




# Hochschule Darmstadt

## Fachbereiche Mathematik und Naturwissenschaften & Informatik


# Fine-tuning of Gaussian Denoising CNNs for Image Reconstruction in Fluorescence Microscopy

Abschlussarbeit zur Erlangung des akademischen Grades
Master of Science (M.Sc.)
im Studiengang Data Science


vorgelegt von

**Leopold Groznov**

Matrikelnummer: 762049


| | | |
|---|---|---|
| Referent(in) | : | Prof. Dr. Thomas März |
| Korreferent(in) | : | Prof. Dr. Elke Hergenröther |

| | | |
|---|---|---|
| Ausgabedatum | : | 12.07.2024 |
| Abgabedatum | : | 11.01.2025 |

# DECLARATION

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

*Darmstadt, 11. Januar 2025*

Leopold Groznov

# ABSTRACT

Fluorescent microscopy is an established tool in life sciences, enabling the visualization of biological structures and processes. This imaging technique works by illuminating the specimen and thus exciting fluorophores located within the specimen with light of specific wavelengths, causing them to emit light that can be detected and used to generate detailed images. However, it requires balancing key imaging parameters, including speed, spatial resolution, light exposure, and imaging depth. Major challenges in fluorescence microscopy are photobleaching and phototoxicity, which necessitate minimizing photon exposure by reducing the duration or intensity of light. This, in turn, degrades image quality, highlighting the importance of computational methods for improving image reconstruction.

This thesis investigates the use of pre-trained Gaussian denoising networks for fluorescence microscopy image reconstruction. Originally trained on large natural image datasets, these networks are fine-tuned on fluorescence microscopy images to evaluate their suitability for this specialized task. Given the limited size of microscopy datasets, this process can be framed as few-shot learning. The work highlights appropriate denoising networks based on diverse architectures and selects suitable fluorescence microscopy datasets for training and evaluation.

The findings demonstrate that pre-trained Gaussian denoisers can effectively reconstruct fluorescence microscopy images, achieving performance comparable to or exceeding state-of-the-art methods. Additionally, this thesis identifies key hyperparameters that significantly influence the fine-tuning process, providing insights to maximize network performance. The approach is straightforward to implement and can be readily adapted for new denoiser networks as they emerge, leveraging the ongoing advancements in this active field of research. Furthermore, by utilizing pre-trained networks, the technique reduces the computational costs and time associated with training, since the pre-training is already performed in different previous studies. These contributions aim to advance the field of computational imaging and provide a basis for future developments that could contribute to practical improvements in fluorescence microscopy techniques.

# ZUSAMMENFASSUNG

Fluoreszenzmikroskopie ist ein etabliertes Werkzeug in den Lebenswissenschaften und ermöglicht die Visualisierung biologischer Strukturen und Prozesse. Diese Bildgebungstechnik basiert auf der Anregung von Fluorophoren im Präparat durch Licht bestimmter Wellenlängen. Die Fluorophore emittieren dabei Licht, das detektiert und zur Erstellung detaillierter Bilder genutzt werden kann. Dabei müssen jedoch wichtige Parameter wie Geschwindigkeit, räumliche Auflösung, Lichtintensität und Eindringtiefe sorgfältig ausbalanciert werden. Zu den zentralen Herausforderungen der Fluoreszenzmikroskopie gehören Photobleichen und Phototoxizität, die eine Minimierung der Photonendosis durch Verringerung der Belichtungsdauer oder Lichtintensität erforderlich machen. Dies beeinträchtigt die Bildqualität und verdeutlicht die Bedeutung rechnergestützter Methoden zur Verbesserung der Bildrekonstruktion.

In dieser Masterarbeit wird der Einsatz vortrainierter Gaußscher Denoising-Netzwerke für die Rekonstruktion von Fluoreszenzmikroskopie-Bildern untersucht. Ursprünglich auf großen Datensätzen natürlicher Bilder trainiert, wird durch Fine-Tuning auf Fluoreszenzmikroskopie-Bildern ihre Eignung für diese spezialisierte Aufgabe untersucht. Aufgrund der begrenzten Größe der benutzten Mikroskopiedatensätze kann dieser Ansatz als Few-Shot-Learning betrachtet werden. Geeignete Denoising-Netzwerke mit unterschiedlichen Architekturen werden vorgestellt, und passende Datensätze aus der Fluoreszenzmikroskopie werden für Training und Evaluierung ausgewählt.

Es wird gezeigt, dass vortrainierte Gaußsche Denoising-Netzwerke Fluoreszenzmikroskopie-Bilder effektiv rekonstruieren können und dabei eine Leistung erzielen, die mit dem aktuellen Stand der Technik vergleichbar ist oder diesen übertrifft. Zudem werden wichtige Hyperparameter identifiziert, die den Feinabstimmungsprozess wesentlich beeinflussen, und es werden wertvolle Erkenntnisse zur Optimierung der Netzwerkleistung gewonnen. Der Ansatz lässt sich einfach umsetzen und an neue Denoising-Netzwerke anpassen, die aus den kontinuierlichen Fortschritten in diesem aktiven Forschungsfeld hervorgehen. Durch den Einsatz vortrainierter Netzwerke werden außerdem die mit dem Training verbundenen Rechenkosten und Zeitaufwände reduziert, da das Vortraining bereits in anderen Studien durchgeführt wurde.

# ACKNOWLEDGMENTS

I sincerely thank Tim Selig for his co-supervision and tireless efforts in guiding me through this master thesis. From proposing the topic and helping with technical challenges to providing regular consultation and refining the text, his generous investment of time and expertise has been invaluable.

# CONTENTS

# LIST OF FIGURES

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| BM3D | Block-Matching and 3D Filtering |
| CARE | Content-Aware Image Restoration |
| CNN | Convolutional Neural Network |
| CPU | Central Processing Unit |
| CT | Computed Tomography |
| GAN | Generative Adversarial Network |
| GDDR6 | Graphics Double Data Rate 6 |
| GPU | Graphics Processing Unit |
| GT | Ground Truth |
| KBA | Kernel Bases Attention |
| LSCM | Laser Scanning Confocal Microscopy |
| MAE | Mean Absolute Error |
| MB | Megabyte |
| MFF | Multi-axes Feature Fusion |
| ms | Millisecond |
| MSE | Mean Squared Error |
| N2V | Noise2Void |
| NA | Numerical Aperture |
| NLP | Natural Language Processing |
| PSF | Point Spread Function |
| PSNR | Peak Signal-to-Noise Ratio |
| ReLU | Rectified Linear Unit |
| SNR | Signal-to-Noise Ratio |
| SSIM | Structural Similarity Index Measure |

# Part I

# THESIS

# 1 INTRODUCTION

## 1.1 MOTIVATION AND RESEARCH QUESTIONS

Fluorescence microscopy is a cornerstone tool in biological sciences, providing visualization and a means for the investigation of the spatio-temporal dynamics within cells and tissues [50]. Unlike traditional light microscopy, which relies on variations in transmitted or reflected light, fluorescence microscopy can isolate and visualize specific components within complex samples, even in living cells. By leveraging the ability of specific fluorophores to emit light under excitation, this form of light microscopy allows high-resolution imaging of biological processes at both cellular and subcellular levels. However, the quality and scope of fluorescence microscopy are inherently constrained by several factors, including the physical properties of the optical system, the photophysical behavior of fluorophores, and the tolerance of biological samples to photon exposure. These limitations necessitate trade-offs between critical imaging parameters, such as spatial resolution, imaging speed, light exposure, and imaging depth, making it challenging to obtain meaningful results without compromising the sample's natural structure or viability [10]. A "maximal photon budget", i.e. the dose of light to which the sample is exposed to, can be seen as a constraint for such a trade-off, which is illustrated in Fig. 1.1.

Two major challenges in fluorescence microscopy are photobleaching and phototoxicity. Photobleaching arises when fluorescent probes degrade under prolonged excitation, leading to a loss of fluorescence emission intensity after exposure [10, 20, 45]. Phototoxicity stems from the generation of reactive oxygen species during light exposure, which can damage or kill living cells [15, 20]. These issues are particularly pronounced in long-term or high-intensity imaging, where they significantly compromise both image quality and sample viability. Additionally, background noise from nonspecific staining or autofluores-

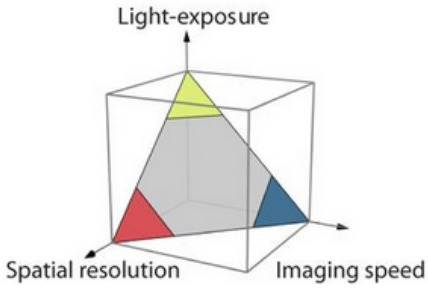


Figure 1.1: A trade-off between parameters of fluorescence microscopy (imaging speed, spatial resolution and light exposure), constrained by the maximal number of photons used. Only three parameters are shown as dimensions, the rest of them is omitted for the clarity. The illustration is from [50].

cence in the sample can reduce image clarity [33]. Addressing these challenges often involves reducing excitation light intensity or exposure times, but these adjustments can result in images with low signal-to-noise ratios (SNR), posing additional challenges for data interpretation [10].

To mitigate these limitations, a range of physical system modifications and software-based strategies — encompassing both conventional and deep-learning-based ones — have been developed. The approaches involving modification of the physical setup include pulsed excitation [45], controlled light exposure microscopy [13] and the use of specialized culture media to minimize photodamage [6]. On the other hand, the software-based strategies include computational procedures, such as deconvolution [43] and classical denoising techniques, e.g. block matching and 3D filtering (BM3D) [9]. Recent advances in machine learning have introduced convolutional neural networks (CNNs) for image denoising and reconstruction. Some of these methods have already been evaluated on fluorescence microscopy image data and have demonstrated the ability to improve low-SNR images acquired with shorter exposure times [10]. However, training such models from scratch can be challenging in practical applications, particularly due to the limited availability of fluorescence microscopy training data [10] and the computational effort involved.

To address this, pre-trained CNN-models can be utilized as an intermediate step. A pre-trained model, initially trained on a large dataset of generic image pairs, consisting of a clean image and a cor-

responding image corrupted by additive Gaussian noise, can then be fine-tuned using the available fluorescence microscopy image pairs. It should be noted that the pre-training step utilizes large datasets consisting of regular photographs, which are significantly easier to obtain compared to the relatively scarce microscopy images. Furthermore, this approach leverages the benefits of prior training, since the computational load for model pre-training has already been incurred in a different study and is therefore amortized. Fine-tuning such pre-trained models requires considerably fewer resources while still offering the potential to improve image quality in low-SNR conditions.

In this thesis, the focus is restricted to a specific class of models: Gaussian denoising networks. These models are well-studied and their fine-tuning has demonstrated competitive performance in reconstructing computed tomography (CT) image data [39]. Additionally, Gaussian denoisers provide a strong foundation for further improvements, as they constitute an active field of research [32]. Future advancements in this class of models can be seamlessly integrated into the proposed framework, offering potential performance enhancements without requiring substantial modifications.

Based on the considerations discussed above, this thesis aims to investigate the potential of pre-trained Gaussian denoiser networks for fluorescence microscopy image reconstruction. The following research questions are formulated to guide this investigation:

RQ1. *Can the use of pre-trained Gaussian denoisers be beneficial for fluorescence microscopy image reconstruction?*
This question explores whether pre-trained Gaussian denoiser models, fine-tuned on fluorescence microscopy data, can achieve competitive or better performance in restoring the quality of low-SNR fluorescence microscopy images compared to existing approaches.

RQ2. *How do different Gaussian denoising network architectures perform when fine-tuned for fluorescence microscopy image reconstruction?*
By comparing various Gaussian denoiser architectures, this question seeks to identify structural or design differences that influence performance in the context of fluorescence microscopy data.

RQ3. *Which optimization techniques have the most significant impact during the fine-tuning process of Gaussian denoisers?*
This question focuses on exploring such optimization approaches as hyperparameter tuning and data augmentation strategies, to identify the factors that play a critical role in improving model performance and enhancing the training process.

Addressing these research questions can provide insights into the practical applicability of pre-trained Gaussian denoisers for image processing in fluorescence microscopy and highlight strategies for maximizing their performance.

## 1.2 OVERVIEW OF THE STRUCTURE

To address the research questions formulated above, this thesis is organized into distinct yet interconnected chapters. The Chapter 2 *Theoretical Background and Related Work* introduces the fundamental concepts necessary to contextualize the approach proposed in this thesis. It begins with an explanation of fluorescence microscopy imaging process, highlighting the key factors contributing to the image degradation and the models used to represent it. This is followed by a concise overview of image restoration approaches developed or applied for fluorescence microscopy data, encompassing both conventional methods and those based on deep learning. Next, Gaussian denoising networks, which form the core components of the proposed approach, are introduced, along with a rationale for selecting three state-of-the-art architectures. The chapter then highlights the fine-tuning strategy employed in this work, providing the necessary background for its selection. It concludes by discussing data augmentation techniques, including a brief overview of the options considered and justifying the selection of a single approach, which is used in certain experiments to enable comparisons with training on unaugmented data.

The next part of the thesis — Chapter 3 *Method and Experimental Setups* — outlines the experimental methodology and setups aimed at investigating the research questions. It begins with a description of the datasets selected for fine-tuning the networks, followed by details of the training and testing procedures, including clarification of the

evaluation metrics used to compare model performance. This chapter serves to provide a clear understanding of the experimental process and to establish a foundation for the reproducibility of the results.

The outcomes of the experiments are presented and discussed in Chapter 4 *Evaluation of Results*. The analysis includes comparison with the baseline method and investigates whether the optimization techniques explored had a meaningful influence on fine-tuning performance.

The thesis concludes by summarizing the key findings, offering direct answers to the research questions. The final chapter also explores potential directions for future research, highlighting the opportunities to expand upon the proposed approach.

# 2

# THEORETICAL BACKGROUND AND RELATED WORK

## 2.1 THEORETICAL FRAMING OF FLUORESCENCE MICROSCOPY IMAGE RESTORATION

To address the origins of distortions in fluorescence microscopy images, it is necessary to first outline key aspects of this imaging technique.

### 2.1.1 *Basics of Imaging Process*

In fluorescence microscopy, an image is acquired by detecting the light emitted from the fluorescent molecules known as fluorophores located within the specimen. After preparation of the specimen, the following four main stages can be distinguished in the process of the image acquisition [4]:

1. *Fluorophore excitation.* A fluorophore must absorb a photon to be raised to an excited state. To achieve this, the specimen is illuminated with light of a specific wavelength suitable for transitioning the fluorophore from its ground state to an excited state.

2. *Photon emission.* The fluorophore molecule returns to its ground state by emitting a photon. This emitted photon has lower energy and a longer wavelength compared to the absorbed photon.

3. *Photon detection.* Some of the emitted photons reach the objective lens, which focuses them onto a detector. Since the photons are emitted in different directions, only a small fraction is ultimately captured by the detector.

4. *Quantification and storage.* After hitting the detector, an incoming photon triggers the release of an electron. The accumulated charges from these electrons are measured at certain time intervals, with a higher number of photons resulting in a greater charge and, consequently, a higher pixel value.

By tagging biological molecules with fluorescent dyes or proteins, it is possible to selectively highlight cellular components, enabling high-contrast imaging of intricate structures. A simplified diagram showing the paths of excitation and emitted light can be found in Fig. 2.1.

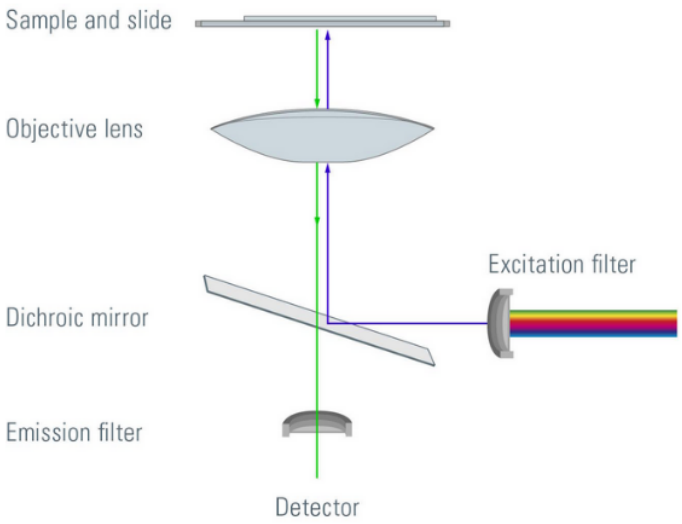


Figure 2.1: Excitation (violet) and emitted (green) light in a fluorescence microscope. Dichroic mirror is used to reflect the light of the excitation wavelength, directing it to the specimen, and to filter the light traveling to the detector [30]. The illustration is from [30].

To capture a three-dimensional image of a biological sample, a fluorescence microscope can sequentially acquire two-dimensional optical slices, plane by plane [43]. This involves continuously shifting the focal plane along the $z$-axis, also referred to as the optical axis. The process requires precise movement of either the objective lens or the specimen itself [43]. For optimal imaging, it is essential to align the detector as closely as possible to the focal point (see Fig. 2.2); otherwise, the resulting image will suffer from increased blurring and reduced intensity.

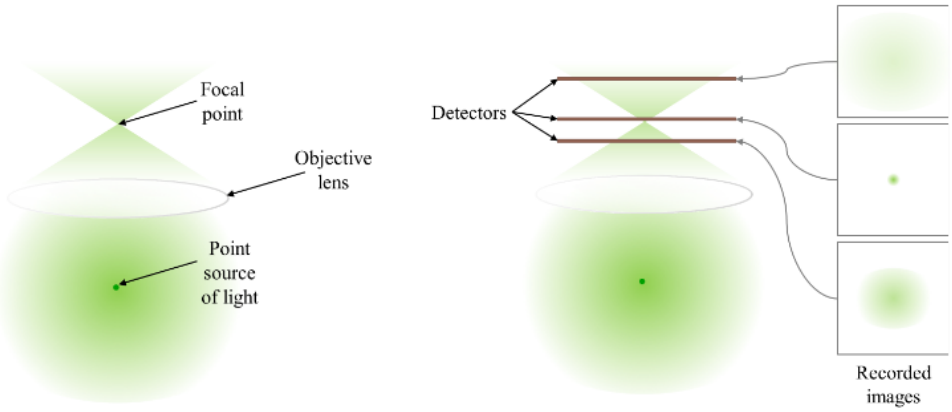


Figure 2.2: Imaging of a light-emitting point with a widefield microscope (simplified diagram). The most accurate recording requires the detector to be placed into the focal point. In this case, the light-emitting point in the specimen is referred to as located in the in-focus plane. If the light-emitting point is placed in an out-of-focus plane, the recorded object is characterized by increased size and decreased intensity [4]. The illustration is from [4].

### 2.1.2 *Degradation with Blur: Point Spread Function*

Even under ideal conditions, where a light-emitting point is perfectly aligned with the in-focus plane, the fluorescent point is not detected as an infinitesimally small volume. Instead, it appears as a larger volume defined by the Point Spread Function (PSF), which depends on the emission wavelength and the properties of the objective lens [4].

The PSF viewed from the side, i.e. in the $xz$ or $yz$ plane, typically has an hourglass-like shape, while in the $xy$ (lateral) plane it appears as a bright central spot, provided that the light-emitting point is located in the in-focus plane. If the light-emitting point shifts to an out-of-focus plane, a characteristic effect occurs due to the interference of light waves. This effect, which produces concentric rings of decreasing intensity, is known as the Airy pattern, named after the mathematician George Biddell Airy (1801–1892) (see Fig. 2.3 a,b) [4]. The intensity of the Airy pattern's central disk, excluding the concentric rings, can be approximated by a Gaussian function (Fig. 2.3 c). Consequently, the blur produced by a light-emitting point in an in-focus plane can be modeled as the application of a Gaussian filter to the undistorted image [4].

Although the PSF continues indefinitely, with low values at large distances from its center, the PSF size can be characterized by the radius of the Airy disk, defined as the distance from its center to the first minimum, just before the first concentric ring begins (Fig. 2.3 c). This radius, $r_{airy}$, is calculated using the formula [4, 43]:

$$r_{airy} = \frac{0.61\lambda}{\eta \sin\theta} = \frac{0.61\lambda}{NA} \tag{2.1}$$

where $\lambda$ is the emission wavelength, $\eta$ is the refractive index of the immersion medium between the sample and the objective (approximately 1.0 for air, 1.34 for water and 1.5 for oil) and $\theta$ is the half cone angle captured by the objective lens (see Fig. 2.4). The product $\eta \sin\theta$ is referred to as the numerical aperture (NA) of the objective. A higher NA is generally preferable, since it can decrease PSF size and therefore reduce blurriness.

The formula (2.1) provides the smallest theoretical Airy disk size under ideal conditions, i.e. aberration-free optics and an in-focus point. In practice, the PSF can be larger in size and more sophisticated in form due to imperfections such as inaccuracies in the objective lens or mismatches between the refractive indices of the immersion medium and the specimen's surrounding medium [4].

Moreover, the PSF size dictates the lateral resolution limit of the optical system: according to the Rayleigh criterion, two equally bright points can only be resolved as separate if their distance in the $xy$ plane equals or exceeds $r_{airy}$. Consequently, a pixel size smaller than $r_{airy}$ does not improve imaging results [4, 43].

Thus, fluorescence microscopy involves mapping of each light- emitting point to its PSF, instead of representing it just as a single pixel of measured intensity. Overlapping PSFs from adjacent fluorescent points result in the addition of intensities, contributing to image blur [4].

In the fluorescence microscopy imaging model, PSFs can be seen as shift-invariant, so that it affects any single point the same way, regardless of its position. The measured data $g(x, y, z)$ can therefore

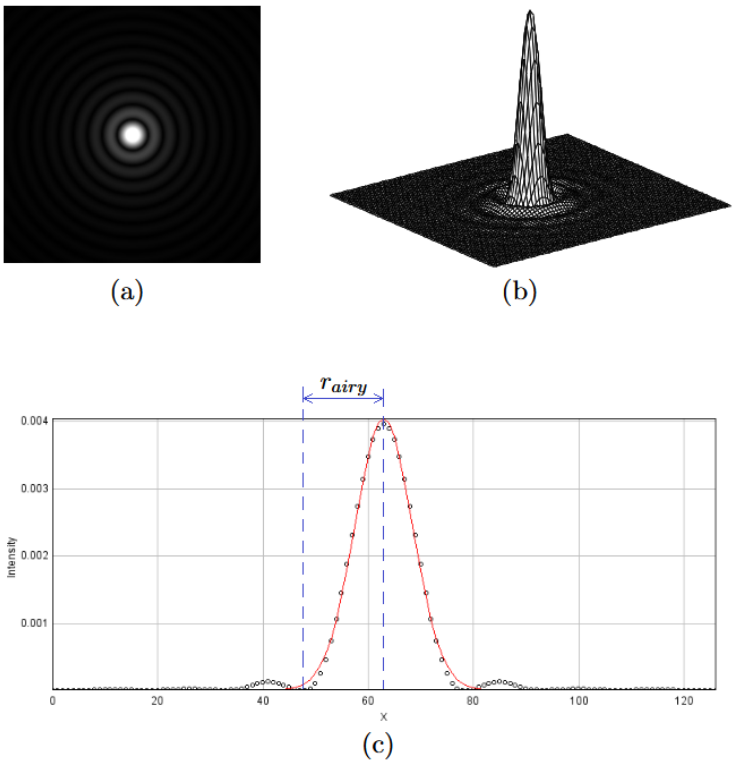


Figure 2.3: An Airy pattern (a) in a lateral plane and (b) as a surface plot. The surface plot demonstrates that the intensity is significantly higher in the central spot compared to the concentric rings. (c) Profile through the center of an Airy disk (black circles) with its Gaussian approximation (red) [4]. $r_{airy}$ is the distance between the center of the Airy disk and the first minimum. The diagrams are from [4].

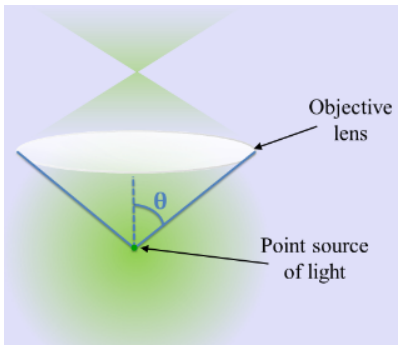


Figure 2.4: Angle $\theta$ is the half-angle of the cone reaching the objective lens. Since $\sin\theta$ cannot be greater than 1, the NA of the objective, which is the product of $\sin\theta$ and the refractive index $\eta$, cannot exceed $\eta$ [4]. It shows that lens with a higher NA can be employed for blur reduction only to a limited extent. The illustration is from [4].

be represented as the result of the convolution operation ($\otimes$) between the undistorted image $f(x, y, z)$ and the PSF $h(x, y, z)$ [28, 43]:

$$g(x, y, z) = f(x, y, z) \otimes h(x, y, z), \qquad x, y, z \in \mathbb{R}. \tag{2.2}$$

### 2.1.3 *Degradation with Noise: Primary Sources*

Along with the distortion from PSF-induced blur discussed above, fluorescence microscopy images are also affected by noise, which often appears as grainy distortions in the recorded image. Two primary types of noise are particularly relevant to fluorescence microscopy:

- *Photon noise* originates from the statistical nature of photon detection. Due to the quantum nature of light, the number of detected photons follows a Poisson distribution [4, 43]. Poisson noise is signal-dependent, with its standard deviation being determined as square root of the expected number of photons recorded per pixel [33]. Consequently, higher signal levels result in an increased standard deviation. However, they also improve the signal-to-noise ratio (SNR), which is calculated as:

  $$SNR = \frac{Signal}{SD_{noise}} = \frac{\lambda}{\sqrt{\lambda}} = \sqrt{\lambda},$$

  with $\lambda$ representing the expected number of photons detected per pixel [4, 43].

- *Read noise*, also known as electronic noise [43], arises during the quantification stage of the recording process, specifically due to imprecision in the detector's readout [4, 43]. Unlike photon noise, read noise is signal-independent and typically follows a Gaussian distribution with a mean of zero [4]. Its standard deviation is constant and determined by the precision of the detector, making read noise an intrinsic characteristic of the imaging system. The amount of the quantified photons does not affect its

standard deviation, but increasing the number of photons results in higher SNR [4].

The simulation in Fig. 2.5 demonstrates that higher signal levels improve the detectability of structures despite the presence of both photon noise and read noise. It has to be noted that photon noise typically has a greater impact on fluorescence microscopy images, since modern detectors are engineered to minimize read noise, rendering its influence comparatively negligible [43].

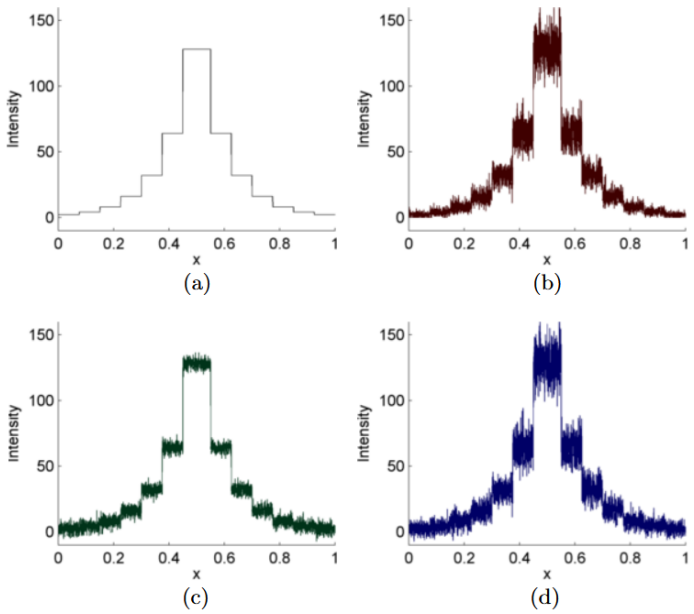


Figure 2.5: Simulated (a) noise-free signal, with each of the adjacent value doubled or halved, (b) with photon noise added, (c) with read noise added, (d) with photon and read noise added [4]. Low signal levels result in bad detectability of the doubling of a signal, with both types of noise. In this simulation, both types of noise are set to approximately the same levels for simplicity. However, it is important to note that modern optical systems significantly reduce read noise [4], thus the photon noise typically has a more pronounced impact. The diagram is from [4].

### 2.1.4 *Fluorescence Microscopy Techniques*

Various types of microscopes are employed for fluorescence imaging, each with distinct operational principles and performance characteristics. Among the numerous available techniques, widefield and confocal microscopy are the most commonly used in practical applications

[38]. Although advanced methods exist that deliver better imaging performance [38], this thesis focuses exclusively on data obtained using widefield and confocal microscopes. These two techniques will therefore be briefly discussed, highlighting their respective advantages and limitations:

- *Widefield microscopy*. In this method, the entire specimen is illuminated simultaneously. This approach lacks optical sectioning capability, meaning that photons from out-of-focus planes contribute to the recorded image alongside those from the in-focus plane [4]. This can result in significant image blur, particularly when imaging thick specimens, where out-of-focus fluorescence obscures fine details [4]. Consequently, detecting small structures in thick samples poses a substantial challenge, limiting the utility of widefield microscopy for high-resolution studies of complex three-dimensional specimens.

- *Confocal microscopy* addresses the limitations of widefield microscopy by employing a pinhole in front of the detector to block out-of-focus light [43]. This enables the system to detect fluorescence from a specific point in the sample at any given time, resulting in improved resolution along the $z$-axis [4, 43]. Typically, confocal microscopes illuminate a small volume of the specimen using a focused laser beam, a technique known as laser scanning confocal microscopy (LSCM) [4]. This method significantly reduces image blur and provides enhanced optical sectioning, making it suitable for imaging thick samples. However, the trade-offs include higher noise in terms of a lower SNR [4, 38], as the pinhole blocks a considerable amount of photons, and increased acquisition time [4, 16], as the system must scan each point sequentially.

### 2.1.5 *Degradation Model*

To take into account the image formation process discussed above, the degradation model should represent at least two distortion elements — blur and noise (see Fig. 2.6 for visual simulation). To obtain the

recorded image $g(x,y,z)$, the undistorted image $f(x,y,z)$ is first being blurred with the PSF $h(x,y,z)$ of the microscope and subsequently affected by noise degradation $N$ [43]:

$$g(x,y,z) = N(f(x,y,z) \otimes h(x,y,z)). \tag{2.3}$$

If all the noise types including the Poisson noise undergo a Gaussian approximation, the model can be simplified to [38]:

$$g(x,y,z) = f(x,y,z) \otimes h(x,y,z) + w(x,y,z), \tag{2.4}$$

where $w(x,y,z)$ is an additive Gaussian noise. However, it has to be mentioned that under low SNR the Gaussian model provides a poor approximation [38]. For a detailed description of a model with Poisson noise distribution, refer to [38].

## 2.2 EXISTING APPROACHES TO FLUORESCENCE MICROSCOPY IMAGES RESTORATION

This section provides an overview of the existing computational approaches used for fluorescence microscopy image restoration. Subsection 2.2.1 discusses techniques that do not involve deep learning, while Subsection 2.2.2 focuses on methods based on deep learning.

### 2.2.1 *Conventional Approaches*

Deconvolution is a well-established approach to enhance fluorescence microscopy images. This term refers to a broad spectrum of algorithms which aim to solve the inverse problem of recovering $f(x,y,z)$ from the equation 2.2, thus obtaining the distribution of intensities in the undistorted image. Traditional deconvolution algorithms require explicit knowledge of the PSF term [21] which acts as a convolutional kernel. The PSF can be either theoretically estimated or empirically measured by acquiring images of small fluorescent beads with the

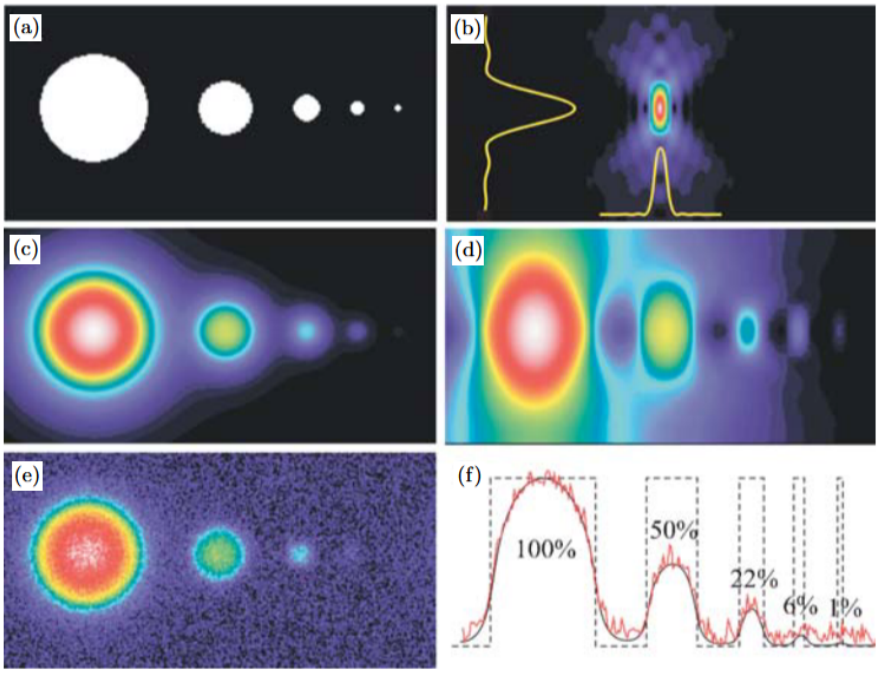


Figure 2.6: Simulation of PSF-induced blur and noise for an object consisting of five spheres in three dimensions: (a) undistorted object; (b) PSF of the microscope; (c, d) result of the blurring by convolution with the PSF, lateral and axial sections; (e) result after blurring and addition of noise; (f) diagram showing maximum intensities of the undistorted (dashed), blurred (black) and blurred and noisy (red) images. The detectability of small details diminishes with the introduction of noise [43]. The diagram is from [43].

given microscope [43]. An example of a discrepancy between a theoretically computed and measured PSF of the similar microscope can be seen in Fig. 2.7. However, it is worth noting that both ways of PSF assessment suffer from aberrations, which can be critical for the ability of deconvolution procedures to deliver meaningful results [43].

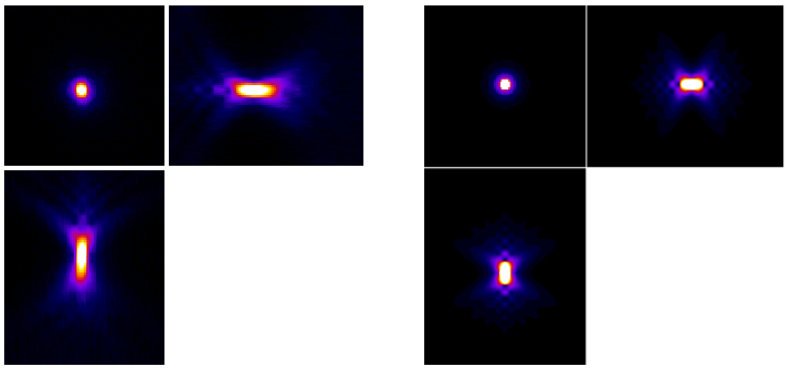


Figure 2.7: PSF of a widefield microscope, after applying linear contrast filter: (a) acquired as an image of a small fluorescent bead; (b) theoretically estimated [4]. The illustration is from [4].

There also have been developed deconvolution algorithms that operate without prior knowledge of the microscope's PSF. They are referred to as blind deconvolution algorithms and aim to estimate PSF iteratively [21, 43].

Deconvolution algorithms are used in a number of commercial software systems for microscopy image reconstruction, such as Huygens [36] and AutoQuant [8].

Another approach is the application of standard denoising methods such as BM3D [9]. It was evaluated in [10] on multiple widefield and confocal microscopy datasets and demonstrated results that, while slightly inferior, were still comparable to the deep learning-based method CARE (see Section 2.2.2 below). Also, in [29] several denoising approaches, including their sparsity-based method, were estimated, however only on confocal microscopy data.

### 2.2.2 *Deep Learning-Based Approaches*

Multiple deep-learning based solutions have been developed for the tasks related to fluorescence microscopy data, such as super-resolution

[22], isotropic restoration [49] or cross-modality transformation [46], to name a few.

A seminal contribution to the problem of fluorescence microscopy image reconstruction was the introduction of a UNet-based Content-Aware Restoration (CARE) network [50]. Tested on widefield and confocal microscope images, this method addresses image reconstruction, while also demonstrating good performance in solving problems of super-resolution, isotropic restoration and image projection of microscopy images [50]. A transfer learning system incorporating this network was introduced in [47]: the first step featuring the processing of a microscopy image with the CARE model pre-trained on a large bulk of general image data is followed by denoising with a self-supervised method Noise2Self [5].

In [10], the researchers attempt to employ an unsupervised UNet-based denoising method Noise2Void (N2V) [19] for restoring fluorescence microscopy data. A considerable advantage of this approach is that it does not require paired training images. However, it performs rather poorly in comparison with other methods evaluated in [10].

Several physics-informed methods leveraging machine learning were introduced, which combine a data-driven approach with the embedding of knowledge about optical processes in microscopy into the image reconstruction systems [17, 21, 24, 41].

## 2.3 GAUSSIAN DENOISING NETWORKS

The pre-trained Gaussian denoising networks selected for this thesis were chosen due to their state-of-the-art status, which means that they have been introduced recently and demonstrated state-of-the-art performance in denoising. Another reason was the availability of a respective pre-trained model suitable for one-channel grayscale images, a requirement dictated by the characteristics of the data used in this thesis. Furthermore, the successful employment of these three pre-trained networks for CT images in [39] allowed to make an assumptions on their applicability to the fluorescence microscopy data.

### 2.3.1 *DRUNet*

DRUNet was introduced by Zhang et al. in 2021 [56]. Notably, the objective of the paper was not to develop a standalone denoising architecture, but rather a component for a plug-and-play solution. In this context, the denoising network itself is supposed to be applied alternately with another plug-and-play component, the data fidelity term, which is dependent on the specific task [56]. However, DRUNet can be used separately for denoising purposes as well and the paper showcases its competitive results by comparing it to several state-of-the-art denoising approaches [56].

**DRUNet Architecture**

DRUNet architecture combines the elements of UNet [35], ResNet [12] and FFDNet [58], with UNet being its design basis. The elements originating from ResNet are additive residual connections. From FFDNet, a previous work by Zhang et al., introduced in 2018 [58], DRUNet adapts the idea of passing a noise level map with the input image.

Fig. 2.8 depicts a non-detailed overview of the DRUNet architecture. Three consecutive downsampling steps, which are implemented using strided convolution, scale down the image size, while the number of channels is increased. Upsampling steps are performed using transposed convolution. In contrast to the classical UNet architecture [35], additive residual connections are employed instead of concatenation-based skip connections. Each downsampling or upsampling step consists of the corresponding convolutional operation, along with four residual blocks. These blocks with additive residual connections alter neither the dimensionality of the image nor the number of channels. At each of the downsampling and upsampling steps, DRUNet has four of those blocks, which are identical within a step, being added up sequentially. Every residual block consists of a convolution operation, followed by a ReLU activation function and an additional convolutional layer, with a residual additive connection over those three operations. Fig. 2.9 presents a detailed view of an upsampling block (a) and a downsampling block (b), both of which include a schematic of a residual block.

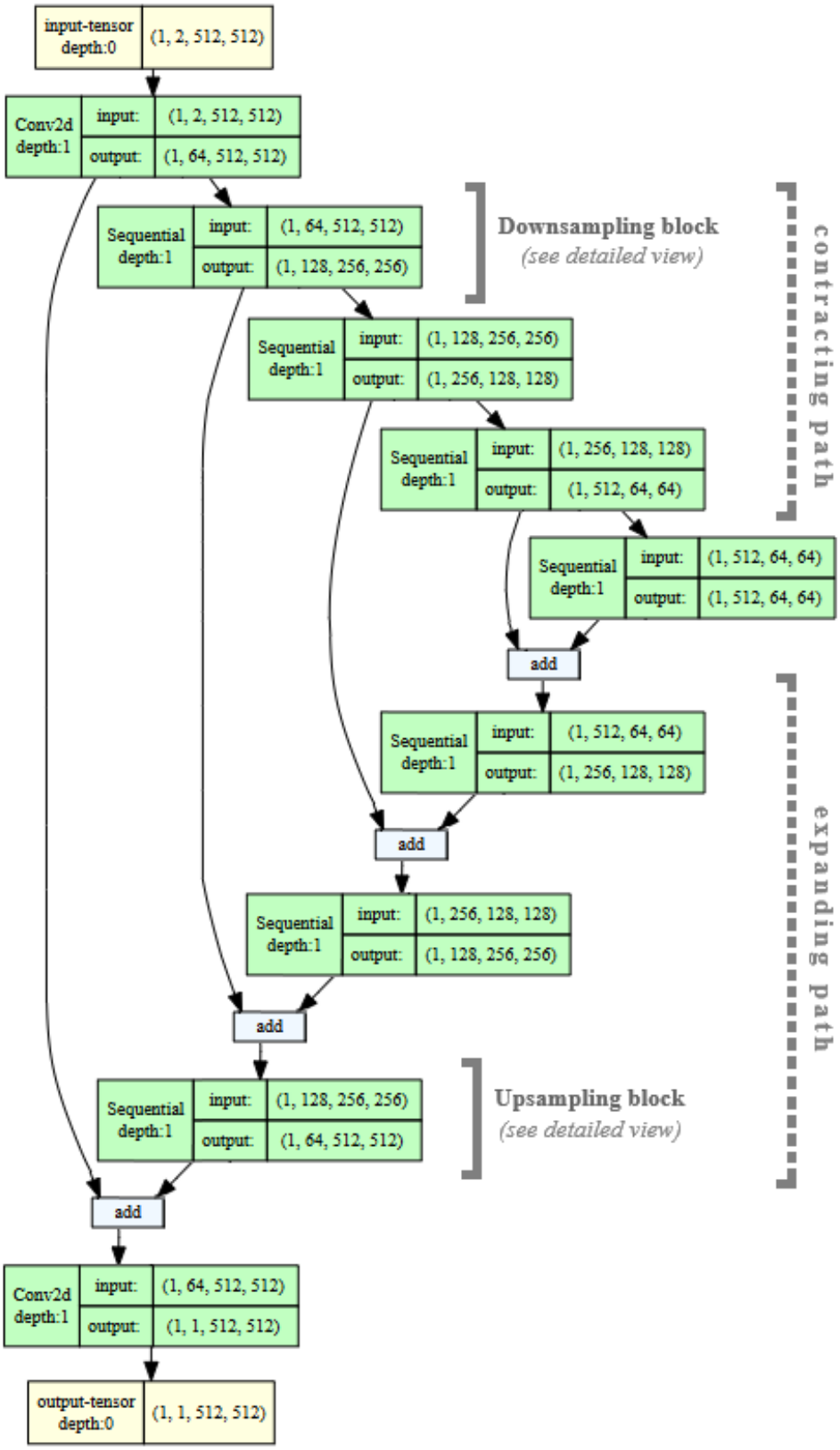


Figure 2.8: Overview of the DRUNet architecture. The graph was generated using the *torchview* library and later modified for clarity. Input image size is 512x512. Downsampling and upsampling steps are highlighted and further detailed in the following figure.

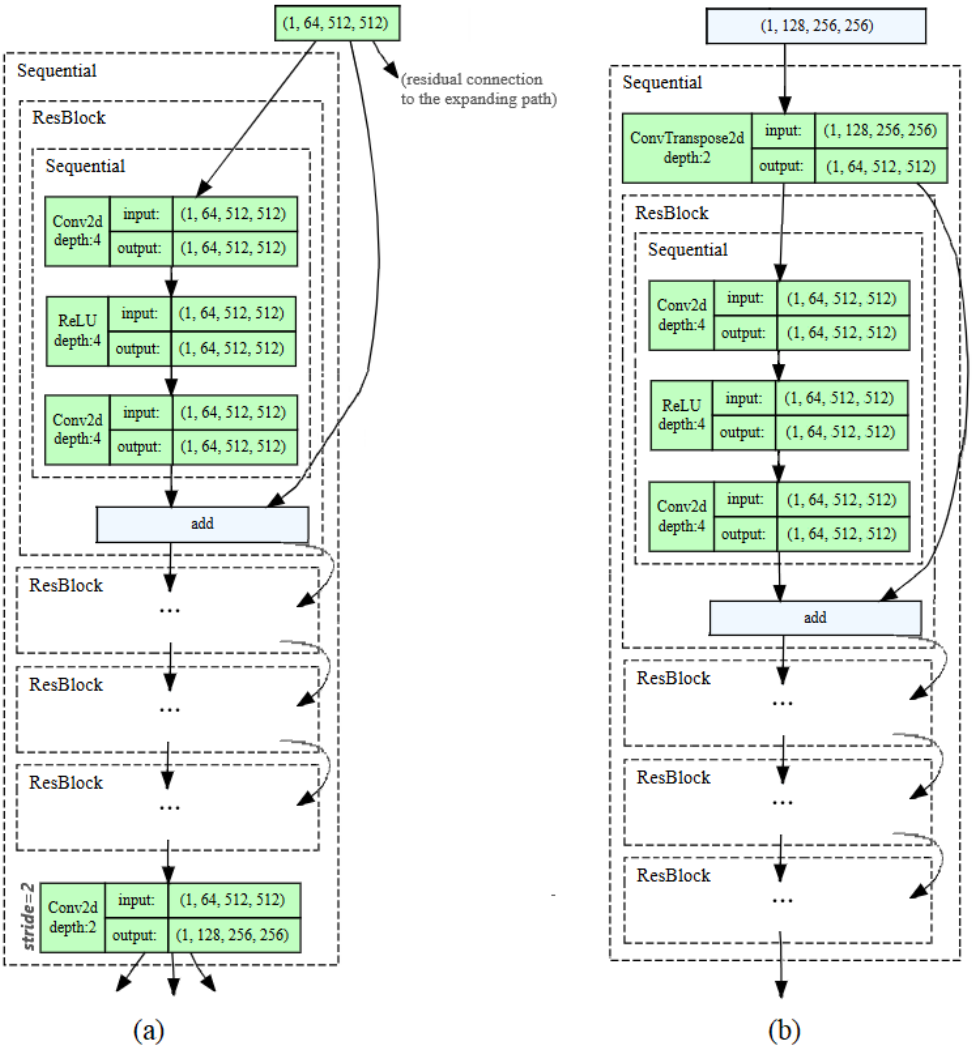


Figure 2.9: Overview of a downsampling (a) and an upsampling (b) step. Only one of the four residual blocks is visualized, since the four residual blocks are identical.

Similarly to the terminology employed for describing the UNet architecture [35], the sequence of the downsampling steps can be referred to as the contracting path, while the corresponding upsampling steps build up the expanding path of the model. The bottleneck layer located between the contracting and the expanding path incorporates four residual blocks and does not change image size and the number of channels. The outer blocks of the model — i.e., the initial and the last one — each perform a single convolution operation and alter only the number of channels. The number of channels increases from the first scale to the fourth one, getting multiplied by 2 with every downsampling step, thus making up 64, 128, 256 and 512 channels.

It is noteworthy that the model includes bias only in the convolution operations in the residual blocks. Outside of these blocks — i.e., in the strided and transposed convolution operations, as well as in both the initial and the last block — bias is not used. The authors deliberately limit the use of additive constants in the convolution operations in DRUNet. This decision is based on the findings of Mohan et al. [31], who demonstrated the advantages of bias-free CNN denoising models in terms of their generalizability across different noise levels.

Another detail of the DRUNet design is that one can find activation functions solely in the residual blocks: every residual block only employs one ReLU function [56], as shown in Fig. 2.9.

The number of trainable parameters in the DRUNet architecture is 32.638.656 [39].

**Pre-training Details**

The DRUNet model employed in this thesis was pre-trained as a Gaussian denoiser on multiple datasets listed in Table 2.1. In total, the combined training dataset comprises 8794 images. Those are natural images containing a wide range of objects including humans, animals, plants, transportation objects, buildings etc.

The input images for training were generated by adding Gaussian noise with a standard deviation sigma to the clean image, with sigma being randomly chosen from the range [0, 50] [56]. Furthermore, the noisy training images were not clipped to the range [0, 255] to avoid altering the noise distribution [56].

| Dataset | Number of images |
|---|---|
| Berkeley Segmentation Dataset [3] | 400 |
| Waterloo Exploration Database [26] | 4744 |
| DIV2K [1] | 900 |
| Flick2K [23] | 2750 |

Table 2.1: Datasets used for pre-training the DRUNet model.

The DRUNet model was trained by minimizing the L1 loss with Adam optimizer and a decreasing learning rate [56]. In each training iteration the model was trained on 16 patches of size 128x128, which were randomly sampled from the training data [56].

### 2.3.2 *Restormer*

Proposed by Zamir et al. in 2021 [53], this model employs an encoder-decoder transformer architecture. It shows capability of grasping distant image context due to learning on whole high-resolution images, without cropping them into local areas. This feature enables Restormer to work with high-resolution images. Its key building block is a modified multi-head attention module that in contrast to the conventional self-attention mechanism has linear complexity. Another essential design solution of Restormer is the introduction of a gating mechanism into the regular feed-forward network, which is supposed to transform features. The convolution operation used at various parts of the network is bias-free [53]. The network contains 26.109.076 trainable parameters [39].

The pre-training was conducted on a dataset comprising 8594 images from several datasets (see Table 2.2). Noisy images for pre-training were produced by introducing additive white Gaussian noise to the undistorted images, with different noise levels $\sigma \in \{15, 25, 50\}$. In this thesis, the network trained with noise level $\sigma = 25$ is utilized.

Apart from showing competitive performance in Gaussian denoising, the model also demonstrated effectiveness in other image restora-

| Dataset | Number of images |
|---|---|
| DIV2K [3] | 800 |
| Flickr2K [23] | 2650 |
| Berkeley Segmentation Dataset 500 (BSDS500) [27] | 400 |
| Waterloo Exploration Database [26] | 4744 |

Table 2.2: Data used for pre-training Restormer model [55].

tion tasks, such as motion deblurring, defocus deblurring, and deraining. [53].

### 2.3.3 *KBNet*

KBNet was introduced in 2023 by Zhang et al. [59]. Its key ingredient is kernel bases attention (KBA) module, which includes a series of learnable kernel bases. Those bases are trained and fused linearly. Therefore, in contrast to previous kernel prediction methods, KBNet does not involve predicting a large number of weights for each pixel, but only the fusing coefficient map determining the fusion of the kernel bases needs to be predicted. KBA module is supposed to be the mechanism effectively aggregating the spatial information.

In addition to being processed by the KBA module, the input features are handled concurrently by two other modules: channel attention and depth-wise convolution. The outputs of those three operations are subsequently multiplied point-wise. This group forms an element called Multi-axes Feature Fusion (MFF) block. The MFF blocks are in turn stacked in both encoder and decoder modules [59].

The number of trainable parameters in KBNet is 141.964.353 [39]. The model was trained on the ImageNet validation dataset with three noise levels $\sigma \in \{15, 25, 50\}$ [59]. In this thesis, the pre-trained model with $\sigma = 25$ is used.

KBNet proved effective on a range of tasks, including denoising, deblurring and deraining. The estimations conducted by the authors

demonstrate that it yields slightly better results than Restormer, while having lower computational demands [59].

## 2.4 FINE-TUNING

Networks were commonly trained from scratch in the earlier stages of deep learning [39], which tended to require extensive data and computational resources. More recently, developments in fields like computer vision and natural language processing (NLP) have demonstrated that fine-tuning pre-trained models can substantially enhance performance, offering a more efficient and effective approach to model design. In [34], a concept of generative pre-training with the following discriminative, i.e. task-specific, fine-tuning was introduced. This approach, estimated on a range of NLP tasks, involved only minimal changes in the model architecture itself [34].

Continued exploration of fine-tuning possibilities has led to the development of a wide array of techniques. In [62], the diverse range of model refinement strategies is referred to as the 'Learn From Model' paradigm, which categorizes these approaches into the following types:

- *Model Tuning* aims at reorientation of the model to downstream tasks by modifying its parameters. It requires understanding of the structure and dynamics of the original model, and thus the ability to identify its components which have the greatest impact on the predictions.
- *Model Distillation* embraces strategies that transform a model in order to make it less computationally demanding, that is, to reduce the runtime speed or to save storage space, without weakening the performance of the original model.
- *Model Reuse* involves combining multiple models either in form of established ensemble methods such as bagging, boosting or stacking or by merging their weights.
- *Meta Learning* focuses on leveraging experience, typically in the form of data distribution, from multiple related tasks to enhance a model's performance on a new task.

- *Model Editing* is a spectrum of techniques updating the knowledge within the model but not necessarily focusing on its usage for a specific downstream task. However, it can be a prerequisite for shifting to a new task.

In this classification, retraining a pre-trained model on new data to adapt it to a new task — the strategy for conducting experiments in this thesis — is most closely aligned with the 'Model Tuning' category. However, the fine-tuning approach employed here may not fully align with the broader 'Learn From Model' paradigm and could be considered outside of its scope.

## 2.5 DATA AUGMENTATION APPROACHES

In scenarios where datasets are limited, data augmentation has proven to be an effective strategy for artificially expanding the training set. Augmentation is supposed to introduce variations to existing data and thus enhance the model's generalization ability [42, 52]. A survey by Shorten et al. [42] categorizes existing image data augmentation techniques into three main groups:

- *Basic image manipulations.* This group encompasses simple and widely used techniques such as geometric transformations (flipping, rotation, scaling and cropping), as well as other straightforward manipulations like sharpening, noise injection and color space transformations.

- *Deep learning-based approaches.* Advanced generative techniques such as Generative Adversarial Networks (GANs) and Neural Style Transfer can synthesize new images which can be used for training.

- *Meta learning approaches.* These strategies involve using neural networks to optimize augmentation policies automatically. For example, AutoAugment [7] searches for augmentation methods that maximize validation accuracy.

Not all augmentation strategies are suitable for every type of data or task. As Shorten et al. [42] emphasize, the "safety" of an augmen-

tation technique, which is understood as the likelihood that the transformed images preserve the correct label, is a critical consideration. For example, rotation is not a "safe" augmentation technique for digit recognition task, where a rotated digit "6" could be misinterpreted as a "9". Therefore, the selection of augmentation methods must be domain- and task-specific [42].

In a broader evaluation of augmentation strategies, Yang et al. [52] explored their effectiveness across such computer vision tasks as semantic segmentation, image classification and object detection. Their findings stress the importance of designing, implementing and testing augmentation methods carefully for each unique dataset and task [52].

In the context of this thesis, a simple augmentation strategy — rotation — has been adopted as an additional exploration to complement training without augmentation. While data augmentation is not the primary focus of this research, it is included to training on some datasets to investigate whether this technique can provide any advantages in improving model performance. The choice to use rotation is justified by the radial symmetry of the point spread function (PSF) — at least of the theoretical one — in fluorescence microscopy data. This symmetry ensures that the rotated images remain consistent with the underlying physical properties of the data, which assumes that rotation is a "safe" augmentation technique for this specific application. Furthermore, the simplicity of rotation augmentation aligns with the practical constraints of this research, since developing a comprehensive, generalizable augmentation policy would extend beyond the scope of the thesis.

# 3

# METHOD AND EXPERIMENTAL SETUPS

## 3.1 DATASETS

Obtaining high-quality training data for fluorescence microscopy is a challenging and labor-intensive process. For this thesis, several publicly available datasets were considered for model training, which originate from the following studies:

- Zhang et al. [61] provided a dataset which contains 12,000 images of various samples, including zebrafish and mouse brain tissues, obtained using three different types of microscopes. However, the images are in 8-bit quality, which limits their utility for high-quality image reconstruction tasks.
- Zhou et al. [63] published a dataset of images of human cells captured using a widefield microscope, offering a narrower scope of samples.
- Weigert et al. [50] made available images of diverse biological samples, including planaria, tribolium, flywing, and liver.
- Hagen et al. [10] provided 10 paired datasets of low-SNR and high-SNR images from two biological samples, captured using widefield and confocal microscopes.

To meet the requirements for the experiments planned in this thesis, the selected dataset needed to consist of paired images suitable for training a denoiser network for the task of image reconstruction and include a sufficient number of images to support robust training.

The dataset of Hagen et al. [10] was chosen as the most suitable option for this thesis due to several compelling factors. First, it provides 32-bit images, offering significantly higher quality compared to the 8-bit datasets used in some previous fluorescence microscopy studies.

Additionally, the dataset features images obtained from two different microscopy techniques — widefield and confocal — allowing for greater versatility and relevance across diverse imaging scenarios. The data from [10] includes paired whole low-SNR and high-SNR images instead of small patches, which is critical for training some models. Furthermore, the dataset is already pre-split into training and test sets in a well-structured manner, which can facilitate the training process. Lastly, [10] provides extra output results generated using three image restoration methods, which can be used as baselines. This eliminates the need for additional overhead in computing baseline results used for comparison. These characteristics collectively make [10] an ideal choice for the objectives of this thesis.

Ten datasets available in [10] are listed in the Table 3.1 including their key characteristics: number of images, image size, microscopy type and the biological structure which is labeled by the imaging. Further details considering microscopes, exposure times, stains and other technical characteristics can be found in [10].

| Dataset | Image size, pixels | Number of images | Microscopy type | Biological structure labeled |
|---|---|---|---|---|
| `Nucleus` | 512x512 | 104 | widefield | Nucleus |
| `ActinConf` | 1024x1024 | 79 | confocal | Actin |
| `MitoConf` | 1024x1024 | 79 | confocal | Mitochondria |
| `Actin60n2` | 2048x2048 | 100 | widefield | Actin |
| `Mito60n2` | 2048x2048 | 100 | widefield | Mitochondria |
| `Actin60n1` | 2048x2048 | 100 | widefield | Actin |
| `Mito60n1` | 2048x2048 | 100 | widefield | Mitochondria |
| `Actin20` | 2048x2048 | 100 | widefield | Actin |
| `Mito20` | 2048x2048 | 100 | widefield | Mitochondria |
| `Membrane` | 2048x2048 | 84 | widefield | Membrane |

Table 3.1: Basic characteristics of the datasets from [10].

It is worth noting that among all ten datasets, `Nucleus` is the one with the smallest image size, 512x512 pixels. Due to the hardware limitations, it is not possible to carry out all the experiments with each dataset. This consideration makes `Nucleus` an ideal candidate in this thesis for the procedures requiring multiple re-trainings such

as hyperparameter tuning, for training the networks with high computational demands as well as for the experiments which are considered optional. However, choosing this dataset for certain experiments has a possible drawback, since the images in this dataset present a structure which is geometrically specific — the high-intensity pixels are centered, building an approximately round form (See Fig. 3.1 a). Given that the denoiser networks may act content-dependent, the selecting of this specific dataset for most of the experiments can limit the generalizability of the conclusions drawn from the experimental results.

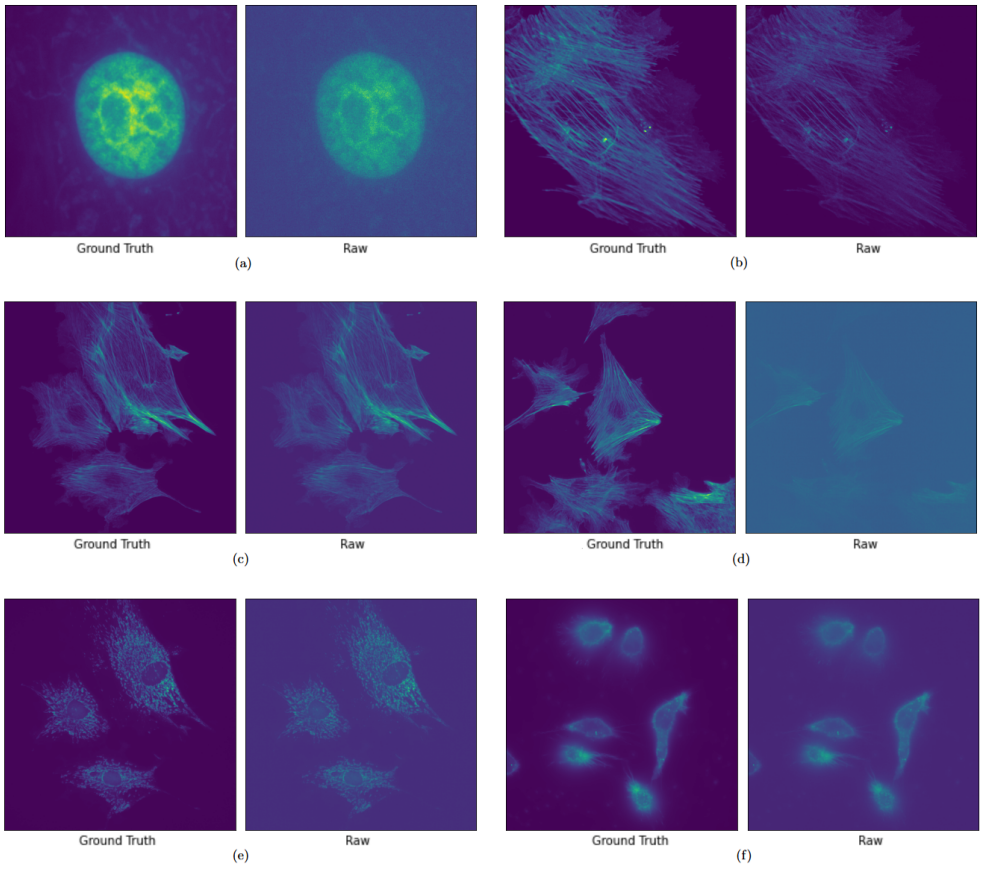


Figure 3.1: Examples of paired images from the datasets used for the training: (a) `Nucleus`, (b) `ActinConf`, (c) `Actin60n1`, (d) `Actin60n2`, (e) `Mito60n1` and (f) `Membrane`.

Another practical consideration for prioritizing certain datasets in the experiments is the noise level of the images. The datasets `Actin60n2` and `Mito60n2` contain the most degraded low-SNR images, due to the shortest exposure times (15 and 10 ms). Because of these imaging

settings, the structures on those images are sometimes indistinguishable for a human eye, which makes their reconstruction especially challenging. On the contrary, produced by the confocal microscope, the images from the datasets `ActinConf` and `MitoConf` have very little blur and noise in comparison to the other datasets, so that the structures on their low-SNR images are quite visible and their image reconstruction might be seen as merely adjusting the overall intensity of the picture. Fig. 3.1 illustrates the diversity of quality of the image pairs drawn from some of the datasets.

## 3.2 DETAILS ON TRAINING AND TEST SETUPS

In each dataset, the last 10% of images — with the number rounded up — are reserved for testing. This specific split is chosen to match the methodology used in [10], where the restored images for the final 10% are provided as publicly available images [11]. By aligning with this approach, the selection of the test set in this thesis remains fully consistent with that in [10], which allows to recalculate the metrics for the test dataset and compare the results of the reconstruction methods presented in [10] and in this thesis.

Of the remaining images, additional 10% are used for validation purposes. This ensures that hyperparameters can be tuned independently of the test set.

Multiple hyperparameter sets are defined, and the model is trained sequentially on the training data. The performance of the model using the respective hyperparameter configuration is then evaluated on the validation data. While an exhaustive search is preferred, training runs with certain hyperparameter sets may be skipped if a particular hyperparameter value shows consistently poor performance compared to others. For the two most computationally demanding networks, hyperparameter tuning is not conducted on all datasets. Instead, the models are fine-tuned using the hyperparameters that demonstrated the best performance on other datasets.

After evaluating multiple hyperparameter sets using the validation data, the best-performing model is selected and retrained on a combined dataset consisting of both training and validation images. The

performance of this final model is then reported as the definitive score. While this approach involves the risk of selecting suboptimal hyperparameters due to the absence of an independent validation set during the final training, it is considered preferable given the small dataset sizes, as it allows for the use of more data during training. Importantly, the test datasets are reserved exclusively for final performance evaluation and are never included in the training process.

Cross-validation was not employed in this thesis due to its high computational cost. Given the size of the datasets and the complexity of the models, performing cross-validation would have significantly prolonged the experiments, making it impractical within the scope of this study.

The number of training epochs for the models is determined by two main factors: resource constraints and model performance. For the most computationally demanding networks, the number of epochs is restricted to fit within the available time. In most cases, however, the number of epochs is selected to ensure that peak model performance is achieved. This is assessed by monitoring the graphs of validation loss and relevant performance metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), discussed in Chapter 3.3. By observing these indicators, the point of optimal accuracy is estimated, aiming for a balance between avoiding underfitting and preventing overfitting.

## 3.3 EVALUATION METRICS

This master thesis adopts two evaluation metrics used in [10]: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), both of which quantify reconstruction quality with higher values indicating better results.

Since the output images yielding from applying the CARE method to the test data are provided [11], the baseline evaluation metrics are not directly copied from [10]. Instead, they are re-calculated using the provided output images. The re-calculated values for both PSNR and SSIM metrics showed slight variations compared to those reported in [10]. These differences could be attributed to variations in metric

computation, scaling procedures or differences in parameter settings of a metric, particularly for SSIM.

### 3.3.1 *Peak Signal-to-Noise Ratio (PSNR)*

Peak Signal-to-Noise Ratio (PSNR) is an established metric that has been widely used for image quality measurement [40]. It is computed as [37]:

$$PSNR = 10 \log_{10} (\frac{peakval^2}{MSE}), \tag{3.1}$$

where *peakval* is the maximal possible value in the image data and $MSE$ is Mean Squared Error between two images, calculated for one-channel images as:

$$MSE = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} [(I_{m,n} - I'_{m,n})^2], \tag{3.2}$$

with $M$ and $N$ representing image resolution, and $I_{m,n}$ and $I'_{m,n}$ the pixel values of the images at the coordinates $m, n$ [40].

PSNR is expressed in decibels. The metric is simple to calculate, however, it is known to have a limited correlation with human perception of image quality [2].

### 3.3.2 *Structural Similarity Index Measure (SSIM)*

Introduced in [48], Structural Similarity Index Measure (SSIM) aims to assess three aspects of images being compared: luminance, contrast and structure. Therefore its computation includes three respective terms [37, 48]:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \tag{3.3}$$

where the terms for luminance, contrast and structure comparison are separately expressed as follows:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{3.4}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{3.5}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{3.6}$$

where $\mu_x$ and $\mu_y$ represent the local means, $\sigma_x$ and $\sigma_y$ the standard deviations and $\sigma_{xy}$ the cross-covariance [37]. Constants $\alpha$, $\beta$ and $\gamma$ can be adjusted to balance the relative importance of each of the three components [48]. In a general case, it is proposed [48] to set $\alpha = \beta = \gamma = 1$ and to set the constant $C_3 = C_2/2$. After these simplifications, the formula for SSIM results in [48]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{3.7}$$

The constants $C_1$ and $C_2$ are included to avoid instability and are dependent on the range of pixel values [48].

The result of SSIM computation is a number in the range [0, 1].

## 3.4 EXPERIMENTAL SETUP

### 3.4.1 *Image Preprocessing*

The objective of this thesis is to use unprocessed, complete images for model training, as this aligns closely with the original data characteristics. However, due to the high computational costs of training certain models, preprocessing steps were necessary for some datasets

to ensure feasible training times and compatibility with hardware limitations.

On the hardware used in this thesis, it is possible to train the DRUNet model without disintegrating the images. However, for the training of Restormer and KBNet, which require greater computational resources, the datasets containing 1024x1024 and 2024x2024 images had to be preprocessed. The images of size 1024x1024 from `ActinConf` and `MitoConf` datasets are therefore cropped into four subimages of size 512x512. For the 2048x2048 datasets, only the four inner fragments of size 512x512 were used for training (see Fig. 3.2). This decision was guided by the observation that, in several of the datasets, e.g. `Actin20` and `Mito20`, the cell structures of interest are predominantly located in the center of the images. By selecting the inner subimages, the training process focuses on regions containing meaningful structures, while excluding peripheral fragments that often consist of background noise. This approach is supposed to maximize the possibility of the training to be conducted on the relevant data, which is especially important given the relatively small sizes of these datasets.

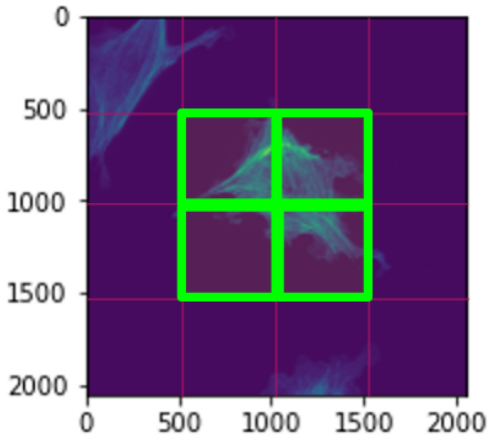


Figure 3.2: For the training of the most computationally demanding models (Restormer, KBNet), only four inner subimages of a 2048x2048 image are used (highlighted in this example image in green).

For consistency and better comparability with the results of the CARE method which is adopted as a baseline, the test images provided in [10] were preprocessed in the same manner. In cases where training images had to be cut into quarters or smaller fragments for

computational reasons, test images underwent the same preprocessing to ensure comparability of results.

### 3.4.2 *Hyperparameter Selection*

The aim of these experiments is to assess the performance of the three Gaussian denoising models discussed in Section 2.3 on the fluorescence microscopy datasets described in Section 3.1, as well as to determine which hyperparameters play the most decisive role in their fine-tuning. To address the research question as comprehensively as possible within this scope, it is first essential to define the specific hyperparameters that will be optimized.

The selected hyperparameters, which can be adjusted during the training of a pre-trained model without altering its architecture, fall into two categories: those related to the loss function and those related to the learning rate.

HYPERPARAMETERS RELATED TO LOSS FUNCTION The selection of the loss function can be regarded as a hyperparameter of the model. For parametric loss functions, the specific choice of parameters directly influences the training process and should therefore be optimized. Hence, the experiments involve training the models using various loss functions:

- Mean absolute error (MAE) loss, also known as L1-loss, is defined as:

$$L_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{x}_i - y_i|,$$

  where $N$ is the number of pixels in an image, $\hat{x}_i$ and $y_i$ are the $i$-th pixels of the output and ground truth images respectively [39].

- Mean squared error (MSE) loss, also known as L2-loss, is defined as [10]:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - y_i)^2,$$

- The composite loss function $L1SSIM$, combining MAE and SSIM, which is inspired by [39], is defined as:

$$L = L_{MAE} + \alpha \cdot (1 - SSIM),$$

where $\alpha$ is the coefficient that controls the balance between MAE and $(1 - SSIM)$ components. The SSIM metric is calculated here according to the same algorithm as discussed in Chapter 3.3.

HYPERPARAMETERS RELATED TO LEARNING RATE Learning rate determines the step size of gradient descent updates during training and should therefore be considered as a critical hyperparameter. Multiple learning rate schedulers have been proposed, such as hyperbolic-tangent decay [14], cyclical learning rate [44], cosine annealing [25], piecewise arc cotangent decay [51]. However, stepwise decay approach is chosen for this master thesis because of its simplicity and intuitive parameters.

The stepwise decay is defined by the following parameters:

- *Initial value* (*lr*): Learning rate value at the beginning of the training.
- *Decay Factor* (*df*): A coefficient by which the learning rate is multiplied after a specified number of epochs. Its value has to be between 0 and 1.
- *Decay Epoch* (*de*): The number of epochs that must pass before the learning rate is reduced by the decay factor.
- *Decay Limit*: The threshold value for the learning rate. Once the learning rate reaches this value or lower, it remains this constant value and is no longer reduced.

Examples of graphs featuring stepwise decay of a learning rate with its parameters designated as described above are shown in Fig. 3.3.

Initial values for most of the tunable parameters are taken from the study [39].

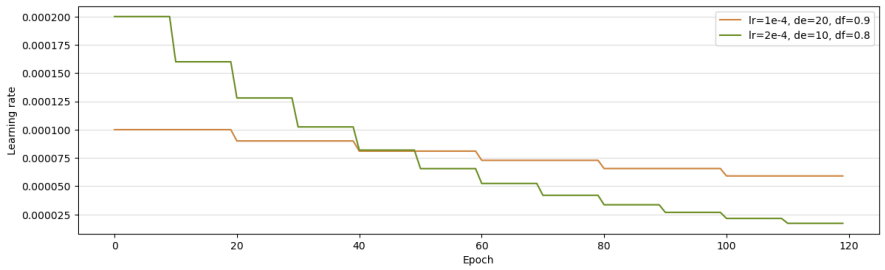


Figure 3.3: Stepwise decay of the learning rate over 120 epochs. The graphs illustrate two parameter configuration.

Another important hyperparameter to consider is the batch size. A small batch size results in frequent weight updates, which can introduce noise due to susceptibility to random fluctuations and thus lead to less coherent optimization. Conversely, large batch sizes may negatively affect generalizability, as they tend to converge to sharp minimizers of the training function [18]. In this thesis, hardware limitations significantly restrict the choice of batch size. For the training of the model DRUnet on 512x512 images, the largest feasible batch size of 16 is used. For models Restormer and KBNet, the batch size is limited to 1. Any attempt to increase the batch size either results in infeasible training times or exceeds the available memory of the GPU. Given these constraints, batch size is not treated as a tunable hyperparameter in this work.

### 3.4.3 *Data Augmentation*

Given the relatively small dataset sizes (between 79 and 104 images), this thesis investigates, whether data augmentation can improve model performance. Among the various augmentation strategies considered (see Section 2.5), a simple rotation-based augmentation was selected. This was achieved by enriching the training dataset with three additional copies of each image, rotated by 90, 180, and 270 degrees. Training with data augmentation is performed only for some combinations of datasets and models.

### 3.4.4 *Training from Scratch*

To evaluate whether pre-trained denoiser models offer advantages over training the same models from scratch, additional experiments are conducted with the denoiser models trained from scratch on the fluorescent microscopy datasets. Training from scratch involves random initialization of the weight values. To limit the number of training procedures and thus the overall computational load of the experiment, the training from scratch is performed not on all of the datasets.

### 3.4.5 *Overview of the Methodology*

A diagram in Fig. 3.4 represents an overview of the experimental procedure, illustrating some of the details mentioned in this Section.

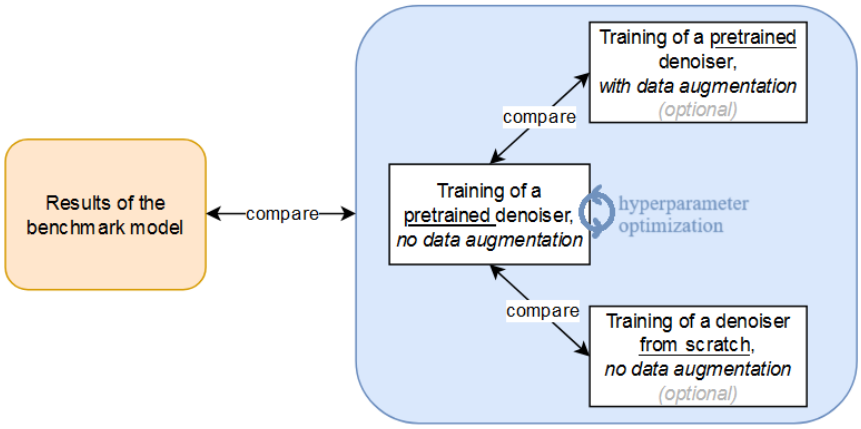


Figure 3.4: Overview of the experimental procedure, performed for a combination model-dataset. Training with data augmentation and training from scratch are performed only for some model-dataset combinations.

### 3.4.6 *Technical Infrastructure*

The model training in this thesis is conducted using an NVIDIA RTX A6000 graphics card, which features 49140 MB of GDDR6 dedicated GPU memory.

All computations, including data preprocessing, model training, testing, and imaging, are performed using Python programming lan-

guage. The denoiser models are implemented with the PyTorch framework, with their pre-trained versions obtained from their respective official repositories. The training and testing scripts are primarily adapted from the scripts provided in [39].

The pre-trained denoiser networks have been downloaded from the following publicly available repositories:

- DRUNet: [57].
- Restormer: [54], with noise level $\sigma = 25$.
- KNBet: [60], with noise level $\sigma = 25$.

# 4

# EVALUATION OF RESULTS

This chapter presents the results of the training procedures conducted based on the setups and considerations outlined in Section 3.4. It begins with an overview of the pre-trained models' performance, including comparisons against the baseline approach, followed by a detailed analysis of specific aspects.

## 4.1 PRE-TRAINED DENOISER MODELS: PERFORMANCE OVERVIEW

The results for each of the three pre-trained Gaussian denoiser models — DRUNet, Restormer and KBNet — are given in the Tables 4.1 (PSNR) and 4.2 (SSIM), where the score for the best performing model in each case is compared to that of the CARE method.

A direct comparison of the results from all three networks is not feasible due to differences in dataset preprocessing required by hardware limitations (see Section 3.4). Specifically, datasets with larger images had to be cropped to accommodate the computational demands of two of the three models. Only the `Nucleus` dataset, which contains smaller images, could be used for training all the models without cropping. The performance of the models trained on this dataset is presented in Table 4.3.

The results indicate that for the given datasets, the pre-trained denoiser networks show slightly better performance in comparison to the CARE method across many metrics, while achieving comparable performance on the remaining ones. Considering the results for the datasets `Actin60n2`, `Mito60n2`, `Actin60n1` and `Mito60n1`, it can be noted that the trained networks demonstrate relatively better performance on data with more severely degraded noisy images. It can be observed that for the datasets containing the large images, the performance of the transformer-based networks is less impressive, which

may be due to the inability to conduct a proper hyperparameter optimization, since their training requires high computational resources.

To illustrate performance of the trained models, some of the output images are presented in Appendix C.

## 4.2 IMPACT OF LOSS FUNCTION

The experiments conducted in this thesis evaluated the impact of different loss functions on the performance of pre-trained image restoration networks for the downstream task of fluorescence microscopy image reconstruction.

Among the tested approaches, the Mean Absolute Error (MAE) loss consistently showed better results than the Mean Squared Error (MSE) loss across both metrics. An example of this performance difference is illustrated in Fig. 4.1, which compares the training results of the pre-trained DRUNet model on `Nucleus` dataset with MAE and MSE loss functions. This observation can be attributed to the distinct properties of the two loss functions: MAE penalizes errors linearly, making it less sensitive to outliers, whereas MSE amplifies larger errors due to its quadratic nature (see Section 3.4.2). In the context of image restoration, where extreme pixel deviations may not represent the overall quality of the output image, MAE appears to better align with quality improvement expressed in the PSNR and SSIM metrics.

Additionally, the use of a combined loss function that incorporates both MAE and Structural Similarity Index (SSIM), denoted as L1SSIM, showed slightly improved performance over MAE in most experiments (see Fig. 4.2). This improvement highlights the benefit of integrating a perceptual quality measure like SSIM, which accounts for structural similarities that are crucial in image restoration tasks. However, the application of the L1SSIM loss function presents practical challenges due to the necessity of tuning additional parameters that balance the contribution of each term in the combined loss. This added complexity may limit its practicality in scenarios requiring streamlined and efficient training processes. Therefore, MAE is chosen as the default loss function for all of the trainings in this thesis.

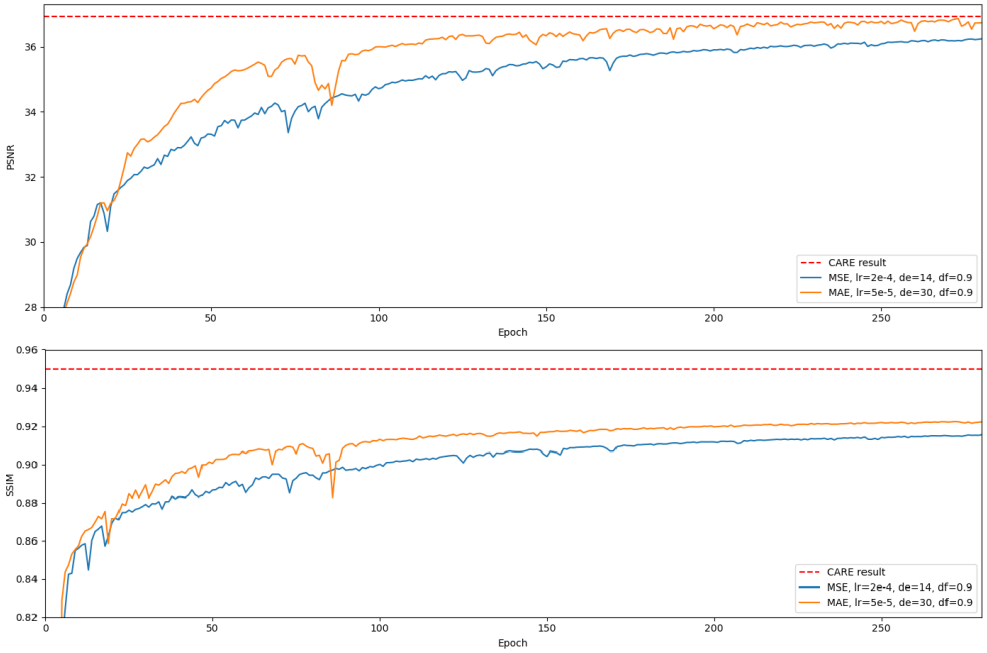


Figure 4.1: PSNR and SSIM: Training of the pre-trained DRUNet network with MAE and MSE loss functions on the `Nucleus` dataset.

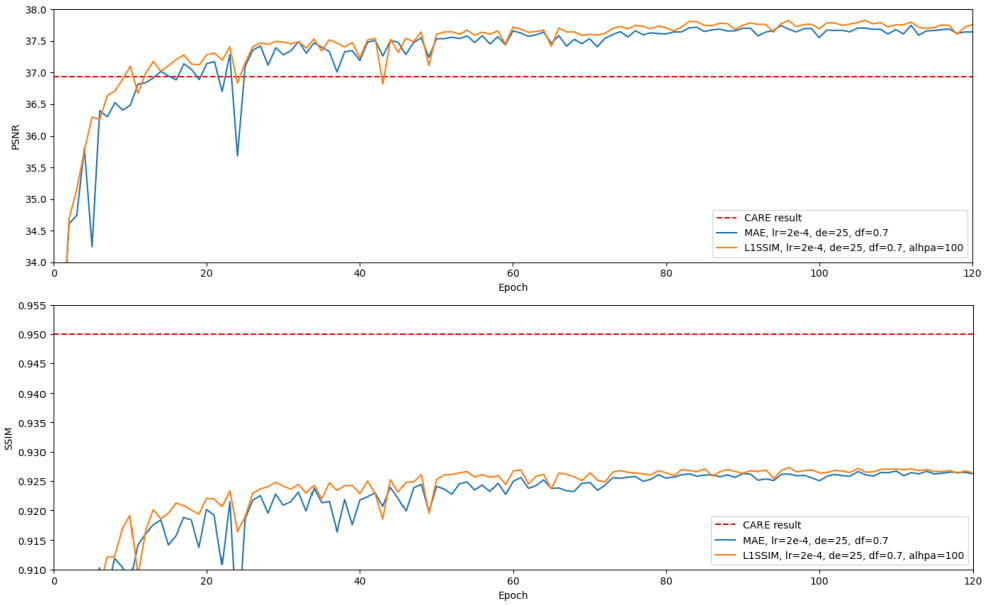


Figure 4.2: PSNR and SSIM: Training of the pre-trained KBNet network with the combined L1SSIM loss function on the `Nucleus` dataset with coefficient $\alpha = 100$ compared to training with MAE loss function ($\alpha = 0$).

## 4.3 IMPACT OF LEARNING RATE

An analysis of the results highlights that the learning rate is a hyperparameter that significantly influences the performance of all evaluated models. As illustrated in Fig. 4.3, which shows the PSNR metric for training the KBNet network, selecting a learning rate smaller than the optimal value results in slower performance improvements, as the model requires more time to converge effectively. Conversely, using a learning rate that exceeds the optimal value introduces instability in the training process, leading to greater fluctuations in the performance metrics. This underscores the importance of carefully tuning the learning rate to balance convergence speed and model stability.

The starting learning rate chosen for training on the most of the datasets is 5e-5 for DRUNet and 1e-4 for Restormer and KBNet.

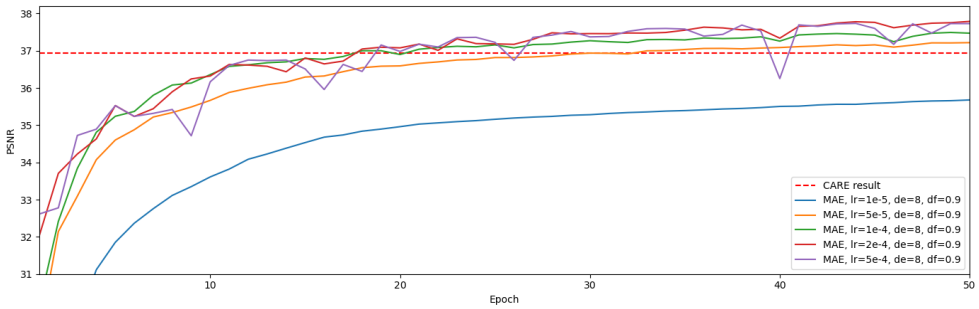


Figure 4.3: PSNR: Training of the pre-trained KBNet network with MAE loss function on the `Nucleus` dataset for various starting learning rates.

## 4.4 IMPACT OF DATA AUGMENTATION

Data augmentation through rotation, as discussed in Section 3.4.3, is applied in certain experiments during network training. The results reveal no substantial improvement in performance achieved through this augmentation technique. Fig. 4.4 provides an example of metric graphs comparing training results with augmented and unaugmented data.

Given the relatively small sizes of the datasets, it was initially expected that augmentation would bring noticeable benefits by increasing the diversity of the training data. Furthermore, the performance

metrics employed, PSNR and SSIM, are sensitive enough to capture even small improvements. Despite these conditions, the impact of rotation-based augmentation appears to be limited.

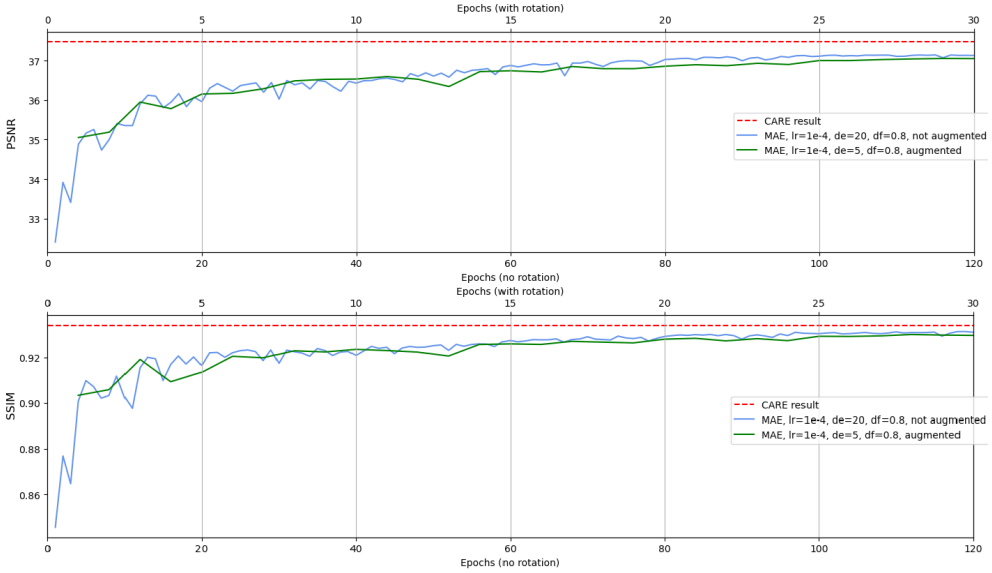


Figure 4.4: PSNR and SSIM: Training of the pre-trained KBNet network with MAE loss function on the augmented and unaugmented data, dataset `Actin60n1`.

One possible explanation for this outcome is that the networks may already generalize well to rotated inputs without requiring explicit augmentation. This could occur if the features relevant to the task are inherently invariant to rotation, allowing the network to learn these patterns effectively even without additional rotated samples. Another factor to consider is the nature of the degradation or noise in the images. If the degradation affects the images uniformly, rotating them may not introduce meaningful variability that contributes to better learning.

These observations suggest that while data augmentation is generally beneficial in many machine learning scenarios [52], its effectiveness depends on the specific characteristics of the task and dataset.

Further trainings on augmented data were performed with DRUNet model, on datasets `Nucleus` and `MitoConf` (see Appendix A).

## 4.5 TRAINING MODELS FROM SCRATCH

The experiments conducted in this thesis compared the performance of pre-trained networks with networks trained from scratch for the task of fluorescence microscopy image restoration. The metric graphs show a significant performance gap at the early stages of training, with pre-trained networks demonstrating much higher initial performance. While networks trained from scratch tend to catch up over time, this process requires substantially more epochs to achieve comparable results.

An example illustrated in Fig. 4.5 (further examples can be found in Appendix B) highlights this difference: the pre-trained network surpassed the baseline result within just 23 epochs, whereas the network trained from scratch required 44 epochs to achieve the same level of performance. This demonstrates the efficiency of pre-trained networks in terms of both time and computational resources. It should be mentioned again that the pre-training does not require additional cost, since it has already been done in other studies.

Further comparisons conducted with various learning rates suggest that adjustments to the learning rate, such as increasing it, can accelerate the convergence of networks trained from scratch, helping them close the performance gap more quickly.

Overall, the results clearly show the advantages of leveraging pre-trained networks for this specific task compared to the training from scratch. By reducing the time and computational resources required for training, pre-trained networks enable faster iteration and experimentation.

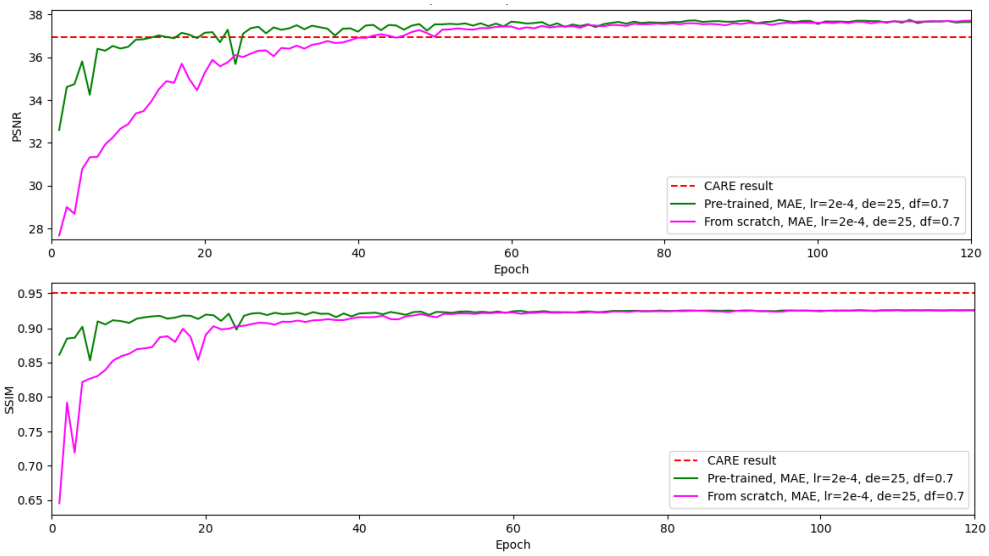


Figure 4.5: PSNR and SSIM: Performance of the pre-trained KBNet network and the KBNet network trained from scratch. MAE loss function, dataset `Nucleus`.

| Dataset | Modus | CARE | DRUNet | Restormer | KBNet |
|---|---|---|---|---|---|
| `Nucleus` | whole | 36.94 | 37.06 ± 1.77 | 37.56 ± 1.96 | **37.73** ± 1.75 |
| `ActinConf` | whole | 33.47 | **33.92** ± 2.18 | | |
| `ActinConf` | quarter | 34.10 | | 33.75 ± 3.02 | **34.44** ± 2.64 |
| `MitoConf` | whole | **30.54** | 30.39 ± 1.46 | | |
| `MitoConf` | quarter | **30.79** | | 29.51 ± 3.50 | 30.30 ± 2.59 |
| `Actin60n2` | whole | 34.18 | **35.39** ± 2.45 | | |
| `Actin60n2` | c-quarter | **30.71** | | 30.30 ± 2.30 | 30.51 ± 2.29 |
| `Mito60n2` | whole | 35.11 | **35.82** ± 2.70 | | |
| `Mito60n2` | c-quarter | 33.77 | | **34.04** ± 2.66 | 33.97 ± 2.63 |
| `Actin60n1` | whole | **44.06** | 42.73 ± 3.25 | | |
| `Actin60n1` | c-quarter | **37.48** | | 36.36 ± 2.83 | 37.08 ± 3.05 |
| `Mito60n1` | whole | **47.42** | 46.53 ± 1.66 | | |
| `Mito60n1` | c-quarter | **40.67** | | 39.83 ± 2.24 | 40.29 ± 2.31 |
| `Actin20` | whole | 41.71 | **42.67** ± 4.65 | | |
| `Actin20` | c-quarter | 36.42 | | 36.40 ± 4.40 | **36.74** ± 4.33 |
| `Mito20` | whole | 41.95 | **43.07** ± 3.41 | | |
| `Mito20` | c-quarter | 37.76 | | **38.27** ± 4.40 | 38.18 ± 4.41 |
| `Membrane` | whole | 41.84 | **42.52** ± 2.91 | | |
| `Membrane` | c-quarter | 38.51 | | 39.42 ± 3.89 | **40.02** ± 4.07 |

Table 4.1: PSNR: Performance of the pre-trained networks compared to the results of CARE method. Preprocessing modi: "whole" - no preprocessing; "quarter" - the images are cropped into quarters; "c-quarter" - only central parts of the images cropped into quarters are used (see Fig. 3.2).

| Dataset | Modus | CARE | DRUNet | Restormer | KBNet |
|---|---|---|---|---|---|
| `Nucleus` | whole | **0.950** | 0.925 ± 0.032 | 0.926 ± 0.032 | 0.926 ± 0.032 |
| `ActinConf` | whole | 0.908 | **0.910** ± 0.022 | | |
| `ActinConf` | quarter | 0.903 | | 0.890 ± 0.053 | **0.908** ± 0.037 |
| `MitoConf` | whole | **0.829** | 0.826 ± 0.031 | | |
| `MitoConf` | quarter | **0.818** | | 0.781 ± 0.102 | 0.810 ± 0.065 |
| `Actin60n2` | whole | 0.923 | **0.939** ± 0.023 | | |
| `Actin60n2` | c-quarter | **0.838** | | 0.825 ± 0.071 | 0.830 ± 0.068 |
| `Mito60n2` | whole | 0.944 | **0.960** ± 0.021 | | |
| `Mito60n2` | c-quarter | 0.923 | | **0.934** ± 0.031 | 0.931 ± 0.032 |
| `Actin60n1` | whole | **0.977** | 0.976 ± 0.012 | | |
| `Actin60n1` | c-quarter | **0.934** | | 0.923 ± 0.040 | 0.930 ± 0.038 |
| `Mito60n1` | whole | 0.991 | **0.992** ± 0.003 | | |
| `Mito60n1` | c-quarter | **0.972** | | 0.970 ± 0.013 | 0.970 ± 0.014 |
| `Actin20` | whole | 0.971 | **0.978** ± 0.013 | | |
| `Actin20` | c-quarter | 0.920 | | 0.921 ± 0.043 | **0.923** ± 0.044 |
| `Mito20` | whole | 0.980 | **0.984** ± 0.009 | | |
| `Mito20` | c-quarter | 0.942 | | **0.944** ± 0.074 | 0.943 ± 0.074 |
| `Membrane` | whole | 0.980 | **0.985** ± 0.006 | | |
| `Membrane` | c-quarter | 0.928 | | 0.936 ± 0.076 | **0.937** ± 0.076 |

Table 4.2: SSIM: Performance of the pre-trained networks compared to the results of CARE method. Preprocessing modi: see Table 4.1.

| Model | PSNR | SSIM | Hyperparameters | Training duration |
|---|---|---|---|---|
| CARE | 36.94 | 0.950 | | |
| DRUNet | 37.06 | 0.925 | MAE, lr=5e-5, de=30, df=0.9, 500 epochs | ∼1.6 hours |
| Restormer | 37.56 | 0.926 | MAE, lr=1e-4, de=20, df=0.9, 120 epochs | ∼13 hours |
| KBNet | 37.73 | 0.926 | MAE, lr=2e-4, de=8, df=0.9, 120 epochs | ∼4.3 hours |

Table 4.3: Performance of the pre-trained denoiser networks, trained on `Nucleus` dataset. Reconstructed images for the CARE method originate from [50].

# 5

# CONCLUSION AND FUTURE WORK

## 5.1 CONCLUSION

This thesis aims to address the Research Questions posed in Section 1.1, focusing on the potential of pre-trained Gaussian denoisers for fluorescence microscopy image reconstruction, the performance of different network architectures when fine-tuned for this purpose, and the impact of key hyperparameters during the fine-tuning process. The findings provide valuable insights into the integration of pre-trained denoisers in the context of fluorescence microscopy.

Firstly, the results demonstrated that pre-trained Gaussian denoisers can significantly enhance fluorescence microscopy image reconstruction. These models, originally developed for general denoising tasks, are not only straightforward to adapt for this specialized domain but also capable of achieving state-of-the-art performance. In some cases, they even exceed the performance of methods specifically designed for microscopy image reconstruction. This adaptability, combined with ease of implementation, suggests a strong case for the continued use and exploration of pre-trained denoisers as the field evolves with improved architectures.

Secondly, an evaluation of UNet- and transformer-based architectures revealed that both are effective for fluorescence microscopy image reconstruction. Transformer-based models may occasionally produce better results, but their computational demands during training pose challenges for rapid and efficient fine-tuning. In contrast, the investigated UNet-based model offered a less resource-intensive alternative while still achieving competitive performance, making this architecture particularly appealing for applications where computational resources are limited or faster training procedures are essential.

Finally, the study highlighted the critical role of hyperparameter selection in fine-tuning Gaussian denoisers. Among these, the choice of

an appropriate loss criterion was found to be a significant factor. Mean Absolute Error (MAE) proved to be a simple yet effective loss function for this task. Additionally, a loss function incorporating Structural Similarity Metrics (SSIM) demonstrates some improvement of the results, though its increased complexity can hinder practical implementation. The learning rate also emerged as a key hyperparameter, with its initial value having a direct impact on training outcomes. Careful tuning of these parameters is essential to maximize the performance of fine-tuned models.

## 5.2 FUTURE WORK

Building upon the findings of this thesis, several directions for future research can be recommended to further explore and enhance the use of pre-trained Gaussian denoiser networks for fluorescence microscopy image reconstruction:

- *Exploring minimal training data requirements.* Producing high-quality training data for fluorescence microscopy is a labor-intensive and resource-demanding process. Therefore, a possible area for future research is to determine the minimal amount of training data required for the fine-tuning approach to still yield satisfactory image reconstruction results. In this case, relying on minimal amount of data for training would move the methodology closer to few-shot learning. Insights from such studies could have practical implications, enabling the adoption of these methods in scenarios where extensive datasets are unavailable.

- *Experimentation with diverse and three-dimensional datasets.* Expanding the scope of experiments to include other fluorescence microscopy datasets, particularly those involving three-dimensional (3D) instead of two-dimensional (2D) images, is a promising direction, since the distribution of distortion in fluorescence microscopy 3D volumes is determined by the specifics of the imaging process.

- *Fine-tuning networks designed for other tasks.* Exploring the fine-tuning of neural network architectures originally developed for

tasks other than denoising, such as deblurring, could be another potential area of research. These networks may bring unique capabilities and advantages when adapted for fluorescence microscopy image reconstruction, broadening the range of tools available for this field.

- *Incorporating denoisers into multistage reconstruction pipelines.* A multistage image reconstruction pipeline integrating the fine-tuning of pre-trained denoiser networks with other techniques, such as conventional deconvolution algorithms or additional neural network-based methods, could bring further improvements in reconstruction quality.

- *Addressing further specific challenges in fluorescence microscopy.* It could be investigated how image reconstruction advantages obtained through fine-tuning of pre-trained denoiser networks can be used for handling other challenges or downstream tasks related to fluorescence microscopy data, such as super-resolution, anisotropy restoration and segmentation (cf. diverse applications highlighted in [50]).

Part II

# APPENDIX

# A

# DATA AUGMENTATION: PERFORMANCE DIAGRAMS

Performance graphs for two extra trainings with data augmentation, compared to the training on unaugmented data.

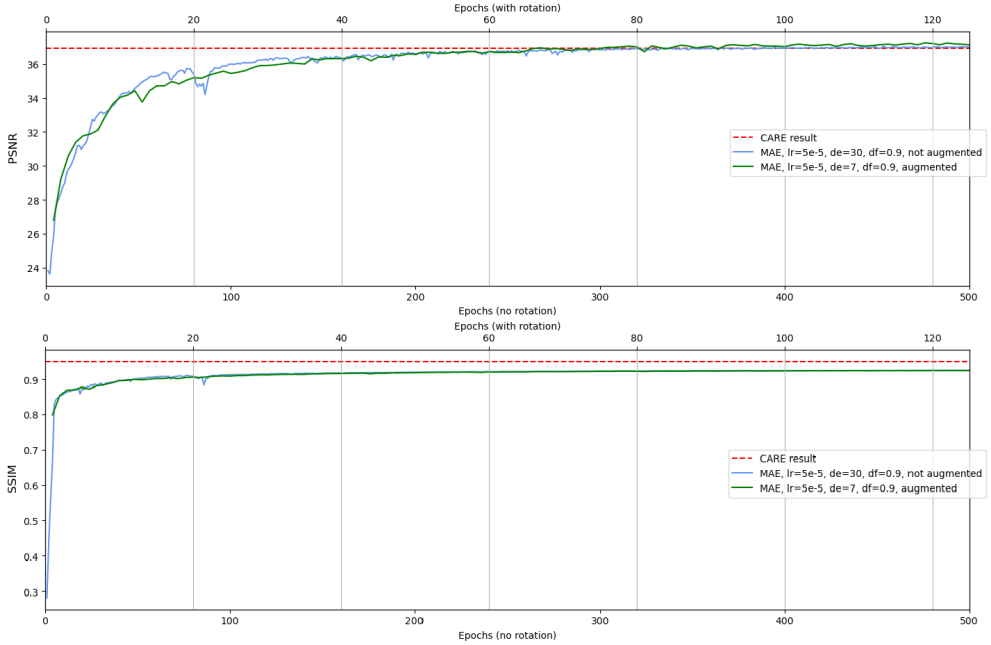


Figure A.1: PSNR and SSIM: Training of the pre-trained DRUNet network with MAE loss function on the augmented and unaugmented data, dataset `Nucleus`.

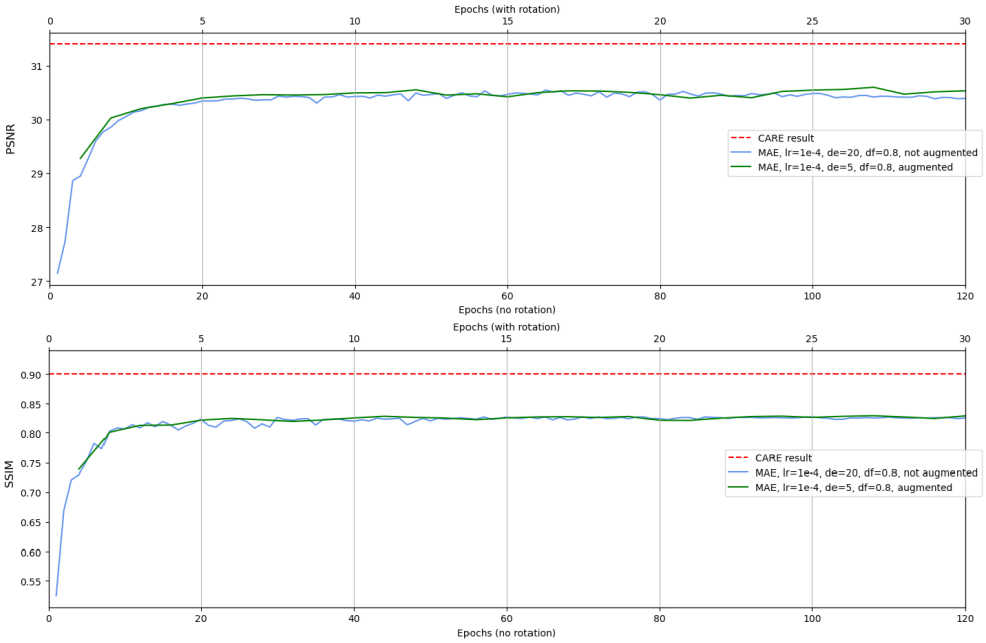


Figure A.2: PSNR and SSIM: Training of the pre-trained DRUNet network with MAE loss function on the augmented and unaugmented data, dataset `MitoConf`.

# B

# TRAINING FROM SCRATCH: PERFORMANCE DIAGRAMS

Performance graphs for the trainings from scratch, compared to pre-trained networks.

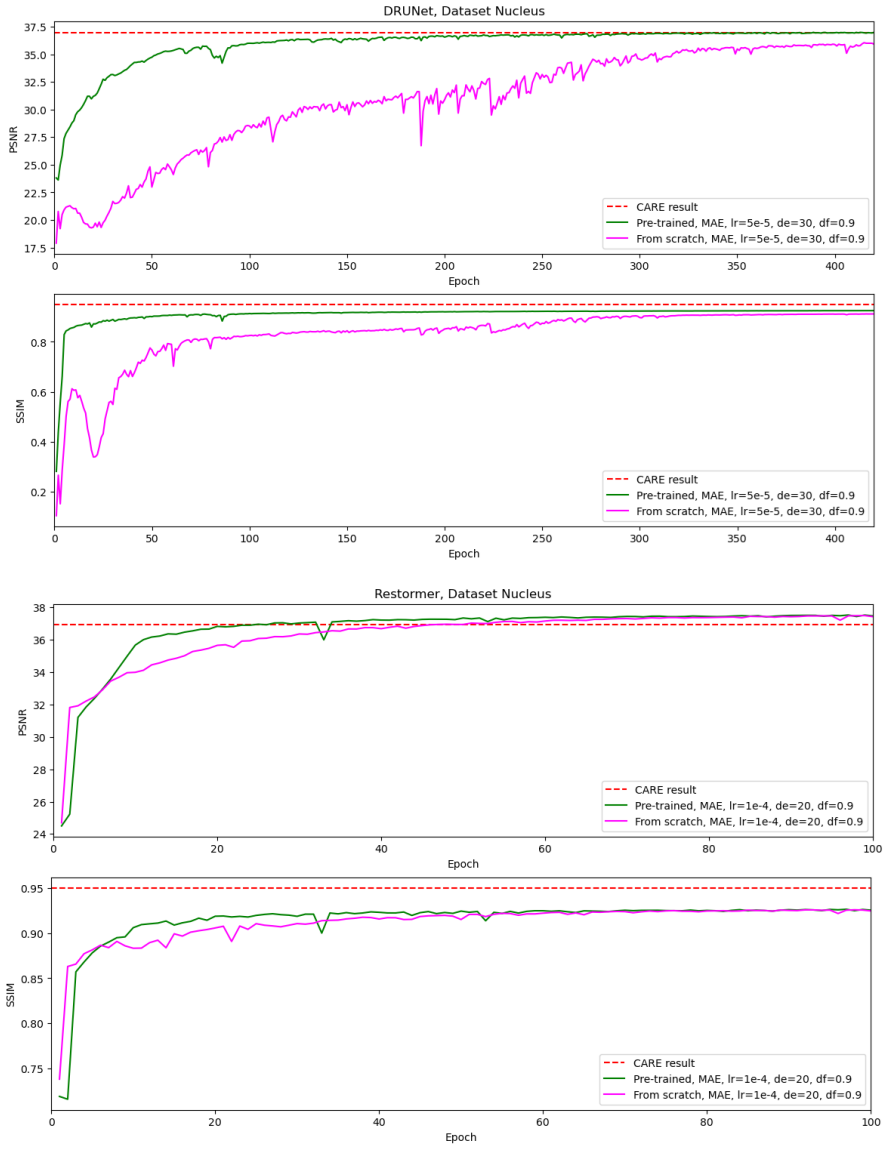


Figure B.1: PSNR and SSIM: Performance of the pre-trained network (above: DRUNet, below: Restormer) and the same network trained from scratch. MAE loss function, dataset `Nucleus`.

# C

# OUTPUT IMAGES

Output images for the three pre-trained networks, compared to Ground Truth (GT), Raw and output for the CARE method.

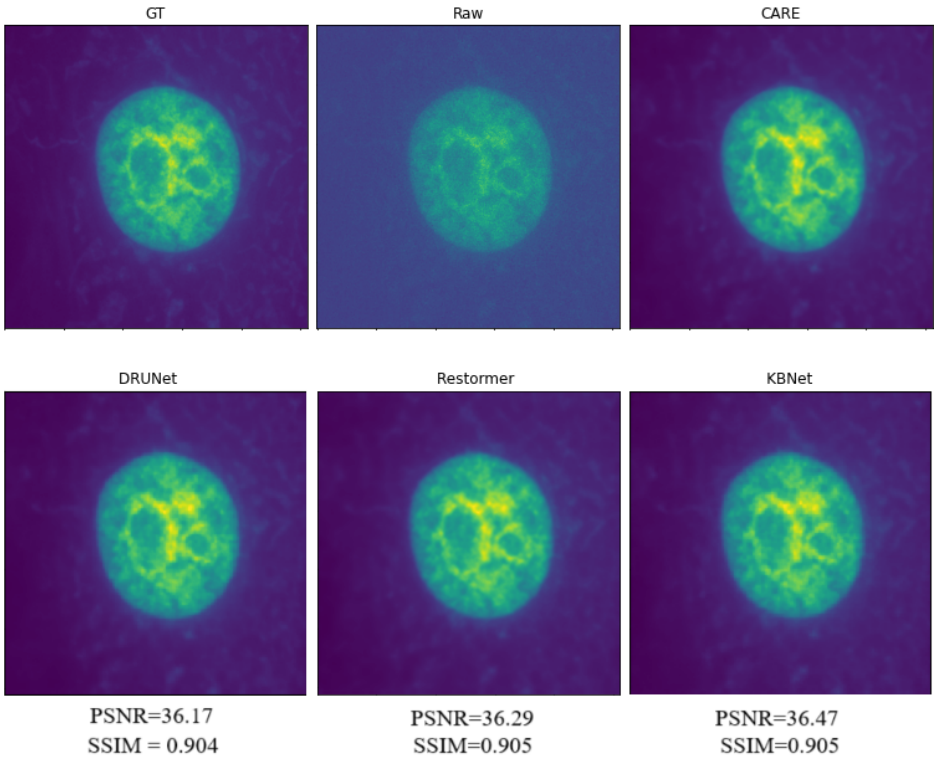


Figure C.1: Output images, drawn from dataset `Nucleus`. MAE loss function.

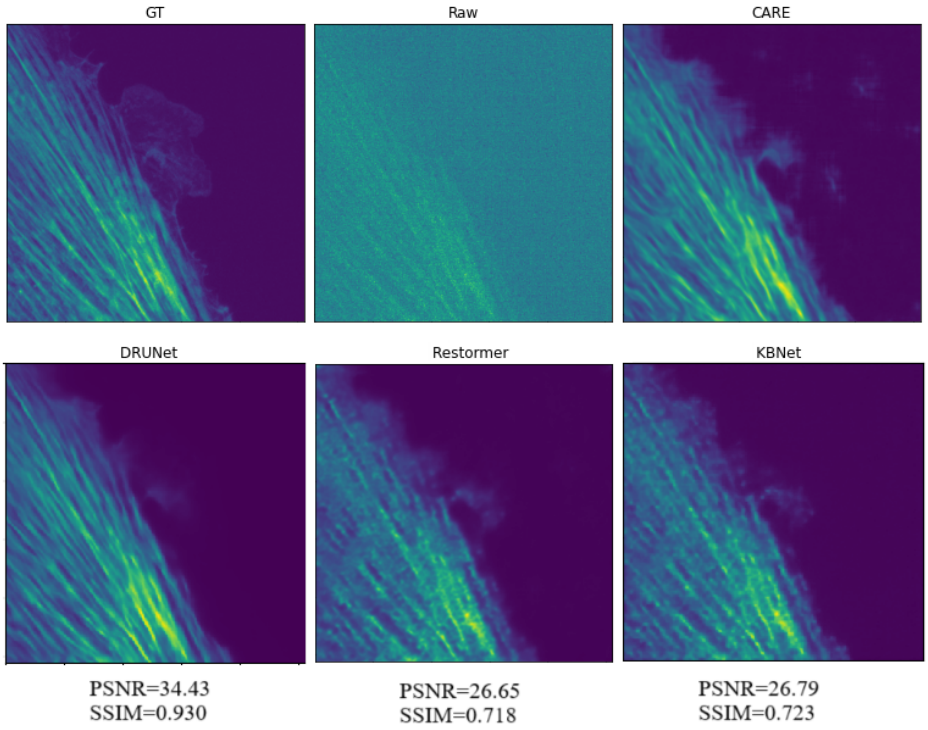


Figure C.2: Output images, drawn from dataset `Actin60n2`. MAE loss function.

# BIBLIOGRAPHY


[1] Eirikur Agustsson and Radu Timofte. "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 1122–1131. DOI: 10.1109/CVPRW.2017.150.

[2] Yusra Al-Najjar. "Comparative Analysis of Image Quality Assessment Metrics: MSE, PSNR, SSIM and FSIM." In: *International Journal of Science and Research (IJSR)* 13 (Mar. 2024), pp. 110–114. DOI: 10.21275/SR24302013533.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. "Contour Detection and Hierarchical Image Segmentation." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.5 (May 2011), pp. 898–916. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.161.

[4] Peter Bankhead. *Analyzing fluorescence microscopy images with ImageJ*. 2014. URL: https://petebankhead.gitbooks.io/imagej-intro/content/ (visited on 01/01/2025).

[5] Joshua Batson and Loic Royer. "Noise2Self: Blind Denoising by Self-Supervision." In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 524–533. URL: https://proceedings.mlr.press/v97/batson19a.html.

[6] Alexey Bogdanov, Ekaterina Bogdanova, Dmitriy Chudakov, Tatiana Gorodnicheva, Sergey Lukyanov, and Konstantin Lukyanov. "Cell culture medium affects GFP photostability: A solution." In: *Nature methods* 6 (Dec. 2009), pp. 859–60. DOI: 10.1038/nmeth1209-859.

[7] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. *AutoAugment: Learning Augmentation Poli-*

*cies from Data*. 2019. arXiv: 1805.09501 [cs.CV]. URL: https://arxiv.org/abs/1805.09501.

[8] Media Cybernetics. *AutoQuant Deconvolution Image-Pro Media Cybernetics*. 2024. URL: https://mediacy.com/image-pro/autoquant-deconvolution/ (visited on 01/01/2025).

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. "Image denoising with block-matching and 3D filtering." In: *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*. Ed. by Nasser M. Nasrabadi, Syed A. Rizvi, Edward R. Dougherty, Jaakko T. Astola, and Karen O. Egiazarian. Vol. 6064. International Society for Optics and Photonics. SPIE, 2006, p. 606414. DOI: 10.1117/12.643267.

[10] Guy M Hagen, Justin Bendesky, Rosa Machado, Tram-Anh Nguyen, Tanmay Kumar, and Jonathan Ventura. "Fluorescence microscopy datasets for training deep neural networks." In: *GigaScience* 10.5 (May 2021), giab032. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab032.

[11] Guy M Hagen, Justin Bendesky, Rosa Machado, Tram-Anh Nguyen, Tanmay Kumar, and Jonathan Ventura. *Supporting data for "Fluorescence Microscopy Datasets for Training Deep Neural Networks"*. 2021. URL: https://doi.org/10.5524/100888 (visited on 01/01/2025).

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[13] Hoebe, RA, Van Oven, CH, Gadella Jr, TWJ, Dhonukshe, Pankaj, Van Noorden, CJF, and Manders, EMM. "Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging." In: *Nature Biotechnology* (Jan. 2007). ISSN: 1087-0156. DOI: 10.1038/nbt1278.

[14] Bo Yang Hsueh, Wei Li, and I-Chen Wu. *Stochastic Gradient Descent with Hyperbolic-Tangent Decay on Classification*. 2018. arXiv: 1806.01593 [cs.CV]. URL: https://arxiv.org/abs/1806.01593.

[15] Jaroslav Icha, Michael Weber, Jennifer Waters, and Caren Norden. "Phototoxicity in live fluorescence microscopy, and how to avoid it." In: *Bioessays* 39 (July 2017), pp. 1700003–15. DOI: 10.1002/bies.201700003.

[16] James Jonkman, Claire M. Brown, Graham Daniel Wright, Kurt I. Anderson, and Alison J. North. "Tutorial: guidance for quantitative confocal microscopy." In: *Nature Protocols* 15 (2020), pp. 1585–1611. URL: https://api.semanticscholar.org/CorpusID:214720026.

[17] Tal Kenig, Zvi Kam, and Arie Feuer. "Blind Image Deconvolution Using Machine Learning for Three-Dimensional Microscopy." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (2010), pp. 2191–2204. DOI: 10.1109/TPAMI.2010.45.

[18] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*. 2017. arXiv: 1609.04836 [cs.LG]. URL: https://arxiv.org/abs/1609.04836.

[19] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. *Noise2Void - Learning Denoising from Single Noisy Images*. 2019. arXiv: 1811.10980 [cs.CV]. URL: https://arxiv.org/abs/1811.10980.

[20] Philippe Laissue, Rana Alghamdi, Pavel Tomancak, Emmanuel Reynaud, and Hari Shroff. "Assessing phototoxicity in live fluorescence imaging." In: *Nature Methods* 14 (June 2017), pp. 657–661. DOI: 10.1038/nmeth.4344.

[21] Rui Li, Mikhail Kudryashev, and Artur Yakimovich. *Solving the inverse problem of microscopy deconvolution with a residual Beylkin-Coifman-Rokhlin neural network*. 2024. DOI: https://doi.org/10.48550/arXiv.2407.03239. arXiv: 2407.03239 [q-bio.QM].

[22] Jianhui Liao, Junle Qu, Yongqi Hao, and Jia Li. "Deep-learning-based methods for super-resolution fluorescence microscopy." In: *Journal of Innovative Optical Health Sciences* 16.03 (2023), p. 2230016. DOI: 10.1142/S1793545822300166.

[23] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. "Enhanced Deep Residual Networks for Single Image Super-Resolution." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 1132–1140. DOI: 10.1109/CVPRW.2017.151.

[24] Sungjun Lim, Sang-Eun Lee, Sunghoe Chang, and Jong Chul Ye. *Blind Deconvolution Microscopy Using Cycle Consistent CNN with Explicit PSF Layer*. 2019. arXiv: 1904.02910 [cs.LG]. URL: https://arxiv.org/abs/1904.02910.

[25] Ilya Loshchilov and Frank Hutter. *SGDR: Stochastic Gradient Descent with Warm Restarts*. 2017. arXiv: 1608.03983 [cs.LG]. URL: https://arxiv.org/abs/1608.03983.

[26] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. "Waterloo Exploration Database: New Challenges for Image Quality Assessment Models." In: *IEEE Transactions on Image Processing* 26.2 (2017), pp. 1004–1016. DOI: 10.1109/TIP.2016.2631888.

[27] D. Martin, C. Fowlkes, D. Tal, and J. Malik. "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics." In: *Proc. 8th Int'l Conf. Computer Vision*. Vol. 2. 2001, pp. 416–423.

[28] James Mcnally, Tatiana Karpova, John Cooper, and José Conchello. "Three-Dimensional Imaging by Deconvolution Microscopy." In: *Methods (San Diego, Calif.)* 19 (Dec. 1999), pp. 373–85. DOI: 10.1006/meth.1999.0873.

[29] William Meiniel, Jean-Christophe Olivo-Marin, and Elsa D. Angelini. "Denoising of Microscopy Images: A Review of the State-of-the-Art, and a New Sparsity-Based Method." In: *IEEE Transactions on Image Processing* 27 (2018), pp. 3842–3856. DOI: 10.1109/TIP.2018.2819821.

[30] Leica Microsystems. *Fluorescene in Microscopy*. 2011. URL: https://www.leica-microsystems.com/science-lab/life-science/fluorescence-in-microscopy/ (visited on 12/26/2024).

[31] Sreyas Mohan, Zahra Kadkhodaie, Eero P. Simoncelli, and Carlos Fernandez-Granda. *Robust and interpretable blind image denoising via bias-free convolutional neural networks*. 2020. arXiv: 1906.05478 [eess.IV]. URL: https://arxiv.org/abs/1906.05478.

[32] Nahida Nazir, Abid Sarwar, and Baljit Singh Saini. "Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges." In: *Micron* 180 (2024), p. 103615. ISSN: 0968-4328. DOI: https://doi.org/10.1016/j.micron.2024.103615.

[33] Jim Pawley. "The 39 Steps: A Cautionary Tale of Quantitative 3-D Fluorescence Microscopy." In: *BioTechniques* 28.5 (2000), pp. 884–887. DOI: 10.2144/00285bt01.

[34] Alec Radford and Karthik Narasimhan. "Improving Language Understanding by Generative Pre-Training." In: 2018. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]. URL: https://arxiv.org/abs/1505.04597.

[36] Scientific Volume Imaging (SVI). *Huygens Software by Scientific Volume Imaging*. 2025. URL: https://svi.nl/Huygens-Software (visited on 01/01/2025).

[37] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study." In: *Journal of Computer and Communications* 07 (Jan. 2019), pp. 8–18. DOI: 10.4236/jcc.2019.73002.

[38] P. Sarder and A. Nehorai. "Deconvolution methods for 3-D fluorescence microscopy images." In: *IEEE Signal Processing Magazine* 23.3 (2006), pp. 32–45. DOI: 10.1109/MSP.2006.1628876.

[39] Tim Selig, Thomas März, Martin Storath, and Andreas Weinmann. *Enhanced Low-Dose CT Image Reconstruction by Domain and Task Shifting Gaussian Denoisers*. 2024. arXiv: 2403.03551 [eess.IV]. URL: https://arxiv.org/abs/2403.03551.

[40] De Rosal Igantius Moses Setiadi. "PSNR vs SSIM: imperceptibility quality assessment for image steganography." In: *Multimedia Tools Appl.* 80.6 (Mar. 2021), 8423–8444. ISSN: 1380-7501. DOI: 10.1007/s11042-020-10035-z.

[41] Adrian Shajkofci and Michael Liebling. "Semi-Blind Spatially-Variant Deconvolution in Optical Microscopy with Local Point Spread Function Estimation by Use of Convolutional Neural Networks." In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. 2018, pp. 3818–3822. DOI: 10.1109/ICIP.2018.8451736.

[42] Connor Shorten and Taghi Khoshgoftaar. "A survey on Image Data Augmentation for Deep Learning." In: *Journal of Big Data* 6 (July 2019). DOI: 10.1186/s40537-019-0197-0.

[43] Jean-Baptiste Sibarita. "Deconvolution Microscopy." In: *Advances in biochemical engineering/biotechnology* 95 (Feb. 2005), pp. 201–43. DOI: 10.1007/b102215.

[44] Leslie N. Smith. *Cyclical Learning Rates for Training Neural Networks*. 2017. arXiv: 1506.01186 [cs.CV]. URL: https://arxiv.org/abs/1506.01186.

[45] Kogiku Shiba Shoji A. Baba Takuya Nishigaki Chris D. Wood and Alberto Darszon. "Stroboscopic illumination using light-emitting diodes reduces phototoxicity in fluorescence cell imaging." In: *BioTechniques* 41.2 (2006), pp. 191–197. DOI: 10.2144/000112220.

[46] Hongda Wang, Yair Rivenson, Yiyin Jin, Zhensong Wei, Ronald Gao, Harun Gunaydin, Laurent Bentolila, Comert Kural, and Aydogan Ozcan. "Deep learning enables cross-modality super-resolution in fluorescence microscopy." In: *Nature Methods* 16 (Jan. 2019). DOI: 10.1038/s41592-018-0239-0.

[47] Yina Wang, Henry Pinkard, Emaad Khwaja, Shuqin Zhou, Laura Waller, and Bo Huang. "Image denoising for fluorescence microscopy by supervised to self-supervised transfer learning." In: *Opt. Express* 29.25 (2021), pp. 41303–41312. DOI: 10.1364/OE.434191.

[48] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. "Image quality assessment: from error visibility to structural similarity." In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[49] Martin Weigert, Loic Royer, Florian Jug, and Gene Myers. *Isotropic reconstruction of 3D fluorescence microscopy images using convolutional neural networks*. 2017. arXiv: 1704.01510 [cs.CV]. URL: https://arxiv.org/abs/1704.01510.

[50] Martin Weigert et al. "Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy." In: *bioRxiv* (2018). DOI: 10.1101/236463.

[51] Haixu Yang, Jihong Liu, Hongwei Sun, and Henggui Zhang. "PACL: Piecewise Arc Cotangent Decay Learning Rate for Deep Neural Network Training." In: *IEEE Access* 8 (2020), pp. 112805–112813. DOI: 10.1109/ACCESS.2020.3002884.

[52] Suorong Yang, Weikang Xiao, Mengchen Zhang, Suhan Guo, Jian Zhao, and Furao Shen. *Image Data Augmentation for Deep Learning: A Survey*. 2023. arXiv: 2204.08610 [cs.CV]. URL: https://arxiv.org/abs/2204.08610.

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. *Restormer: Efficient Transformer for High-Resolution Image Restoration*. 2022. arXiv: 2111.09881 [cs.CV]. URL: https://arxiv.org/abs/2111.09881.

[54] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. *Restormer Repository on GitHub*. 2022. URL: https://github.com/swz30/Restormer (visited on 01/01/2025).

[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. *Supplementary Material – Restormer: Efficient Transformer for High-Resolution Image Restoration*. 2022. URL: https://openaccess.thecvf.com/content/CVPR2022/supplemental/Zamir_Restormer_Efficient_Transformer_CVPR_2022_supplemental.pdf (visited on 01/01/2025).

[56] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. *Plug-and-Play Image Restoration with Deep Denoiser Prior*. 2021. arXiv: 2008.13751 [eess.IV]. URL: https://arxiv.org/abs/2008.13751.

[57] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. *Plug-and-Play Image Restoration with Deep Denoiser Prior (IEEE TPAMI 2021) (Repository on GitHub)*. 2022. URL: https://github.com/cszn/DPIR (visited on 01/01/2025).

[58] Kai Zhang, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising." In: *IEEE Transactions on Image Processing* 27.9 (Sept. 2018), 4608–4622. ISSN: 1941-0042. DOI: 10.1109/tip.2018.2839891.

[59] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. *KBNet: Kernel Basis Network for Image Restoration*. 2023. arXiv: 2303.02881 [cs.CV]. URL: https://arxiv.org/abs/2303.02881.

[60] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. *KBNet Repository on GitHub*. 2023. URL: https://github.com/zhangyi-3/KBNet (visited on 01/01/2025).

[61] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. "A Poisson-Gaussian Denoising Dataset With Real Fluorescence Microscopy Images." In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 11702–11710. DOI: 10.1109/CVPR.2019.01198.

[62] Hongling Zheng, Li Shen, Anke Tang, Yong Luo, Han Hu, Bo Du, and Dacheng Tao. *Learn From Model Beyond Fine-Tuning: A Survey*. 2023. arXiv: 2310.08184 [cs.AI]. URL: https://arxiv.org/abs/2310.08184.

[63] Ruofan Zhou, Majed El Helou, Daniel Sage, Thierry Laroche, Arne Seitz, and Sabine Süsstrunk. *W2S: Microscopy Data with Joint Denoising and Super-Resolution for Widefield to SIM Mapping*. 2020. arXiv: 2003.05961 [eess.IV]. URL: https://arxiv.org/abs/2003.05961.