) and (Javanmardi et al., 2023) The thesis culminates in Chapter 6, where we encapsulate our findings and contemplate the broader ramifications of our research. We summarize the pivotal discoveries and reflect on their implications while also casting light on prospective research trajectories that build upon the innovative methodologies and successful implementations documented in our study.

1.4 Contribution of this Thesis

- Javanmardi, S., S.-H. M. Ashtiani, F. J. Verbeek, and A. Martynenko (2021).Computer-vision classification of corn seed varieties using deep convolutional neural network. Journal of Stored Products Research 92, 101800.
- Ashtiani, S.-H. M., S. Javanmardi, M. Jahanbanifard, A. Martynenko, and F. J. Verbeek (2021). Detection of mulberry ripeness stages using deep learning models. IEEE Access 9, 100380–100394.
- Javanmardi, S., X. Tang, M. Jahanbanifard, and F. J. Verbeek (2023). Unsupervised Segmentation of High-Throughput Zebrafish Images Using Deep Neural Networks and Transformers. In International Conference on Data Science and Artificial Intelligence, pp. 213–227. Springer.
- Javanmardi, S., A. M. Latif, M. T. Sadeghi, M. Jahanbanifard, M. Bonsangue, and F. J. Verbeek (2022). Caps captioning: a modern image captioning approach based on improved capsule network. Sensors 22 (21), 8376.
- Javanmardi, S, M. Jahanbanifard, M. Bonsangue and F. J. Verbeek (2023). Using a Novel Capsule Network for an Innovative Approach to Image Cap-

tioning, The Third AAAI Workshop on Scientific Document Understanding, CEUR.