

Universität Osnabrück  
Fachbereich Humanwissenschaften  
Institute of Cognitive Science

Masterthesis

**AI-based Reconstruction and Denoising for Robust  
Structured Illumination Microscopy at Low  
Signal-to-Noise Ratios**

Argha Sarker  
990398  
Master's Program Cognitive Science

First supervisor: Prof. Dr. Kietzmann, Tim Christian  
Institute of Cognitive Science  
Universität Osnabrück

Second supervisor: Prof. Dr. rer. nat. Jacob Piehler  
Institute of Biologie/Chemie  
Universität Osnabrück

# **Declaration of Authorship**

---

**Name, Vorname / Full Name:**

---

**Matrikelnummer / student number:**

The content of this thesis represents my own knowledge, my own understanding, and my own perspective on the topic. In case artificial intelligence tools were used, their way and purpose of usage have been made transparent. In addition, I have cited all my sources in accordance with academic standards. I am ready and able to explain and defend the positions developed in this thesis. This thesis has not been submitted, either in part or whole, at this or any other university.

---

city, date

---

signature

## Abstract

Fluorescence microscopy is a powerful imaging technique that enables the visualization of biological specimens at cellular and subcellular level with a resolution limit of 200-300 nm. The development of structured illumination microscopy (SIM) has further enhanced resolution by exploiting the interference of light patterns to achieve SIM reconstructions with two times the resolution. However, imaging biological specimens under microscopy faces fundamental challenges, such as imaging speed, light exposure, and spatial resolution. Specifically, imaging biological specimens often requires rapid acquisition with high temporal resolution and low light exposure to reduce photo-bleaching and photo-toxicity. These requirements typically lead to low-SNR imaging, which results in poor SIM reconstruction with a high level of reconstruction artifacts. In contrast, high-SNR imaging produces superior quality at the cost of slower speeds and increased light exposure, which can potentially jeopardize the sensitive specimens.

In this work, we explore the feasibility of recovering spatial and temporal resolution hidden in low-SNR images by leveraging state-of-the-art deep learning techniques. Our approach is centered on a novel direct SIM reconstruction methodology using Projection Upsampling Network (PU-Net), which replaces the Deep Fourier Channel Attention Network (DFCAN) module in the current RDL-SIM training pipeline, which consists of predicting SIM reconstruction images with a pre-trained network (DFCAN) and convolving this image with a prior knowledge-based sinusoidal pattern to simulate moiré fringes for training the final denoising module, as proposed by Qiao et al., 2022.

# Acknowledgments

I would like to express my sincere gratitude to everyone who has contributed their knowledge, guidance, and support in shaping this project. Each of these contributions has played a significant role in the completion of this work, and I am profoundly grateful for the support and inspiration provided by all.

I am deeply thankful to Prof. Dr. Tim Christian Kietzmann for generously sharing his expertise and providing invaluable direction throughout the development of this work. His thoughtful insights have been essential in refining the project's scope and methodology.

I am also grateful to Dr. Varun Kapoor, whose enthusiastic contributions were instrumental in bringing this project to fruition. His robust code base and extensive experience in bio-image analysis provided the foundation upon which I built and extended my own work and experiments. Varun has enabled me to mature my understanding, better conceptualize the problem, oversee the experiments, and refine the machine learning aspects with his timely feedback.

Special thanks to Prof. Dr. Jacob Piehler and Dr. Rainer Kurre for introducing me to bioimage analysis and microscopy. Dr. Kurre's innovative project idea and expert guidance have allowed me to navigate the biological aspects of the project. His practical insights into bioimage analysis have significantly enriched my understanding and approach to this challenging subject.

Finally, I would like to acknowledge Dr. Susanne Kunis for her support in facilitating my involvement in this project during my student assistantship at CellNanOs. Her assistance and commitment were crucial to the continuation of this project throughout my thesis.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Structured Illumination Microscopy . . . . .	2
1.2	Implementation of SIM and SIM reconstruction . . . . .	5
1.3	Biological Specimens . . . . .	6
1.3.1	Microtubules . . . . .	6
1.3.2	F-actin . . . . .	6
1.3.3	Structural Complexity of Different Specimens . . . . .	7
1.4	Model Architectures . . . . .	9
1.4.1	Content Aware Image Restoration (CARE) . . . . .	9
1.4.2	Deep Fourier Channel Attention Network (DFCAN) . . . . .	11
1.4.3	ScU-Net . . . . .	12
1.4.4	Rationalized Deep Learning Super-Resolution Microscopy (RDL-SIM) . . . . .	13
1.4.5	Projection Upsampling Network (PU-Net) . . . . .	15
1.5	Aims and Objectives . . . . .	17
<b>2</b>	<b>Methods</b>	<b>19</b>
2.1	Dataset . . . . .	19
2.1.1	BioSR . . . . .	19
2.1.2	Acquisition of our F-actin Dataset . . . . .	19
2.1.3	Training Data Generation for SIM Reconstruction Module . .	20
2.1.4	Training Data Generation for Image Denoising Module . . .	21
2.2	quantitative assessment of image reconstruction . . . . .	21
2.2.1	Mean Squared Error . . . . .	21
2.2.2	Normalized Root Mean Square Error . . . . .	22
2.2.3	Peak Signal-to-Noise Ratio . . . . .	23
2.2.4	Structural Similarity Index (SSIM) . . . . .	24
2.2.5	Multi-Scale Structural Similarity Index (MS-SSIM) . . . . .	26
2.2.6	Decorrelation Analysis . . . . .	27
2.2.7	FairSIM . . . . .	28

2.2.8	Quantitative assessment summary . . . . .	29
2.3	Training Details . . . . .	30
2.3.1	Training Details of SIM Reconstruction Module . . . . .	30
2.3.2	Training Details of Image Denoising Module . . . . .	32
<b>3</b>	<b>Results</b>	<b>34</b>
3.1	SIM Reconstruction Module . . . . .	34
3.2	Image Denoising Module . . . . .	43
<b>4</b>	<b>Discussion and Outlook</b>	<b>52</b>
<b>5</b>	<b>Declaration of used AI tools</b>	<b>57</b>
<b>6</b>	<b>Bibliography</b>	<b>58</b>

# List of Figures

1.1	<b>Schematic overview of structured illumination microscopy (SIM).</b> Adapted from [5] [6], the figure illustrates: (a) Schematic representation of the working principle of Structured Illumination Microscopy as a TIRF-SIM configuration. . . . .	3
1.2	<b>Examples of raw images from microscopy and SIM reconstruction.</b> Acquired raw images of (a) Microtubules and (b) F-actin under low-SNR (left) and high-SNR (middle) conditions, both limited by the Abbe diffraction limit, are compared to the SIM-reconstructed super-resolution images (right) that achieve twice the classical resolution through reconstruction from pattern-modulated data. . . . .	4
1.3	<b>Assessment of structural complexity of different specimens.</b> (a) Representative GT-SIM images of CCPs, ER, MTs, and F-actin. . . . .	8
1.4	<b>The network architecture of CARE.</b> (a) Image reconstruction model (b) surface projection model, as presented in [10]. . . . .	10
1.5	<b>The network architecture of Deep Fourier Channel Attention Network (DFCAN),</b> which utilizes the FCA mechanism in image reconstruction, is adapted from [9]. . . . .	11
1.6	<b>The network architecture of ScU-Net,</b> adapted from [12]. . . . .	12
1.7	<b>The network architecture of RDL-SIM,</b> adapted from [2], generates a SIM reconstruction prediction, uses it to create pattern-modulated images, and takes a noisy input image and pattern-modulated image as input for the final prediction of a denoised high-SNR image. . . . .	14
1.8	<b>The standard architecture of the Projection-Upsampling Network.</b> Adapted from [12], the framework integrates two 3D U-Net-based architectures, followed by an upscaling block that aligns the network output with the ground truth. . . . .	15

<b>3.1 SIM reconstruction superiority of PU-Net on Microtubules.</b>	35
(a) The SIM reconstructions of Microtubules from DFCAN and various PU-Net variants, along with the ground truth image(below input) and the difference image. . . . .	
<b>3.2 SIM reconstruction from low-SNR Microtubules data.</b> Comparison of image reconstruction fidelity of DFCAN vs. . . . . .	36
<b>3.3 Model comparison of DFCAN and PU-Net variants on Microtubules.</b> (a) Model performance comparison on image quantity evaluation metrics. . . . .	37
<b>3.4 SIM Reconstruction superiority of PU-Net on F-actin.</b> (a) The network prediction form DFCAN and PU-Net configurations, with the notable differences pointed by the arrow, demonstrates the superiority of PU-Net ( $9 \times 3 \times 3$ ) in SIM reconstruction. . . . .	39
<b>3.5 SIM reconstruction from low-SNR F-actin data.</b> (a) SIM reconstruction of F-actin from mid-SNR raw data shows that DFCAN and PU-Net variants have similar capabilities. . . . .	40
<b>3.6 Model comparison of DFCAN and PU-Net variants on F-actin.</b> (a) Model performance comparison on image quantity evaluation metrics. . . . .	41
<b>3.7 SIM image reconstruction of our own F-actin data.</b> (a) The notable differences are indicated by the orange marks, showing that PU-Net ( $9 \times 3 \times 3$ ) captures more information than DFCAN. . . . .	42
<b>3.8 Model comparison of low-SNR Microtubules data.</b> (a) Image reconstruction of Microtubules from low-SNR data, using DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL approaches, is illustrated alongside the corresponding SIM reconstruction. . . . .	44
<b>3.9 Model comparison of mid-SNR Microtubules data.</b> (a) Image reconstruction of Microtubules from mid SNR data, using DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL approaches, is illustrated alongside the corresponding SIM reconstruction. . . . .	45
<b>3.10 Training loss comparison of mid-SNR Microtubules data.</b> Comparison of training loss for Microtubules using mid-SNR raw data. 46	
<b>3.11 Model comparison of low-SNR F-actin data.</b> (a) F-actin reconstructions from low-SNR raw data and SIM images for the DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL methods are shown. . . . .	47

3.12	<b>Training loss of mid-SNR F-actin data.</b> Comparison of training loss for F-actin using mid-SNR raw data, highlighting the impact of PU-Net in the loss curve. . . . .	48
3.13	<b>Model comparison of mid-SNR F-actin data.</b> (a) Image reconstruction of F-actin was done using mid-SNR noisy raw data and corresponding SIM reconstruction. . . . .	49
3.14	<b>Model comparison of our own low-SNR F-actin data.</b> (a) Image reconstruction of F-actin from noisy data, featuring DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL outputs. . . . .	51

## **List of Tables**

2.1	Overview of image quantity evaluation metrics.	. . . . .	29
2.2	Summary of Pu-Net variants applied to the MT and F-actin datasets.	. . . . .	32

# 1 Introduction

Imaging biological samples on the cellular and subcellular scales is essential in modern biology. Conventional optical microscopy is fundamentally limited by the diffraction of light, restricting the resolution to roughly 200 nm. Advanced techniques, such as structured illumination microscopy (SIM), overcome this barrier by using periodical sinusoidal illumination patterns to extract spatial information that would otherwise remain unobservable. SIM is especially attractive for live cell imaging due to its requirement of fewer raw images and lower light intensities than alternative super-resolution methods [1].

Achieving high-quality SIM reconstructions critically depends on acquiring images with a high signal-to-noise ratio (SNR). However, several competing factors make this goal difficult to attain in practice. Rapid acquisition rates naturally decrease the average number of photons captured per frame, leading to a lower signal-to-noise ratio (SNR). Increasing excitation intensity to compensate for the risk of concealing information introduces the pitfalls of damaging the samples. Moreover, the intrinsic structural complexity of many biological specimens further degrades SIM performance under low-SNR conditions, as fine features become obscured by noise. Conversely, optimizing SNR through prolonged exposure times or elevated laser power inevitably slows imaging speed and elevates the likelihood of photobleaching. Balancing these trade-offs remains a fundamental challenge in the pursuit of high-fidelity SIM imaging.

Recent advances in deep learning have demonstrated remarkable effectiveness in image reconstruction and denoising tasks. These techniques present a promising solution to the challenges of SIM, though replicating the precise illumination patterns essential for accurate reconstruction remains a significant hurdle for current artificial intelligence approaches. In this paper, we introduce a 3D kernel-based

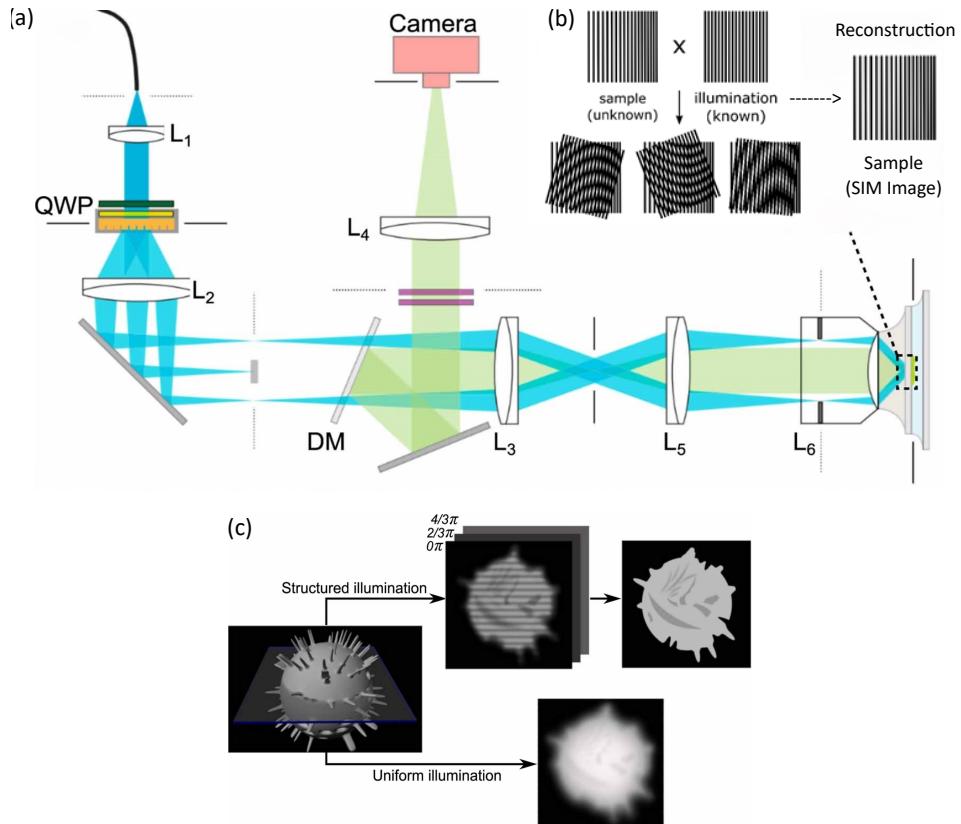
novel projection-upsampling network, a fully supervised deep learning technique to reconstruct the SIM images from noisy low-SNR raw data.

Our ultimate objective is to replace the current DFCAN method in the RDL-SIM pipeline [2] with our PU-Net to recover the lost spatial and temporal resolution from low-SNR images. The proposed framework is designed to be more machine learning centric, straightforward, and user-friendly to the biologist.

## 1.1 Structured Illumination Microscopy

Structured Illumination Microscopy (SIM) is an advanced optical imaging technique that enhances spatial resolution beyond the classical Abbe diffraction limit by exploiting patterned illumination and computational reconstruction algorithms. Unlike conventional wide-field fluorescence microscopy, which is limited [3], SIM uses a known interference pattern, typically sinusoidal, to illuminate the sample. This patterned light interacts with sub-diffraction spatial features of the sample, producing moiré fringes that downshift high spatial frequency information into the detectable passband of the microscope [4].

In practice, the SIM procedure involves several key steps. First, a series of images are captured while illuminating the sample with a structured light pattern at different phases and orientations. Each illumination phase encodes the high-frequency details of the sample in the observable frequency space by mixing it with the known spatial frequencies of the illumination pattern. Mathematically, this process can be described by the convolution of the sample's spatial frequency distribution with the Fourier components of the illumination pattern [4]. By acquiring multiple images with shifted phases and rotated patterns, one can sufficiently sample the frequency space so that these mixtures of frequency components can be disentangled.



**Figure 1.1: Schematic overview of structured illumination microscopy (SIM).** Adapted from [5] [6], the figure illustrates: (a) Schematic representation of the working principle of Structured Illumination Microscopy as a TIRF-SIM configuration. (b) Illustration of the moiré pattern generated through the interference of two coherent light beams, (c) Demonstration of how a stack of 2D SIM images is captured, from the 3D specimen, and detailed information can be recovered using the SIM reconstruction algorithm from images modulated by the sinusoidal patterns in different phases and angles.

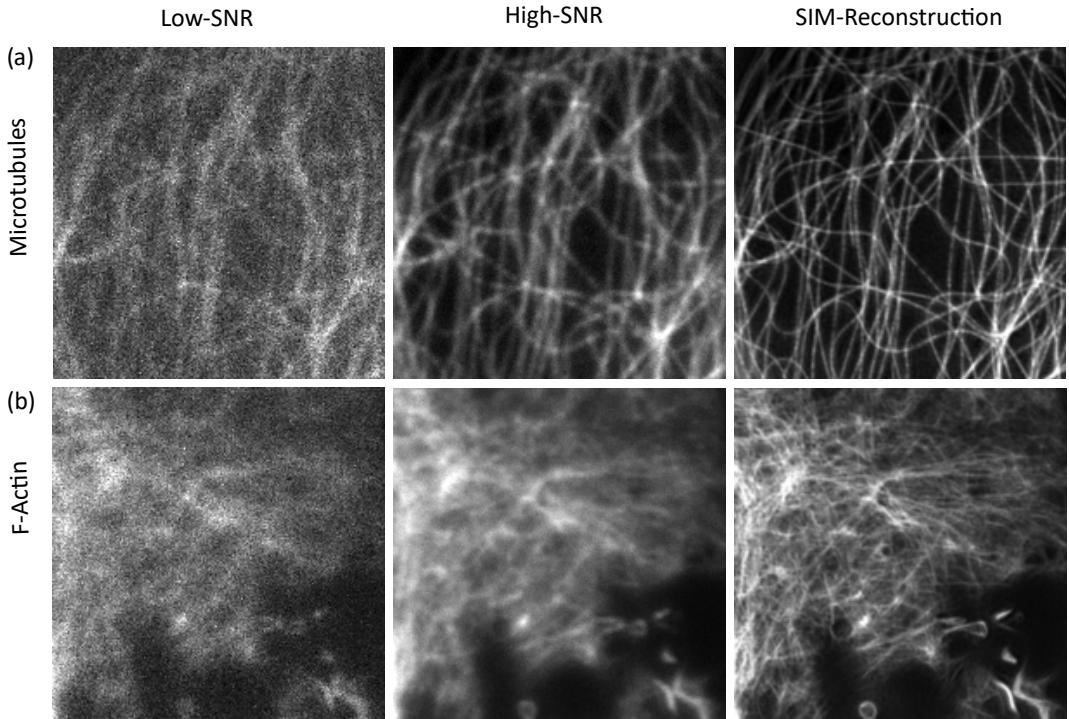


Figure 1.2: **Examples of raw images from microscopy and SIM reconstruction.** Acquired raw images of (a) Microtubules and (b) F-actin under low-SNR (left) and high-SNR (middle) conditions, both limited by the Abbe diffraction limit, are compared to the SIM-reconstructed super-resolution images (right) that achieve twice the classical resolution through reconstruction from pattern-modulated data.

The subsequent computational reconstruction is performed in the Fourier domain. Here, the recorded images are deconvolved to mathematically separate the sample information from the illumination pattern. Through an iterative process involving frequency demodulation and recombination, the high spatial frequencies are repositioned to their correct locations, thereby yielding an image with improved resolution, typically up to twice that of conventional optical microscopy [7].

## 1.2 Implementation of SIM and SIM reconstruction

Structured Illumination Microscopy (SIM) reconstruction is an advanced computational technique designed to enhance image resolution beyond the diffraction limit of conventional light microscopy, which is approximately 200 nm laterally. In conventional microscopy, light emanating from a point source spreads out due to the point spread function (PSF), primarily due to diffraction. When two closely spaced objects emit light that is separated by less than the diffraction limit, their emitted light overlaps, resulting in a blurred region where the two objects are indistinguishable.

SIM addresses this limitation by illuminating the specimen with a series of known sinusoidal illumination patterns, typically using three different angles and three different phases, as shown in (Figure 1.1) [b]. The sinusoidal illumination adds spatial frequency information to the object such that high-frequency details, normally not observable due to the diffraction limit, are modulated into the lower frequency region of the image where they can be detected.

Multiple raw images are acquired under these varying illumination conditions. A specialized reconstruction algorithm then processes these images to extract the high-frequency information that was shifted into the observable frequency range during illumination. By combining this extracted information with the directly acquired low-frequency content, the algorithm synthesizes a final image with enhanced resolution. Thus Figure 1.1 c, the lateral resolution of the reconstructed image is roughly doubled compared to diffraction-limited instruments, achieving a resolution in the range of approximately 100 nm, revealing the fine details of the object.

## 1.3 Biological Specimens

For the scope of this project, we focus only on 2 biological specimens, Microtubules and F-actin, which are important structural components of the cytoskeleton involved in cellular transport and migration. They are highly dynamic structures forming a dense three-dimensional network with structural details smaller than the classical resolution limit. Actin is special in the sense that it forms a quasi two-dimensional network directly above the plasma membrane, which is called the actin cortex. Therefore, its dynamics can be perfectly imaged by TIRF-SIM [8]. The structural complexity of these specimens, measured by the mean gradient shown in Figure 1.3 d, compares the characteristics of the different biological specimens as a benchmark for the SIM performance.

### 1.3.1 Microtubules

Microtubules are hollow, cylindrical, repeating protein filaments formed by the assembly of alpha and beta tubulin heterodimers, and are essential for maintaining cell architecture, intracellular transport, and mitotic spindle formation during cell division [**Microtubules**]. These rigid, hollow cylinders, about 25 nm in diameter, are made up of tubulin protein dimers that form long, spaghetti-like filaments. Unlike other cytoskeletal components, Microtubules are relatively stable, yet they display dynamic instability at their ends, allowing rapid growth or shrinkage. In microscopy images, Microtubules often appear as straight or curved lines emanating from the centrosome. The small diameter in nanometers and the overlapping filaments and sharp edges make it challenging to image and reconstruct. SIM images of this specific biological specimen. As shows in Figure 1.3 a, the structural complexity of Microtubules(MT) has a large mean gradient, proving it significantly challenging for image reconstruction tasks.

### 1.3.2 F-actin

F-actin, the filamentous polymer of globular actin (G-actin) monomers, is a major component of the cytoskeleton that underpins cell shape, mobility, and force generation. The polymerization of G actin into F actin produces dynamic, polarized

filaments with a rapidly growing “barbed” end and a slower “pointed” end that drives processes such as cell migration, endocytosis, and the formation of specialized structures (e.g., lamellipodia and filopodia)[**F-actin**]. The filamentous variant of actin comprises thin, flexible microfilaments approximately 7 nm in diameter that can extend several micrometers in length. The extreme nanoscale dimensions, overlapping filaments, and sharp edges, which can be easily lost by very little noise, present significant challenges for the imaging and reconstruction of SIM data for this biological specimen, specifically for deep learning based methods. As shown in Figure 1.3 a, The structural complexity of F-actin has a larger and wider mean gradient than other specimens, which is introduced by the complex and dense overlapping filaments, making it very difficult to separate this filament in image reconstruction tasks. The Figure 1.2 shows the density and complexity of the F-actin filaments in the image and the SIM recons, compared with Microtubules.

### 1.3.3 Structural Complexity of Different Specimens

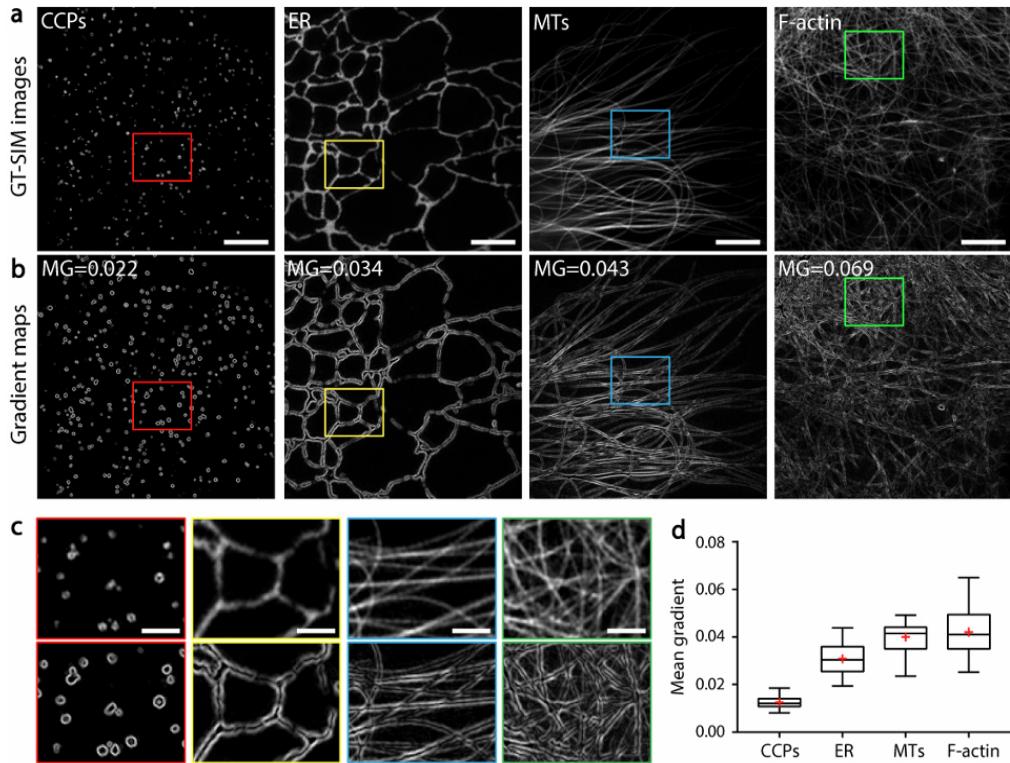
The structural complexity measures the density and intricacy of a specific biological specimen. As presented in [9], for a given biological SIM image  $I$ , consisting of  $M \times N$  pixels, its structural complexity is defined as its grayscale mean gradient (MG):

$$MG(I) = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{\left(\frac{\Delta I}{\Delta x}\right)_{ij}^2 + \left(\frac{\Delta I}{\Delta y}\right)_{ij}^2}{2}} \quad (1.1)$$

Typical gradient maps and corresponding MGs of CCPs, ER, MTs, and F-actin are shown in [2]. For each type of biological specimen, we employed its mean MG (averaged from the MG scores of all its GT-SIM images in the BioSR dataset) to represent its structural complexity. As shown in (1.1), the structural complexity increases from punctate CCPs, reticular ER, and string-like MTs to intricate F-actin.

The mean-gradient method provides a robust metric for quantifying structural complexity. As discussed in Section 1.3.1 and Section 1.3.2, Microtubule filaments measure roughly 25 nm in diameter, whereas F-actin can be as fine as 7 nm. In practice, the noise from the microscope compounds the difficulty of resolving these thinner,

more densely overlapping actin filaments. This contrast between relatively isolated Microtubules and the highly entangled F-actin network represents one of our principal challenges and a rigorous test of the deep-learning approach on such structurally complex and high-density data.



**Figure 1.3: Assessment of structural complexity of different specimens.** (a) Representative GT-SIM images of CCPs, ER, MTs, and F-actin. (b) The corresponding gradient maps to (a). (c) magnified images of the boxed regions from GT-SIM images (upper row) and gradient maps (lower row). Scale bar: 3  $\mu\text{m}$  (a), 1  $\mu\text{m}$  (c). Experiments were repeated with more than 50 images, achieving similar results. d, Box-and-whisker plots of the mean gradients for CCPs ( $n=54$ ), ER ( $n=69$ ), Microtubules ( $n=54$ ), and F-actin ( $n=51$ ). Central line, medians; red cross, mean; limits, 75% and 25%; whiskers, maximum and minimum. This illustration is taken from [9]

## 1.4 Model Architectures

### 1.4.1 Content Aware Image Restoration (CARE)

In the domain of bioimaging, acquiring high signal-to-noise ratio (SNR) images for supervised training is often impractical due to physical and biological constraints. Researchers frequently face trade-offs between imaging speed and spatiotemporal resolution, limiting the quality and quantity of usable training data. The Content-Aware Image Restoration (CARE) framework addresses this challenge by leveraging a deep learning-based approach. Specifically, a symmetric U-Net architecture for both image denoising and joint surface projection tasks.

The denoising model in CARE is based on a modified U-Net architecture that incorporates skip connections between corresponding layers in the encoder and decoder paths. This structure facilitates effective gradient propagation and preserves spatial features during reconstruction. The model performs image downsampling and upsampling to extract multi-scale features, and includes a residual connection from the input to the output layer. This residual path allows high-dimensional features from the input to directly influence the output, improving reconstruction quality and enabling robust denoising even in low-SNR conditions.

For projection tasks, CARE extends the U-Net-based design by employing a chained network architecture. This projection network comprises two cascaded U-Nets connected via skip connections, allowing precise feature localization and effective learning across volumetric inputs. The first U-Net learns to perform a content-aware projection of the 3D volume into a 2D representation, while the second U-Net refines this output through additional denoising and enhancement. The outputs from multiple depths are averaged to produce the final image. This architecture is particularly well-suited for applications such as joint surface projection and super-resolution SIM reconstruction, where high-quality 2D outputs must be derived from 3D datasets acquired under low-light or fast-acquisition conditions.

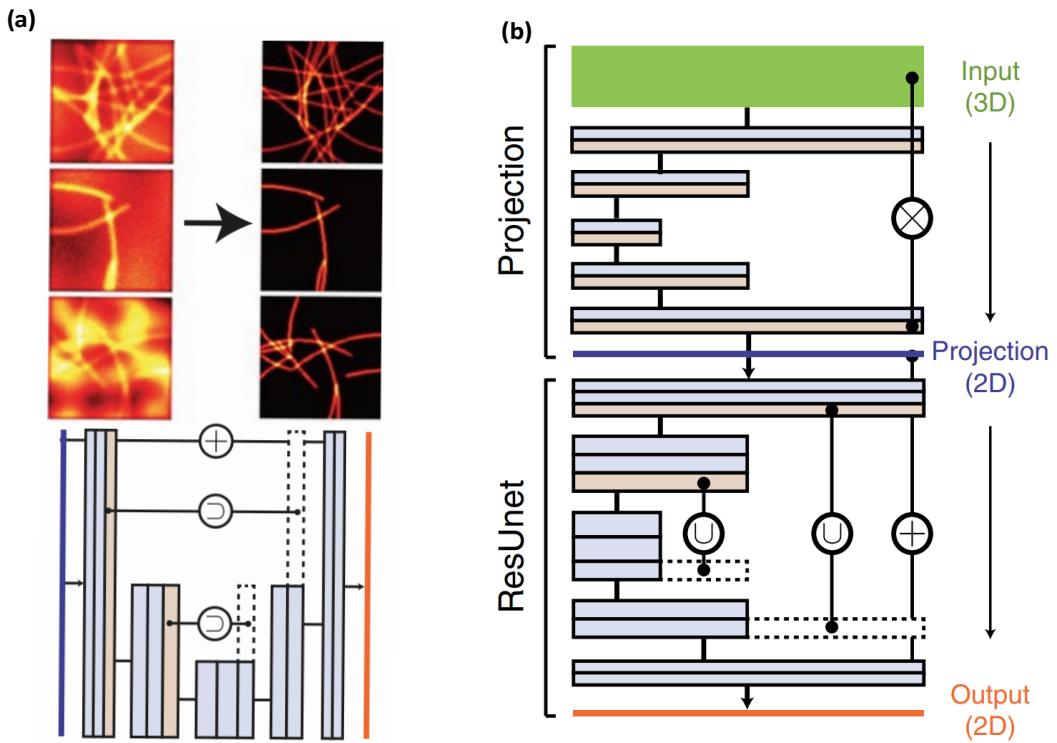


Figure 1.4: **The network architecture of CARE.** (a) Image reconstruction model  
(b) surface projection model, as presented in [10]

The simple design shows that, leveraging the downsampling and upsampling path, connection via skip connection can reconstruct image features from very noisy SNR conditions, producing results that were previously unobtainable, thus pushing the limits of the possible in fluorescence microscopy through machine-learned image computation. [10]

### 1.4.2 Deep Fourier Channel Attention Network (DFCAN)

DFCAN is designed for the reconstruction of a super-resolution image from a noisy, low-SNR image. Instead of just calculating in the spatial domain of the image, it also looks at the image in the Fourier domain. This means it specifically examines the frequency data, where high frequencies are responsible for the fine details, like edges and textures. By focusing on these high-frequency components, DFCAN can restore more accurate details when converting low-resolution images into super-resolution images.

In this method, after extracting the high-level features with a simple layer, it incorporates Fourier Channel Attention Blocks (FCAB), that converts the features into the frequency domain and calculates and enhances the power contribution of high-frequency components (the area that carries the fine details). These extracted features are then carried out via skip connections, keeping the low-frequency information details intact, which is followed by a pixel shuffle layer to merge these feature blocks back into a super-resolution image. [9]

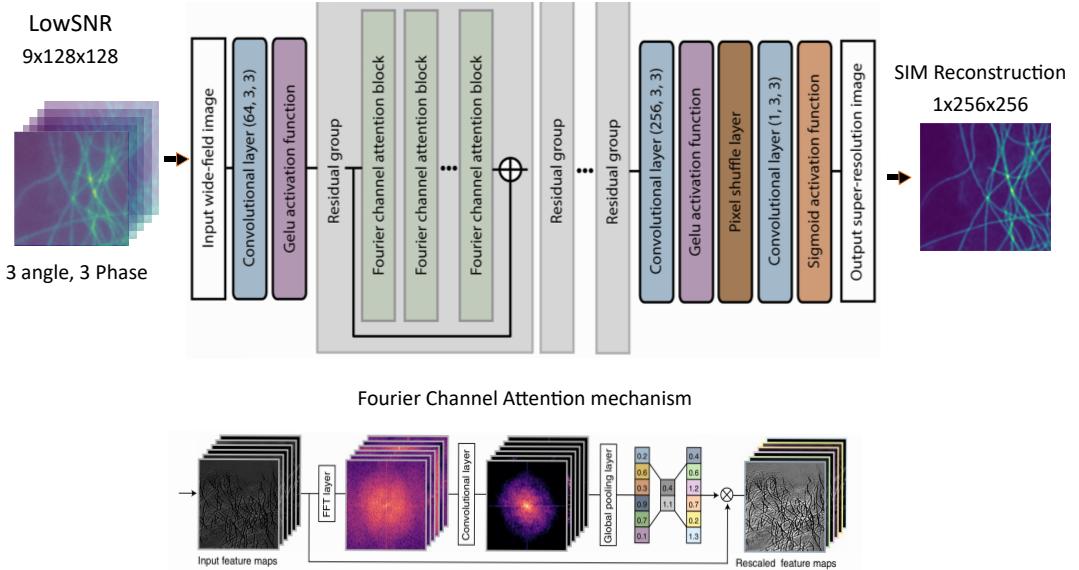


Figure 1.5: The network architecture of Deep Fourier Channel Attention Network (DFCAN), which utilizes the FCA mechanism in image reconstruction, is adapted from [9]

### 1.4.3 ScU-Net

The U-Net architecture has significantly impacted the field of bioimaging since its introduction in May 2015. Its capability for feature localization and effective training on image patches has established it as a preferred method for image segmentation and reconstruction tasks. The fundamental concept of U-Net involves enhancing a conventional contracting network by integrating successive layers, where standard pooling layers are substituted with upsampling layers. This results in a U-like shape, thus giving it the name U-Net. U-Net utilizes an increased number of feature channels in both the downsampling and upsampling parts. Notably, at each down-sampling step, The number of feature channels is doubled, optimizing the model's ability to capture intricate patterns. Furthermore, the architecture incorporates skip connections that transfer features from the 2D max-pooling segment to the 2D up-sampling segment, thereby facilitating improved localization of features throughout the network [11].

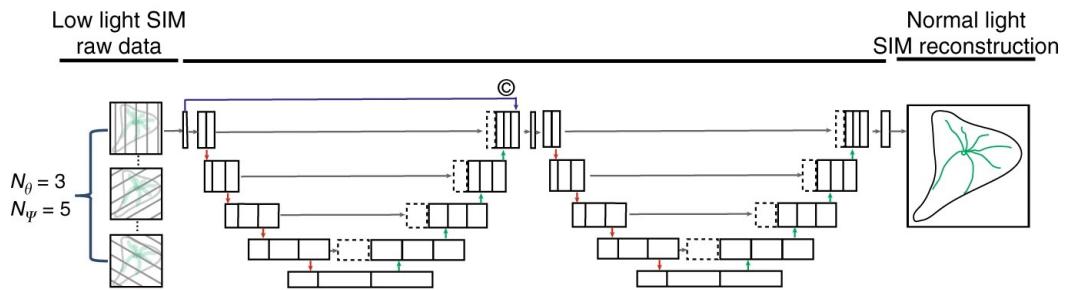


Figure 1.6: **The network architecture of ScU-Net**, adapted from [12]

The scU-Net architecture takes the idea further by chaining two 2D U-Nets together. This design addresses the limitations of a single U-Net, particularly regarding image reconstruction fidelity in low signal-to-noise ratio (SNR) conditions. By utilizing the output of the initial U-Net as input for the second U-Net, the architecture is trained against the ground truth, resulting in enhanced accuracy for reconstructing images affected by low signal-to-noise. This method demonstrates superiority over conventional Structured Illumination Microscopy (SIM) reconstruction techniques, particularly under low-light conditions. [12]

#### **1.4.4 Rationalized Deep Learning Super-Resolution Microscopy (RDL-SIM)**

RDL-SIM is a deep learning-based super-resolution microscopy method that utilizes a convolutional neural network (CNN) to predict a denoised, high-SNR image from a noisy, low-SNR fluorescence images by incorporating the prior knowledge of the modulation and illumination patterns used in structured illumination microscopy (SIM). For low signal-to-noise ratio (SNR) data, conventional SIM reconstruction is severely compromised by noise and artifacts. In this context, RDL-SIM leverages a neural network that is specifically trained to denoise low-SNR images while preserving the high-frequency illumination patterns inherent to SIM. By first applying this denoising procedure, the subsequent SIM reconstruction algorithm operates on images with significantly improved SNR, thereby facilitating the extraction of high-resolution details. The SIM reconstruction is performed on this denoised image can reconstruct high high-resolution SIM image revealing detailed features, which was not possible by deep learning-based SIM reconstruction methods, directly from low-SNR data. [2]

The method introduced 3 branches of the network architecture, primary feature extraction (PFE) branch, Moiré Pattern Extraction (MPE) branch, and Feature Coalescing and Denoising (FCD) branch.

The MPE branch has an important role in SIM reconstruction, as it takes the generated pattern-modulated image as input, where it has prior knowledge about the moiré Patterns. As the illumination patterns and PSF are natural attributes of the SIM system that are independent of observed specimens or available SNR levels in different experiments, these deterministic properties can be used to regulate the denoised raw images in compliance with the physical model of SIM [2]. To generate this pattern-modulated image, the method depends on DFCAN, Section 1.4.2, calculates the modulation from the corresponding high-SNR ground truth images and then uses this as input for the MPE branch of the network.

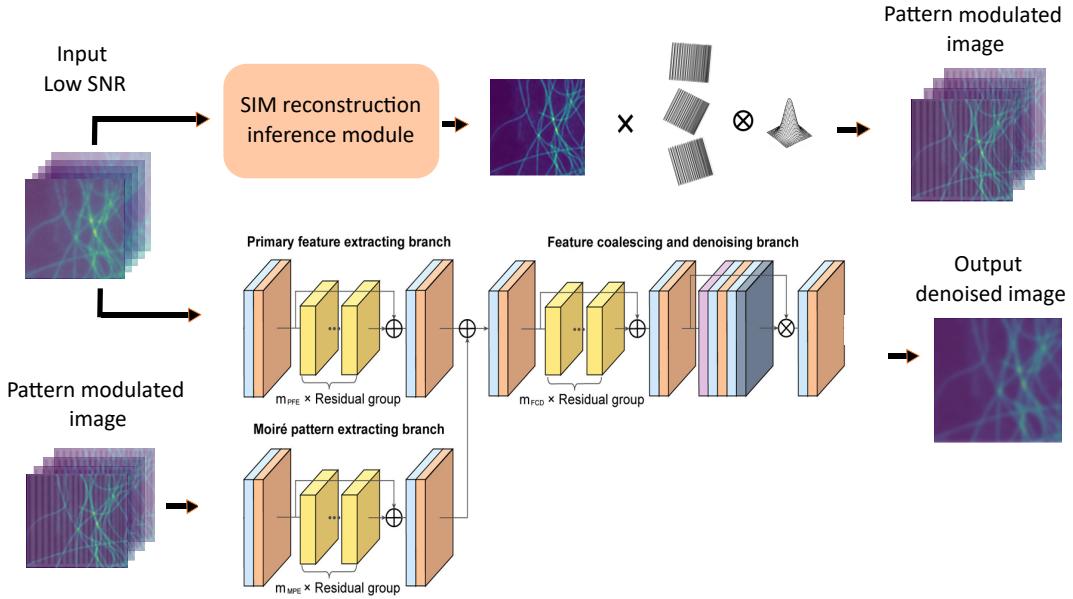


Figure 1.7: **The network architecture of RDL-SIM**, adapted from [2], generates a SIM reconstruction prediction, uses it to create pattern-modulated images, and takes a noisy input image and pattern-modulated image as input for the final prediction of a denoised high-SNR image.

The PFE branch is the first step in the RDL-SIM architecture, where it takes the noisy, low-SNR raw image as input and extracts the primary features.

In the FCD branch, the features from the PFE and MPE branches are fused to form a more comprehensive representation of the input image. The assumption is that the PFE branch captures the general features of the image, while the MPE branch focuses on the specific Moiré patterns. By combining these two sources of information, the FCD branch aims to enhance the overall quality of the image reconstruction. The fused features are then processed through a series of convolutional layers to further refine and denoise the image. This denoising process is crucial for improving the signal-to-noise ratio (SNR) of the final output, making it more suitable for subsequent analysis and interpretation.

### 1.4.5 Projection Upsampling Network (PU-Net)

The U-net architecture is well known in the bioimaging field for its simplicity, reliability, and well-designed architecture. In this paper, inspired by the U-Net design, we propose a chained, 3D kernel-based U-Net architecture, Projection Upsampling network (PU-Net), to explore the deep learning based direct super-resolution SIM reconstruction technique from noisy low-SNR raw data. Conventional SIM typically excites specimens using sinusoidal illumination at various angles and phases, requiring the acquisition of nine images (three angles, three phases) for two-beam SIM and fifteen images (three angles, five phases) for three-beam SIM [13]. In classical SIM reconstruction, all these images are stacked in the channel dimension from different angles and phases, as presented in Section 1.2, which contributes to the final SIM reconstruction image, where the much finer detailed features can be observed, which are hidden in the raw image data.

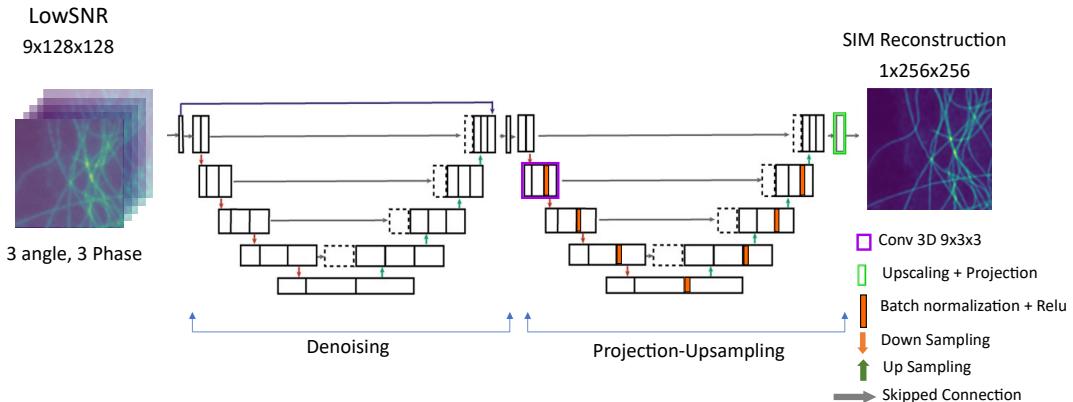


Figure 1.8: **The standard architecture of the Projection-Upsampling Network.** Adapted from [12], the framework integrates two 3D U-Net-based architectures, followed by an upscaling block that aligns the network output with the ground truth.

The concept behind PU-Net is that our imaging data and BioSR were both obtained using a two-beam system, which produces nine-channel images. We sought to see if treating the channel dimension as the z-axis within a three-dimensional convolution framework can potentially gain additional spatial information that might not be captured by traditional two-dimensional kernels.

To evaluate this concept, we developed a Projection Upsampling (PU-Net) network that integrates two 3D U-Net architectures, scaled by an upsampling layer to output predicted super-resolution SIM reconstruction enhanced by two folds in resolution the the input noisy raw image, as illustrated in Figure 1.4.5. In our experimental setup, to test out the capability of capturing spatial-temporal information by incorporating channel dimension as z-axis, we define 3 different PU-Net variants differing by 3 different kernel sizes in the first projection layer only as marked by the purple block in the Figure 1.8. The PU-Net variants, namely PU-Net (3x3x3), PU-Net (7x5x5), and PU-Net (9x3x3), were differentiated by the first projection layer, where the models utilized kernel sizes of 3x3x3, 7x5x5, and 9x3x3, respectively, intuitive by their names, while all subsequent layers across PU-Net variants employed a standard kernel size of 3x3x3.

## 1.5 Aims and Objectives

This thesis aims to explore and enhance deep learning-based image reconstruction methods specifically designed for Structural Illumination Microscopy (SIM), which assess image denoising and reconstruction techniques by leveraging the unique characteristics of moiré patterns. In particular, we will enhance the existing RDL-SIM training pipeline, which is currently reliant on a pre-trained Deep Fourier Channel Attention Network (DFCAN), to incorporate the prior knowledge of the moiré patterns into the network. We aim to replace this DFCAN with our own PU-Net, integrate it with the current workflow, and study its advantages.

In our approach, we hypothesize that our 3D kernel-based projection upsampling network (PU-Net) will achieve significantly enhanced reconstruction fidelity by capturing spatial correlations across channel dimensions, thereby outperforming existing 2D kernel-based methods for SIM reconstruction.

Our research will focus on the Microtubule and F-actin specimens, which are drawn from the publicly available BioSR dataset (see Section 2.1.1). Both specimen types are highly detailed and structurally complex, particularly the dense F-actin network described in Section 1.3.3, which pushes deep-learning methods to their limits in terms of resolution enhancement and image reconstruction. In addition, we will determine the minimum signal-to-noise ratio (SNR) required to produce reliable, artifact-free SIM reconstructions under low-light imaging conditions.

## **Objectives:**

- 1. Architectural comparison:** Design and implement three variants of the PU-Net, namely PU-Net ( $3 \times 3 \times 3$ ), PU-Net ( $7 \times 5 \times 5$ ), and PU-Net ( $9 \times 3 \times 3$ ), with a difference in the varying kernel size of the first convolution layer only. The goal is to systematically integrate information by enhancing spatial resolution using a larger kernel size along the channel dimension (3 angle, 3 phase) of SIM images to study its impact on the image reconstruction performance.
- 2. Parameter optimization:** Hyperparameter optimization and fine-tuning the PU-Net variant on the Microtubules and F-actin datasets to identify optimal hyperparameters for the SIM reconstruction tasks.
- 3. Benchmarking SIM reconstruction:** Train and validate the PU-Net variants and DFCAN using the same training and validation set for each biological specimen, and compare the outcome employing image qualitative methods such as decorrelation analysis, structural similarity, normalized root mean square error, and others.
- 4. PU-Net integration:** Integrate the best-performing PU-Net model into the RDL-SIM pipeline. Separately train and evaluate (a) PU-Net+RDL-SIM and (b) DFCAN+RDL-SIM on each biological specimen on each low and mid signal-to-noise image separately.
- 5. Determining SNR threshold:** Evaluate the reconstruction performance from the separately trained RDL-SIM pipelines and determine the feasibility of generating artifact-free SIM reconstruction at each SNR level for each biological specimen.
- 6. Test our own F-actin dataset:** Apply the optimized PU-Net+RDL-SIM and DFCAN+RDL-SIM workflow to our independently acquired F-actin dataset to identify any specific challenges regarding dataset acquisition and opportunities for further model generalization.

## 2 Methods

### 2.1 Dataset

The efficacy of any deep learning architecture is inherently tied to the quality and relevance of its underlying data. In the absence of a quality dataset, it is impossible to optimize the model toward the desired outputs. Accordingly, the following section provides a comprehensive overview of the dataset employed, the pre-processing protocols applied, and the structured workflow developed for the generation of training samples.

#### 2.1.1 BioSR

The BioSR dataset is a super-resolution image microscopy dataset that consists of five biological structures (CCPs, ER, MTs, F-actin, Myosin-IIA) [14]. In the present study, we focus primarily on Microtubules and F-actin. The microtubule component of the dataset includes 55 cellular images captured at nine different signal-to-noise levels, each accompanied by a paired high signal-to-noise image and its corresponding super-resolved SIM image. Similarly, the F-actin subset consists of 51 cellular images acquired at 12 distinct signal-to-noise levels, with each image paired with high signal-to-noise data and the derived SIM reconstruction. For both biological specimens, the original image stack has a region of interest (ROI) of dimensions  $502 \times 502 \times 9$ , consisting of 3 phases and 3 angles, while the super-resolved SIM image is provided, which is two-fold improved resolution of  $1004 \times 1004$  pixels.

#### 2.1.2 Acquisition of our F-actin Dataset

Our F-actin dataset comprises 51 cellular image stacks (each of size  $300 \times 300 \times 9$ ) acquired under single low signal-to-noise ratio (SNR) conditions, with corresponding high-SNR ground-truth images. Super-resolution SIM ground-truth reconstructions were generated using the publicly available ImageJ fairSIM plugin. Detailed descriptions of the sample preparation and image-acquisition protocols can be found in Winkelmann et al. (2024) [8].

### 2.1.3 Training Data Generation for SIM Reconstruction Module

Firstly, we train our PU-Net model for direct SIM reconstruction from noisy low-SNR images. The generated training data consists of 8 noise levels for Microtubules and 9 different noise levels for the F-actin dataset. We randomly select 40 image cells for training data generation, 7 for validation, and 5 for testing. From each image cell, we create a patch of 128x128x9 from different cell regions, with a total number of 15000 image patches for training and 4500 image patches for validation.

Image normalization and background filtering are critical pre-processing steps in the generation of eligible training data from microscopic images. ensuring consistency in intensity levels across samples and improving overall model performance. For normalization, we use a percentile-based approach that adjusts image intensity using  $p_{low} \in (1, 3)$  and  $p_{high} \in (99.5, 99.9)$ , thereby ensuring robust normalization that reduces the influence of outliers and hot spots commonly known in microscopy images, while preserving the essential structure of the image data. While creating the image patches, we keep selecting the regions with reduced background information with a threshold of 0.4 to make sure the network trains on image patches where maximum information is available.

As very little training data is available in the BioSR dataset for each specimen of cells, we rely on heavy image augmentation. As in biomedical imaging, deformation and rotation are the most common variations, we rely on different affine transformations of data, random crops, and scaling. The data augmentation optimizes the network for desired invariances, thus making it robust in hyperplanes. The necessity of data augmentation for learning invariances has been well-experimented in Dosovitskiy et al [15]. Moreover, to make sure the PU-Net learns the features more robustly, instead of doing the data augmentation when generating the training data, we use the Data Wrapper function inside the training function to create augmented data for each batch on the fly. This makes sure that in each batch, the network sees the augmented variation of the same data, thus making the training process more effective.

Secondly, we train a DFCAN network for the prediction of SIM reconstruction. For DFCAN, we generate the data for both biological specimens, Microtubules and F-actin, using the BioSR dataset similarly as mentioned above, except for the data augmentation part. To augment the data, we use random flip, rotation, and translation, which are then saved on the disk with the original data and fed into the model as a whole dataset, following the training data generation process of DFCAN. [9]

Thirdly, for training both deep learning networks, PU-Net and DFCAN, on our own F-actin dataset, we follow the same procedure described above.

#### **2.1.4 Training Data Generation for Image Denoising Module**

For the training of the models on two different biological specimens separately, we produce 2 different training datasets for each specimen of low and mid SNR levels with respect to its high-SNR data. We divide our training data into 40 cells for training, 7 cells for validation, and 6 cells for testing randomly. For training, from each cell we create 128x128x9 patches from different regions of the image, normalized and filtered by the background as described in Section 2.1.3, further augmented by random flip, rotation, and translation, producing 5500 patches for training and saved on the disk with the original dataset. We use the same training and validation dataset for both DFCAN + RDL and PU-Net (9x3x3) + RDL for model training and performance comparison.

## **2.2 quantitative assessment of image reconstruction**

To evaluate the predicted reconstructed images from the network with the ground truth pair, we use various full-reference images measurement quality techniques such as MSE, NRMSE, PSNR, SSIM, and MS-SSIM. Only the method Decorrelation analysis is a parameter-free resolution test that does not require a reference image.

### **2.2.1 Mean Squared Error**

MSE measures the average squared difference between pixel values of the ground truth and the predicted images. It is calculated as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.1)$$

Where:

- $N$  is the number of pixels in the image.
- $y_i$  is the pixel value of the ground truth image at pixel  $i$ .
- $\hat{y}_i$  is the pixel value of the predicted image at pixel  $i$ .

However, it has limitations in reflecting perceived visual quality, particularly for microscopy images rich in structural details, as it focuses solely on pixel-wise differences. The lower the MSE value is, the better the image reconstruction. This Method is very sensitive to illumination and does not consider the underlying characteristic of the image itself. For example, if the image is higher brightness, though being a better reconstruction, it will yield a higher MSE value.

### 2.2.2 Normalized Root Mean Square Error

NRMSE enhances the Root Mean Squared Error (RMSE) by normalizing it with the image's intensity range, enabling comparisons across datasets with similar scales. It is expressed as:

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\max(y) - \min(y)} \quad (2.2)$$

Where:

- $N$  is the number of pixels in the image.
- $y_i$  is the pixel value of the ground truth image at pixel  $i$ .
- $\hat{y}_i$  is the pixel value of the predicted image at pixel  $i$ .
- $\max(y)$  and  $\min(y)$  are the maximum and minimum pixel values in the ground truth image, respectively.

NRMSE is a more robust metric than MSE, as it accounts for the image's intensity range, making it suitable for comparing images with different brightness levels. It is indeed useful for measuring shifts in pixel intensity because it directly compares the pixel values between two images. If the intensity of pixels changes (e.g., due to brightening, darkening, or other uniform shifts), this will increase the difference between  $I_1(i)$  and  $I_2(i)$ , leading to a higher NRMSE. In this sense, it effectively captures overall mismatches in pixel intensity across the image.

In microscopy images, NRMSE gives insight into the deviation against the dynamic range of the reference image. NRMSE provides a scale-invariant measure of how closely a reconstructed dataset reproduces ground-truth features, including subtle pixel shifts and intensity fluctuations. Its sensitivity to even minor misalignment makes it an indispensable tool for validating algorithmic corrections.

### 2.2.3 Peak Signal-to-Noise Ratio

The PSNR is a signal processing measurement that compares a given received or processed signal to its source signal. This comparison allows us to quantify how much a processed signal is faithful to the original, also allowing us to identify possible noise or distortions to the signal [16]. PSNR is a widely used metric to evaluate the quality of a processed image (e.g., reconstructed, compressed, or enhanced) compared to its original (ground truth) version. It measures the ratio of the maximum possible signal strength to the noise introduced by the processing, expressed in decibels (dB). Essentially, PSNR tells you how much distortion or noise is present in the processed image relative to the clean original. Higher PSNR values indicate better image quality. PSNR is purely pixel-based and doesn't account for how humans perceive quality. Two images with identical PSNR values might look very different if errors are distributed unevenly or affect perceptually important areas (e.g., edges in microscopy images).

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (2.3)$$

where:

- $\text{MAX}_I$  is the maximum possible pixel value of the image ( $\max(I)-\min(I)$ ).
- MSE is the mean squared error between the original and processed images.

In the context of microscopy, PSNR is particularly effective for evaluating the fidelity of CNN-generated images in tasks such as denoising or super-resolution, where signal clarity is paramount [10]. Higher PSNR is considered better image reconstruction quality. This image evaluation method is purely pixel-based and doesn't account for how humans perceive quality. Two images with identical PSNR values might look very different if errors are distributed unevenly or affect perceptually important areas (e.g., edges in microscopy images).

#### 2.2.4 Structural Similarity Index (SSIM)

SSIM evaluates image quality by mimicking human visual perception, focusing on luminance, contrast, and structural similarities. [17] It is defined as:

$$\text{SSIM}(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) \quad (2.4)$$

Where:

- $l(x, y)$  is the luminance comparison function, quantifying the brightness similarity between images  $x$  and  $y$ .
- $c(x, y)$  is the contrast comparison function, measuring the contrast similarity of the two images.
- $s(x, y)$  is the structure comparison function, evaluating the structural information contained within images  $x$  and  $y$ .

The individual components of the SSIM can be articulated mathematically as follows:

##### 1. Luminance Function:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.5)$$

Where:

- $\mu_x$  and  $\mu_y$  denote the average pixel values of images  $x$  and  $y$ , respectively.
- $C_1$  is a constant introduced to mitigate the effect of weak denominators.

## 2. Contrast Function:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.6)$$

Where:

- $\sigma_x^2$  represents the variance of image  $x$ .
- $\sigma_y^2$  represents the variance of image  $y$ .
- $C_2$  is an additional constant that stabilizes the division.

## 3. Structure Function:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (2.7)$$

Where:

- $\sigma_{xy}$  indicates the covariance between images  $x$  and  $y$ .
- $C_3$  is a constant that serves to prevent potential division by zero.

The constants  $C_1$ ,  $C_2$ , and  $C_3$  are critical for ensuring numerical stability, especially in cases where the denominators may approach zero. Typically, these constants are defined as follows:

$$C_1 = (K_1 L)^2, \quad C_2 = (K_2 L)^2$$

- $L$  denotes the dynamic range of the pixel values.
- The commonly used values for the constants are  $K_1 = 0.01$  and  $K_2 = 0.03$ .
- These constants are established to mitigate issues related to numerical instability.

SSIM evaluates how well the processed image preserves the visual structure of the original, considering human perception. It's sensitive to distortions like blurring or artifacts that might not drastically change pixel values (and thus PSNR) but alter the image's appearance. For microscope-reconstructed images, SSIM can highlight how well structural details like cell edges or tissue patterns are preserved, which is

often more scientifically relevant than raw pixel accuracy. It complements PSNR by focusing on visual fidelity. For example, A slightly blurred image might have a high PSNR (small pixel errors) but a lower SSIM due to lost structural detail. SSIM is measured between 0 and 1, where 1 is perfect similarity and 0 means no similarity at all.

### 2.2.5 Multi-Scale Structural Similarity Index (MS-SSIM)

MS-SSIM addresses the limitation of SSIM and takes the idea even further. SSIM index algorithm introduced in [17] is a single-scale approach. MS-SSIM considers this drawback of the method because the right scale depends on viewing conditions (e.g., display resolution and viewing distance). MS-SSIM provides more flexibility than a single-scale approach in incorporating the variations of image resolution and viewing conditions [18]. The overall MS-SSIM evaluation is obtained by combining the measurements at different scales using :

$$\text{MS-SSIM}(x, y) = \prod_{j=1}^J [l_j(x, y)]^{\alpha_j} \cdot [c_j(x, y)]^{\beta_j} \cdot [s_j(x, y)]^{\gamma_j} \quad (2.8)$$

Where:

- $J$  is the number of scales.
- $\alpha_j$ ,  $\beta_j$ , and  $\gamma_j$  are the weights assigned to the luminance, contrast, and structure components at scale  $j$ .
- The product is taken over all scales  $j$ .
- The weights  $\alpha_j$ ,  $\beta_j$ , and  $\gamma_j$  are typically chosen to sum to 1, ensuring that the overall MS-SSIM value is a weighted combination of the contributions from each scale.
- The weights can be adjusted based on the specific application or the characteristics of the images being compared.

MS-SSIM is particularly useful in applications where images may be viewed at different scales or resolutions, such as in Biomedical imaging.

## 2.2.6 Decorrelation Analysis

Decorrelation analysis assessing the resolution of individual images based on image partial phase autocorrelation. The method is a parameter-free estimation and does not require a reference image. The main algorithm is divided into 2 steps. Firstly, the algorithm computes the Fourier transform of an edge-apodized image, normalizes it, and determines its Pearson correlation with the original image. In the next step, the algorithm applies a variable-radius binary circular mask to the normalized transform and recalculates Pearson correlations to produce the resolution function  $d(r)$  .[19] Which is calculated as follows:

$$d(r) = \frac{\int \text{Re}\{I(k)I_n^*(k)M(k; r)\} dk_x dk_y}{\sqrt{\int |I(k)|^2 dk_x dk_y \int |I_n(k)M(k; r)|^2 dk_x dk_y}} \quad (2.9)$$

Where:

- $I(k)$  is the Fourier transform of the image.
- $I_n(k)$  is the Fourier transform of the normalized image.
- $M(k; r)$  is the binary circular mask with radius  $r$ .
- $\text{Re}\{\cdot\}$  denotes the real part of a complex number.
- $*$  denotes the complex conjugate.
- $dk_x$  and  $dk_y$  are the differential elements in the Fourier domain.

We use the published fairSIM ImageJ plugin [20] of this application for calculating the resolution, we use the default parameter setting, which are  $N_r = 50$  and  $N_g = 10$  with a radius of  $r_{min} = 0$  and  $r_{max} = 10$ , as described in the basic user manual of the plugin. [19]

### **2.2.7 FairSIM**

FairSIM is a SIM Reconstruction plugin for Structural Illumination Microscopy for ImageJ. The SIM Reconstruction is performed numerically on the acquired wide-field data from the microscope, these algorithms allow to enhance the resolution by two fold in comparison with the corresponding acquired image. [20] This application can work on two-beam interference, which is utilized by many home-built and total internal reflection-excited fluorescence (TIRF)-based systems. It is based on the well-established SIM illumination technique introduced by Gustaffson and Heintzmann [4] [21], and the corresponding reconstruction algorithms.

This application provides the biologist with the flexibility to approximate the parameter and the OTF automatically. With advanced options to correct phase and angles of modulation, thus enhancing the reconstructed SIM image. The reliability of the SIM reconstruction and the user-friendliness make it the preferred choice for this project.

## 2.2.8 Quantitative assessment summary

Table 2.1: Overview of image quantity evaluation metrics.

Metric	What it Measures?	Limitations
MSE	Quantifies the average squared difference between pixel values of the reference and distorted images; serves as a global measurement.	Highly sensitive to variations in illumination and outliers; do not account for spatial relationships and structures within the image.
NRMSE	Illumination invariant; useful for quantifying pixel-level shifts.	Despite normalization, it is still predominantly pixel-based; it overlooks local structures and features.
PSNR	evaluates the ratio of the maximum possible signal power to the noise power, indicating reconstruction quality relative to the original image.	It does not consistently correlate with human perceptual quality; local distortions may not be adequately captured by this metric.
SSIM	Assesses the image quantitatively by comparing local luminance, contrast, and structural information, thereby offering a measure that is more aligned with human visual perception.	The metric is sensitive to the choice of analysis window size, and it can be adversely affected by shifts in brightness; it is not invariant to changes in scale.
MS-SSIM	Extends SSIM by capturing both fine and coarse details across multiple scales, thereby providing a comprehensive assessment of perceptual quality.	The computational demands are notably high, and the global score obtained may obscure significant local errors that are critical to assess.
Decorrelation Analysis	Utilizes Fourier masking to estimate image resolution based on the statistical decorrelation of frequencies.	does not consider information of the ground truth image; yields a single global value; relies heavily on the accuracy of noise estimation

## 2.3 Training Details

### 2.3.1 Training Details of SIM Reconstruction Module

In this project, to reach our objectives as described in Section 1.5, for the task of SIM reconstruction from noisy raw data, we train two different biological specimens with DFCAN and PU-Net variants on the same training and validation dataset to ensure the evaluation of model performance.

In the Projection-Upsampling Network experiment, we evaluated three distinct PU-Net architectures, denoted by PU-Net ( $3 \times 3 \times 3$ ), PU-Net ( $7 \times 5 \times 5$ ), and PU-Net ( $9 \times 3 \times 3$ ), as detailed in Section 1.4.5. Each three-dimensional convolutional layer incorporates an  $\ell_2$  kernel regularizer, immediately followed by batch normalization and a ReLU activation (illustrated by the orange blocks in Figure 1.8), thereby promoting stable gradient propagation and mitigating overfitting. Model parameters were optimized via stochastic gradient descent using the Adam optimizer, with hyperparameters set to  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . In each training epoch, the mini-batch iterations were set by the ratio of total training samples to the mini-batch size. To enhance generalization, the training data was channeled via an augmentation pipeline to provide the model with augmented data on each mini batch iteration on the fly as described in Section 2.1.3. In our training process, we initialize MSE and MAE as our loss functions defined as:

$$\text{MSE} = \frac{1}{H \times W \times Z} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^Z (y_{ijk} - \hat{y}_{ijk})^2 \quad (2.10)$$

$$\text{MAE} = \frac{1}{H \times W \times Z} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^Z |y_{ijk} - \hat{y}_{ijk}| \quad (2.11)$$

- $H$  is the height of the input.
- $W$  is the width of the input.
- $Z$  is the depth of the input.
- $y_{ijk}$  is the ground truth  $(i, j, k)$ .
- $\hat{y}_{ijk}$  is the prediction  $(i, j, k)$ .

We performed two consecutive hyperparameter explorations to identify configurations that minimize training loss. In the initial sweep, we systematically varied batch size, network depth, the number of filters in the first convolutional layer, the loss functions, and quantified their impact on the loss curve. Statistical analysis revealed that larger batch sizes, increased model depth, and a higher filter count in the first layer each exhibited a strong, negative correlation with training loss, meaning they consistently yielded lower loss values. Leveraging these findings, we designed a second, more focused sweep to determine the most promising hyperparameter ranges. This refined search not only accelerated convergence but also produced markedly smoother loss trajectories, demonstrating more efficient and stable learning, Figure 3.3 b for Microtubules, and Figure 3.6 b for F-actin dataset. All training runs were conducted on a high-performance computing cluster equipped with an NVIDIA A100 GPU (80 GB HBM2e) and 500 GB of system RAM. Each full training required approximately five hours to complete. The trained parameters of our experiment are documented in Table 2.2.

For training the DFCAN model, we use the same training parameters as described in the paper [9], with a batch size of 4, the learning rate of 0.0001, and the loss function, which is a combination of MSE and SSIM weighted by a factor of 0.1. The loss function is described as,

$$Loss = \text{MSE} + 0.1 (1 - \text{SSIM}) \quad (2.12)$$

Below, we have listed the summary of training parameters for the different PU-Net variants on specific biological specimens used in our experiments.

Table 2.2: Summary of Pu-Net variants applied to the MT and F-actin datasets.

Parameter	PU-Net (3x3x3)	PU-Net (7x5x5)	PU-Net (9x3x3)	PU-Net (3x3x3)	PU-Net (7x5x5)	PU-Net (9x3x3)
Dataset	MT	MT	MT	F-actin	F-actin	F-actin
Patch Size	128x128x9	128x128x9	128x128x9	128x128x9	128x128x9	128x128x9
U-Net Depth	4	2	3	4	2	4
U-Net First (filters)	64	48	32	64	48	84
U-Net Kernel	3	3	3	3	3	3
Residual Block	True	True	True	True	True	True
Proj. Kernel (1st-layer)	3x3x3	7x5x5	9x3x3	3x3x3	7x5x5	9x3x3
Proj. Kernel (block)	3x3x3	3x3x3	3x3x3	3x3x3	3x3x3	3x3x3
Proj. Depth	3	2	2	3	2	2
Proj. Conv/Depth	2	3	2	2	3	3
Batch Size	16	32	16	16	32	128
Learning Rate	0.0004	0.004	0.0004	0.0004	0.004	0.0004
Upsampling Factor	2	2	2	2	2	2
Loss Function	MSE	MSE	MSE	MSE	MSE	MSE
Batch Norm	True	True	True	True	True	True
Kernel Regularizer	$l_2(0.0003)$	$l_2(0.0003)$	$l_2(0.0003)$	$l_2(0.0003)$	$l_2(0.0003)$	$l_2(0.0003)$
Total Params.	22,068,354	1,749,938	1,485,506	22,068,354	1,749,938	38,010,686

### 2.3.2 Training Details of Image Denoising Module

As per our objective, PU-Net integration, we train two biological specimens distinguished by low and mid SNR data to test the image reconstruction fidelity of the model based on RDL-SIM with DFCAN, and to obtain our goal of replacing the DFCAN with our proposed model, we train the RDL-SIM pipeline employed with our PU-Net (9x3x3). To ensure the fairness of the training comparisons, we always use the same training and validation dataset for both models employed by the low and mid SNR noisy raw data.

First, we train the RDL-SIM model using the DFCAN module, the same training protocol outlined in [2]. Specifically, the training is conducted with a batch size of 3, a ReLU coefficient of 0.1, and a learning rate of 0.0001. The depths of the branches are set accordingly, while the parameters for moiré pattern extraction (MPE), primary feature extraction (PFE), and feature coalescing and denoising (FCD) are assigned values of 5, 2, and 5, respectively. The loss function is a composite of the mean squared error (MSE) and the structural similarity index measure (SSIM), weighted by a factor of 0.1.

$$Loss = \text{MSE} + 0.1(1 - \text{SSIM}) \quad (2.13)$$

Secondly, we tested our proposed PU-Net integrated with the RDL-SIM pipeline, and we trained a separate RDL-SIM model to replace the original DFCAN module with our pre-trained PU-Net(9x3x3). As a training parameter, we use the same parameter as described in the paper [2]. Under these conditions, we perform two experiments for each specimen by replacing the original SIM reconstruction module based on DFCAN with a PU-Net configuration (9×3×3). To analyze the result, we used 2 different biological specimens, Microtubules and F-actin from the BioSR dataset [14].

In our experiments, we evaluate the performance of our proposed training pipeline using 2 different biological specimen datasets under varying SNR conditions. Specifically, for the Microtubules dataset, each network DFCAN + RDL and PU-Net (9×3×3) + RDL is trained using either low-SNR or mid-SNR raw data as inputs, with the corresponding high-SNR raw data serving as the ground truth. Similarly, for the F-actin dataset, separate networks are trained for low-SNR and mid-SNR inputs, again using high-SNR raw data as the ground truth. This unified approach across different specimens facilitates a comprehensive comparison of network performance under diverse noise conditions.

# 3 Results

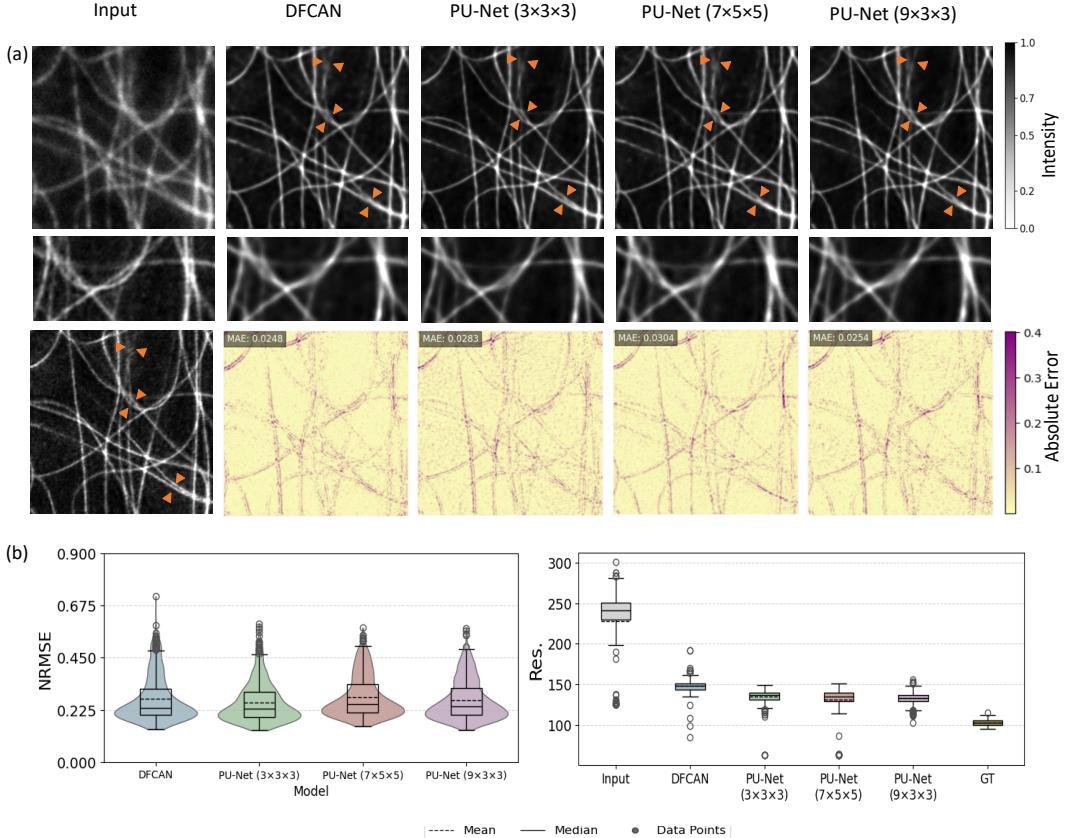
We structured our study around three main aims. First, we trained DFCAN and PU-Net variants to directly reconstruct super-resolution SIM images from noisy, low-SNR raw data. Second, we integrated this pre-trained network within the current RDL-SIM denoising pipeline to denoise noisy low-SNR inputs into noise-free high-SNR image reconstructions. Finally, we process the denoised outputs with the ImageJ fairSIM plugin to produce super-resolved SIM reconstruction images and identify the lowest SNR threshold for artifact-free SIM reconstructions. The results of these experiments are summarized below.

## 3.1 SIM Reconstruction Module

In this section, we present the result of our third objective, the Benchmarking of SIM reconstruction models.

### Mictorubules Assessment

In our experiments with the SIM reconstruction module, we assess the performance of Microtubules using DFCAN and various variants of PU-Net. As shown in Figure 3.1 a, indicated by the orange arrow, PU-Net is more effective at reconstructing detailed features of the filaments, while DFCAN struggles to distinguish these features clearly. This finding is further confirmed by a quantitative resolution analysis in Figure 3.1 b, where the PU-Net variant exhibits a notably lower score, indicating higher resolution compared to DFCAN.



**Figure 3.1: SIM reconstruction superiority of PU-Net on Microtubules.** (a) The SIM reconstructions of Microtubules from DFCAN and various PU-Net variants, along with the ground truth image(below input) and the difference image. The areas where the network predictions significantly differ from those of DFCAN and the PU-Net variants are indicated by the arrow. (b) The median value of NRMSR is lower for PU-Net (3x3x3) and PU-Net (9x3x3). The decorrelation analysis plot demonstrates the robustness of PU-Net, showing a noticeable improvement in resolution.

It is important to note that when dealing with highly noisy data, all network predictions can exhibit significant artifacts and false structure, as shown in Figure 3.2. This makes it difficult to perform a meaningful decorrelation analysis. Therefore, in our resolution analysis, we limit our comparison to images with a resolution greater than 95 nm, as achieved by PU-Net (9x3x3). This approach ensures a fair and reliable assessment of the model predictions.

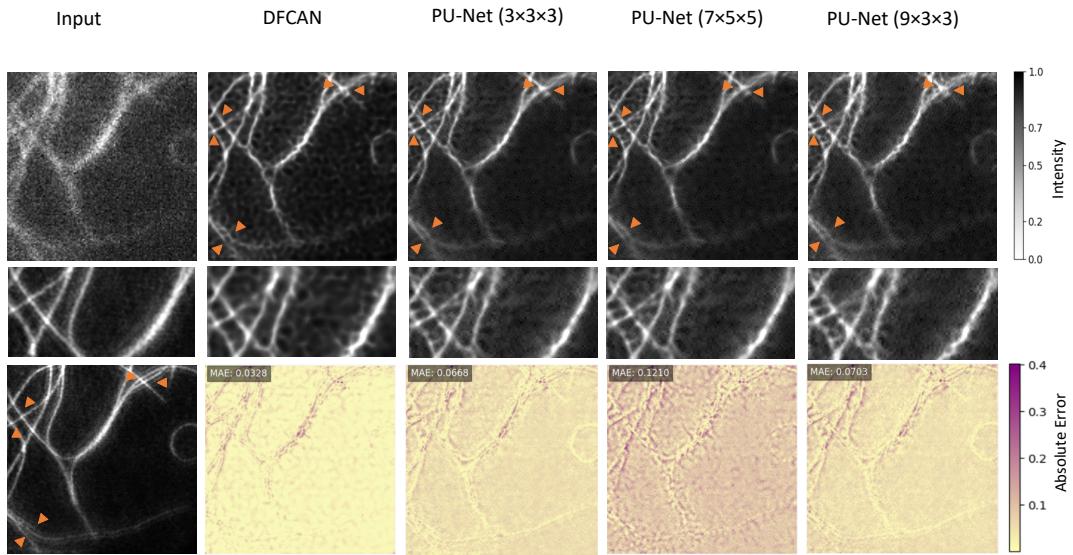
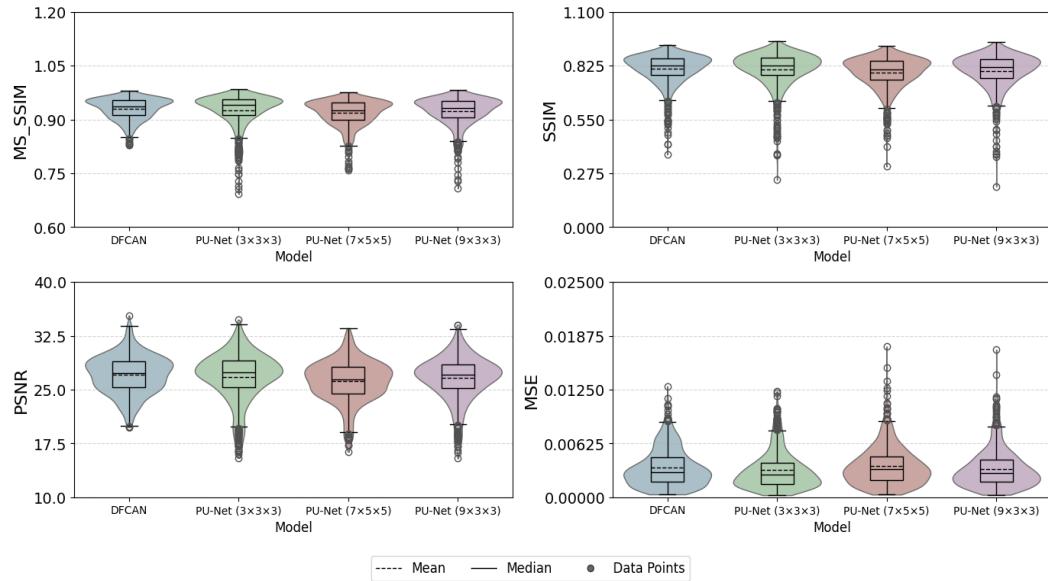


Figure 3.2: **SIM reconstruction from low-SNR Microtubules data.** Comparison of image reconstruction fidelity of DFCAN vs. Pu-Net variants from very noisy, low-SNR data, which is more susceptible to the development of artifacts.

(a)



(b)

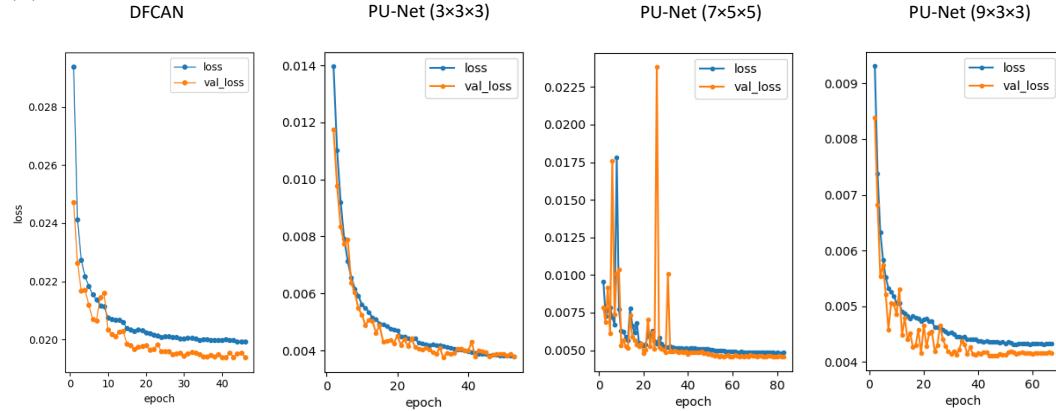
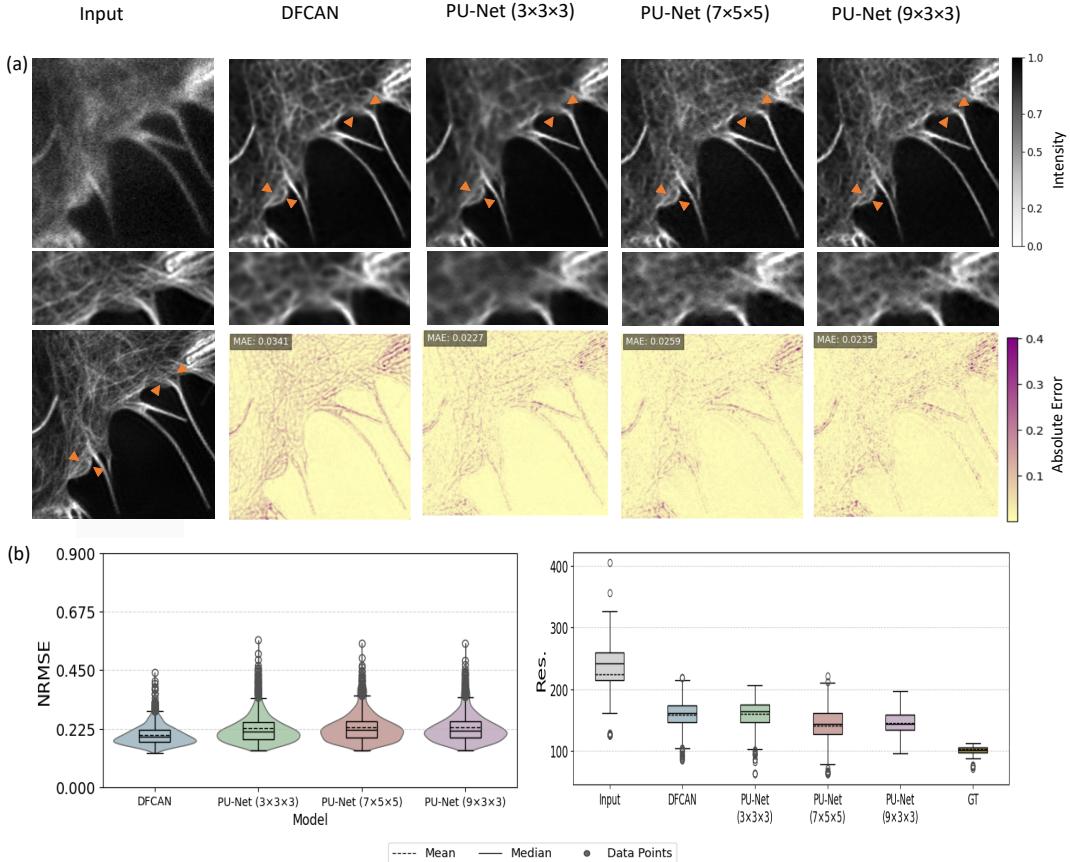


Figure 3.3: **Model comparison of DFCAN and PU-Net variants on Microtubules.** (a) Model performance comparison on image quantity evaluation metrics. (b) Training-loss plots for the DFCAN vs. PU-Net variants.

## **F-actin Assessment**

As detailed in Section 1.3.3, F-actin poses greater challenges than Microtubules because of its densely packed, overlapping fine-scale filaments. In our experiment, Figure 3.4 a, shows that the PU-Net equipped with a  $9 \times 3 \times 3$  kernel in the first layer only, resolves F-actin features far more clearly than either the PU-Net ( $3 \times 3 \times 3$ ) or DFCAN. This advantage is borne out in the quantitative resolution analysis (Figure 3.4 b), where the PU-Net ( $9 \times 3 \times 3$ ) delivers the highest effective resolution. This improved performance is attributed to the PU-Net's ability to capture spatial correlations across all channel dimensions by considering as z-axis, a capability that is limited in the 2D convolution layer of DFCAN method.



**Figure 3.4: SIM Reconstruction superiority of PU-Net on F-actin.** (a) The network prediction from DFCAN and PU-Net configurations, with the notable differences pointed by the arrow, demonstrates the superiority of PU-Net (9x3x3) in SIM reconstruction. (b) The decorrelation analysis plot demonstrates the robustness of PU-Net, showing a noticeable improvement in resolution.

All methods struggle in the very noisy and dense F-actin regions, as shown in Figure 3.5 a, low-SNR inputs lead to reconstructions that miss many of the fine structures seen in the ground truth. Figure 3.5 b demonstrates that PU-Net (9x3x3) recovers more small-scale detail than both DFCAN and PU-Net (3x3x3) (indicated by orange arrow), yet none of the models fully restores the finest filaments.

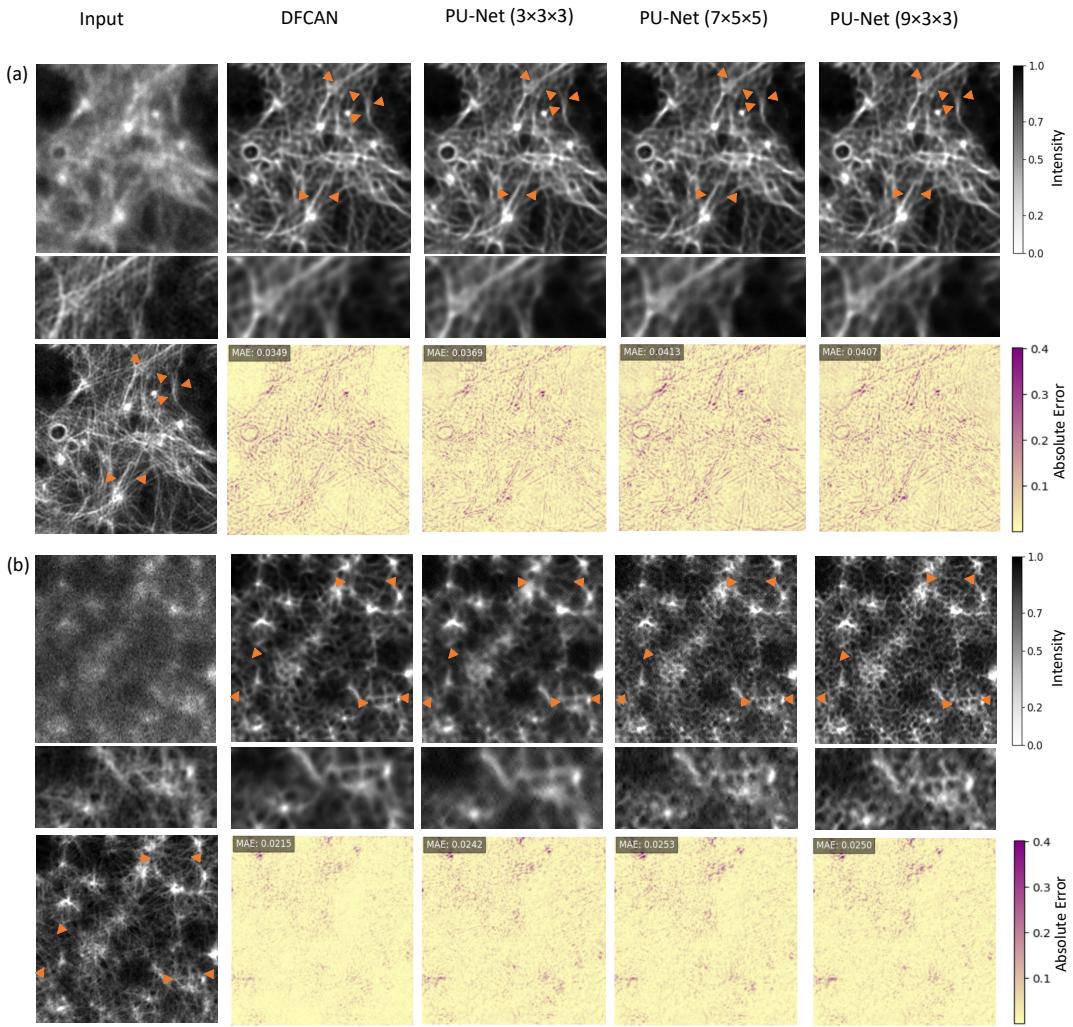
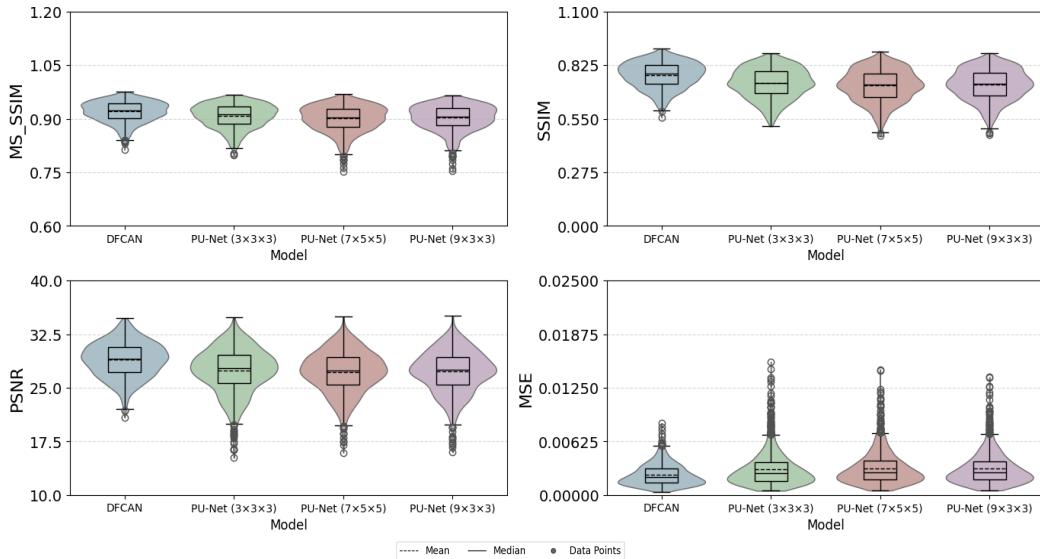


Figure 3.5: **SIM reconstruction from low-SNR F-actin data.** (a) SIM reconstruction of F-actin from mid-SNR raw data shows that DFCAN and PU-Net variants have similar capabilities. (b) For low-SNR data, PU-Net ( $7 \times 5 \times 5$ ) and PU-Net ( $9 \times 3 \times 3$ ) outperform DFCAN in capturing spatial correlations across the channel dimension.

(a)



(b)

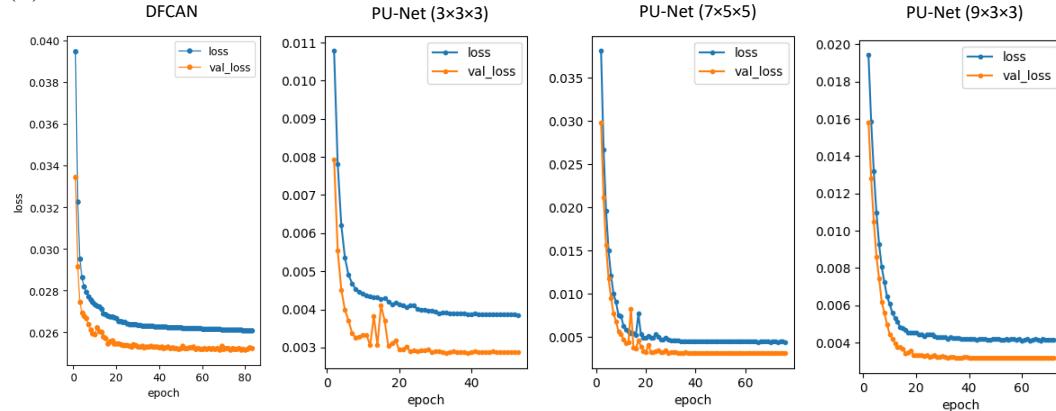
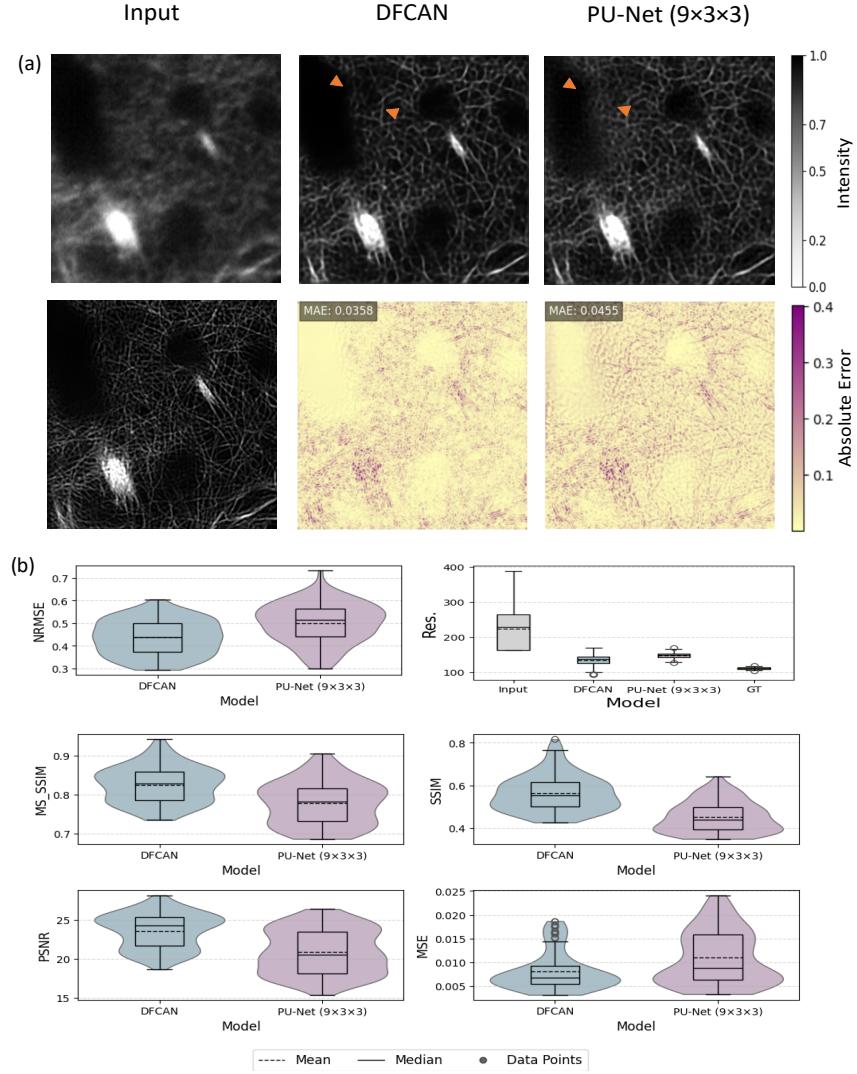


Figure 3.6: Model comparison of DFCAN and PU-Net variants on F-actin.

- (a) Model performance comparison on image quantity evaluation metrics.
- (b) Training-loss plots for the DFCAN vs. PU-Net variants.

We further evaluated the networks using our own F-actin dataset. We observed a bimodal distribution in the evaluation metrics, as illustrated in Figure 3.7 b. This observation indicates a limitation in our training and validation data, which is relatively small and shows limited variation among the structures presented. These findings highlight the critical need for a larger and more diverse training dataset, as it would enhance the network's ability to generalize and achieve finer detail reconstruction.



**Figure 3.7: SIM image reconstruction of our own F-actin data.** (a) The notable differences are indicated by the orange marks, showing that PU-Net ( $9 \times 3 \times 3$ ) captures more information than DFCAN. (b) The statistical evaluation of the network prediction from our own F-actin dataset.

## 3.2 Image Denoising Module

In line with our fourth objective, we employ the RDL-SIM workflow from Section 1.4.4, by replacing the DFCAN module with our proposed PU-Net (9x3x3). First, we generate an initial SIM reconstruction using a pretrained SIM reconstruction model. This reconstruction is then used to synthesize the pattern-modulated training images, which are paired with the original raw, noisy data, and used for the RDI-SIM denoising module. The proposed changes aim to explore the potential of leveraging the spatiotemporal correlations captured by the channel dimension in the 3D kernel of PU-Net (9x3x3), as detailed in section 1.4.5. This replacement is intended to determine whether PU-Net can effectively predict the SIM reconstruction from noisy raw data. Finally, we process those high-SNR images with the fairSIM plugin (Section 2.2.7) to obtain the final super-resolution SIM reconstructions and determine our fifth objective, the threshold for the lowest possible artifact-free super-resolution SIM reconstruction.

### Mictorubules Assessment

We evaluated Microtubule reconstructions at two different signal-to-noise ratio (SNR) levels, low and mid. The results indicate that both the DFCAN + RDL and PU-Net (9 $\times$ 3 $\times$ 3) + RDL approaches yield comparable outcomes regarding image reconstruction quality and evaluation metrics for both low-SNR, Figure 3.8 a and mid-SNR datasets, Figure 3.9 a. The images shown are from the first angle, the first phase of the nine-channel image stack. The image quality measurements presented in Figure 3.8 b and Figure 3.9 b, illustrate an identical performance of the models. Additionally, the quantitative assessment using SIM reconstruction through the FairSIM plugin in ImageJ, Section Section 2.2.7 further confirms that both methods are effective at reconstructing Microtubule structures from low and mid SNR levels.

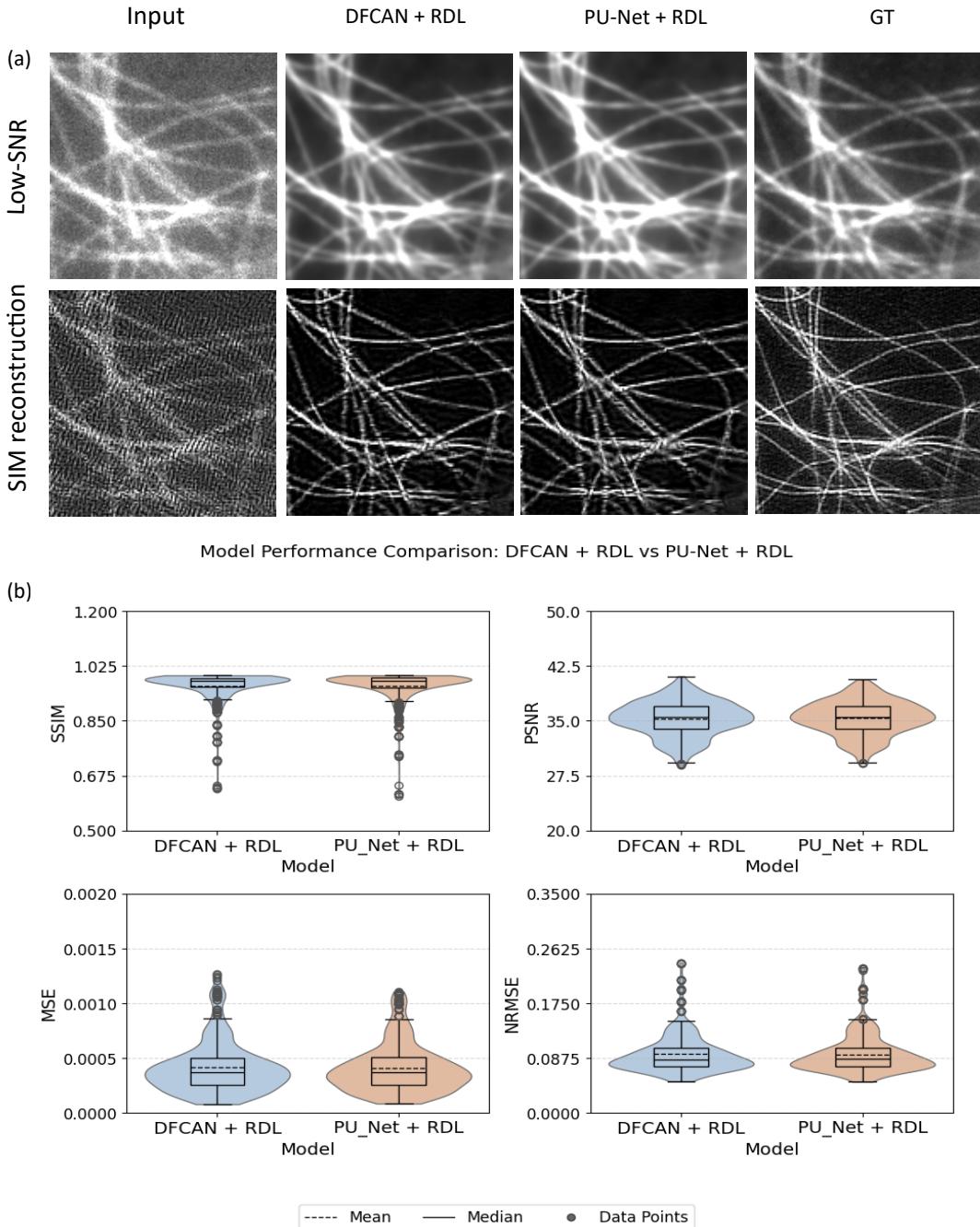


Figure 3.8: **Model comparison of low-SNR Microtubules data.** (a) Image reconstruction of Microtubules from low-SNR data, using DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL approaches, is illustrated alongside the corresponding SIM reconstruction. (b) A comparison of evaluation metrics indicates that both approaches exhibit similar performance.

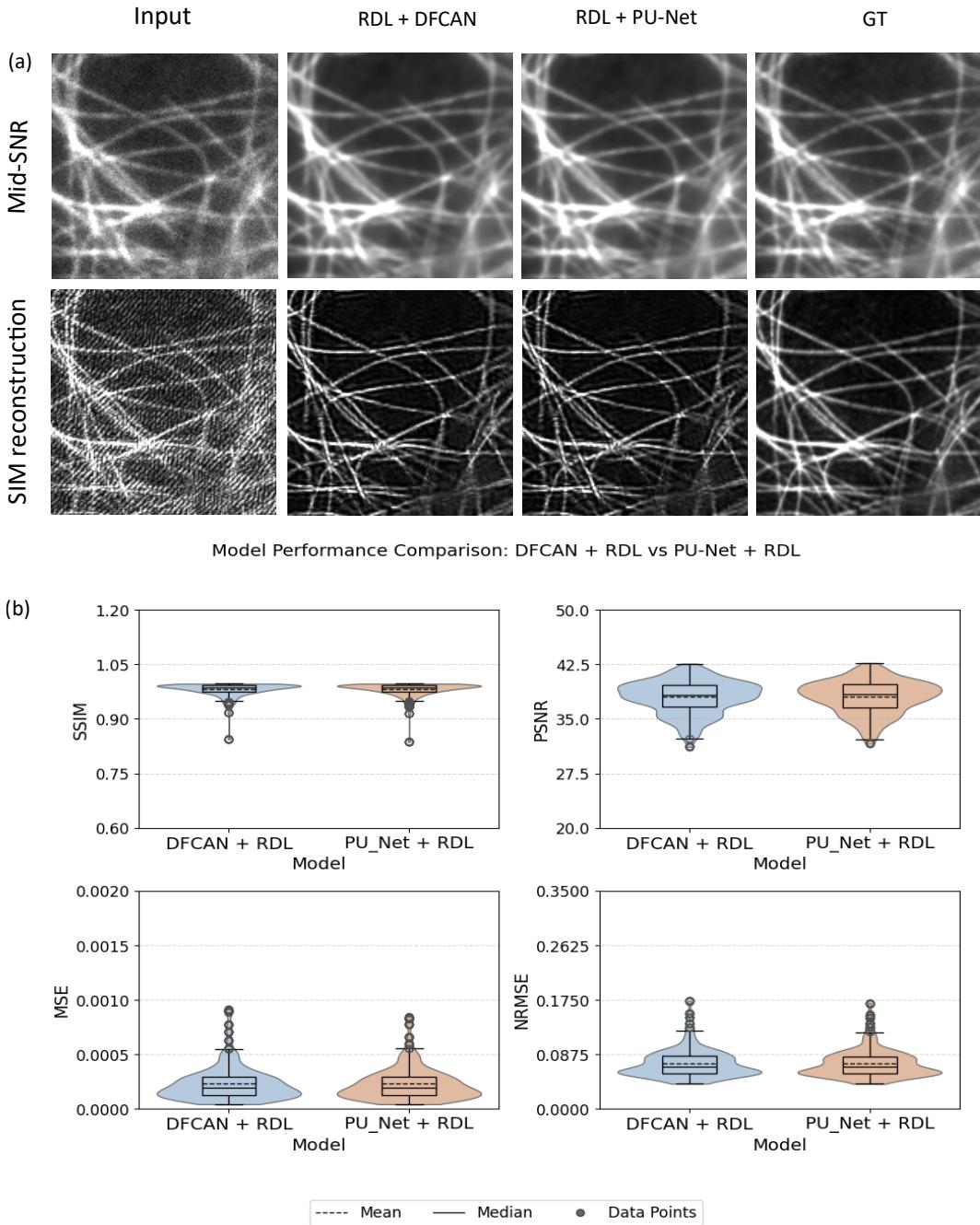
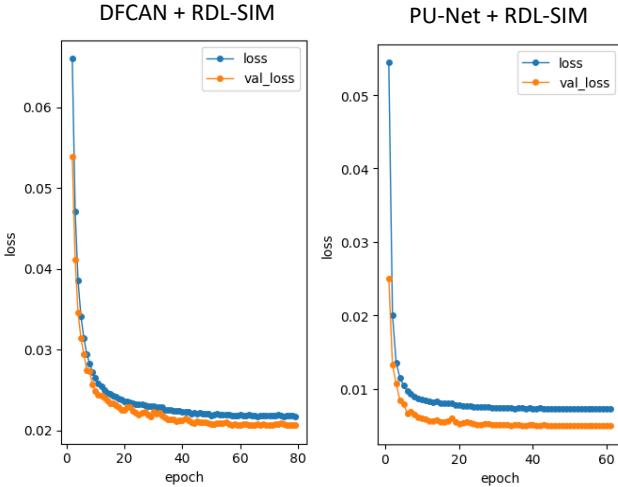


Figure 3.9: **Model comparison of mid-SNR Microtubules data.** (a) Image reconstruction of Microtubules from mid SNR data, using DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL approaches, is illustrated alongside the corresponding SIM reconstruction. (b) A comparison of evaluation metrics indicates that both approaches exhibit similar performance.



**Figure 3.10: Training loss comparison of mid-SNR Microtubules data.**  
Comparison of training loss for Microtubules using mid-SNR raw data.

### F-actin Assessment

First, we evaluated performance at low signal-to-noise ratio (SNR). Although both approaches yielded similar overall reconstruction quality, a closer inspection revealed difficulties in accurately reconstructing fine filament details by both approaches of DFCAN + RDL and PU-Net + RDL. As shown in Figure 3.11 a, the FairSIM plugin struggled with parameter estimation because the illumination patterns were not strongly evident in the predicted image. This issue likely stemmed from the limited, very noisy training data and the high structural density in the F-actin dataset. This conclusion is further supported by the broader distribution observed in the evaluation metrics, particularly the normalized root mean square error (NRMSE) and mean square error (MSE), shown in Figure 3.11b. As NRMSE is a scale-invariant image evaluation metric, the localization of pixels and shifts in the reconstructed image are well observed by this method. These metrics suggest that the model had difficulties generalizing the fine structural features effectively in comparison to the NRMSE plots shown in Figure 3.8 b.

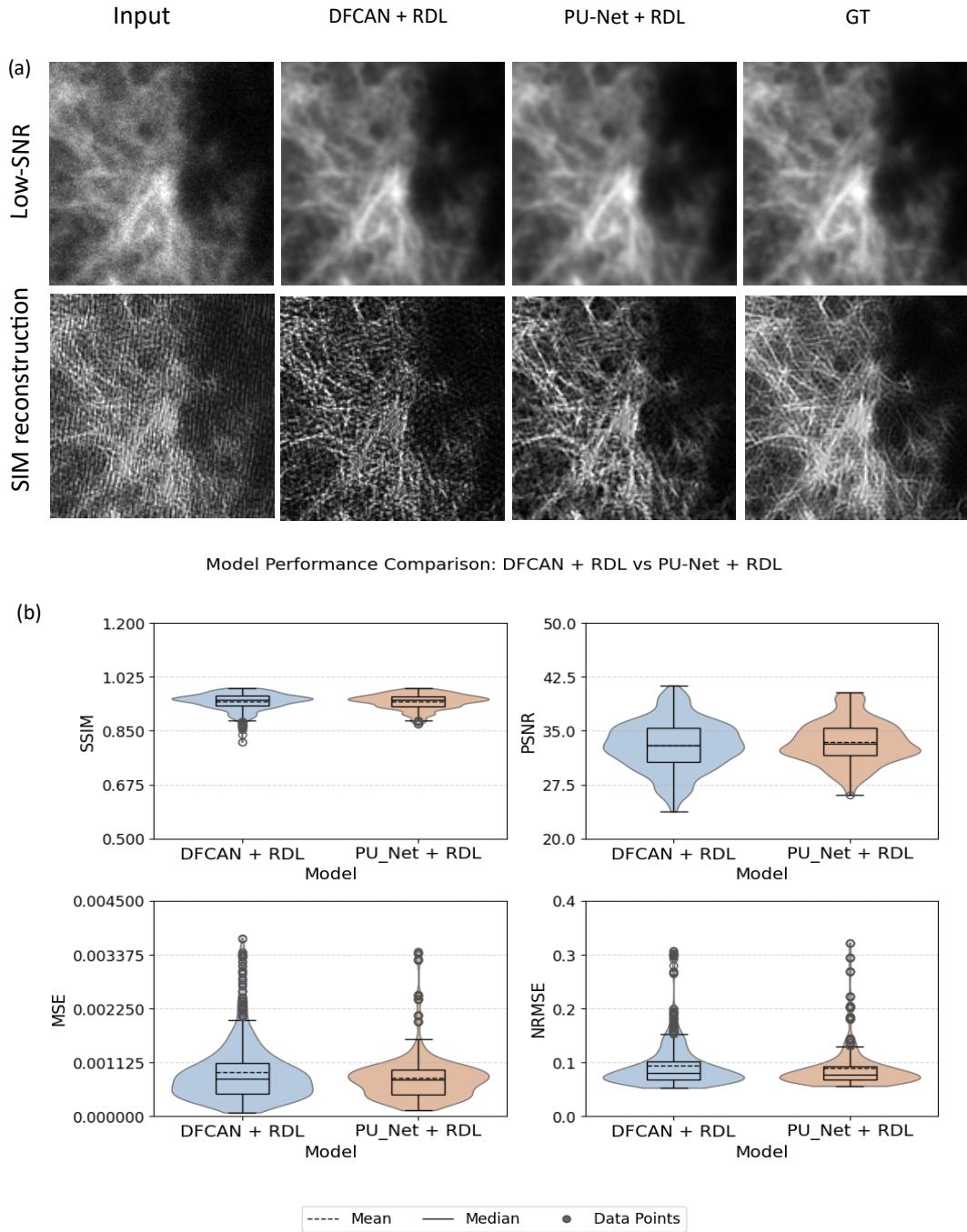
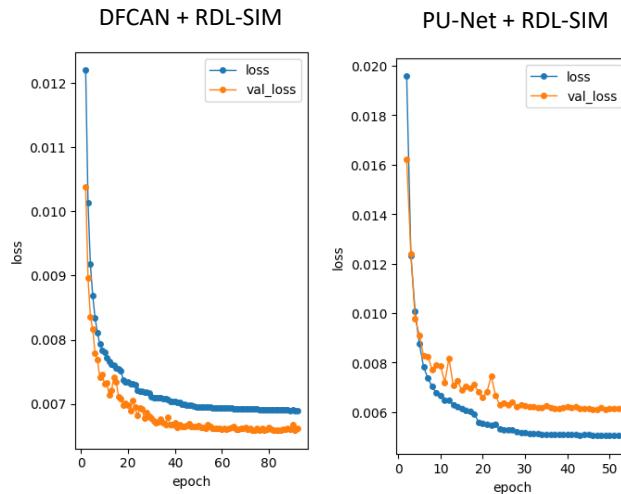
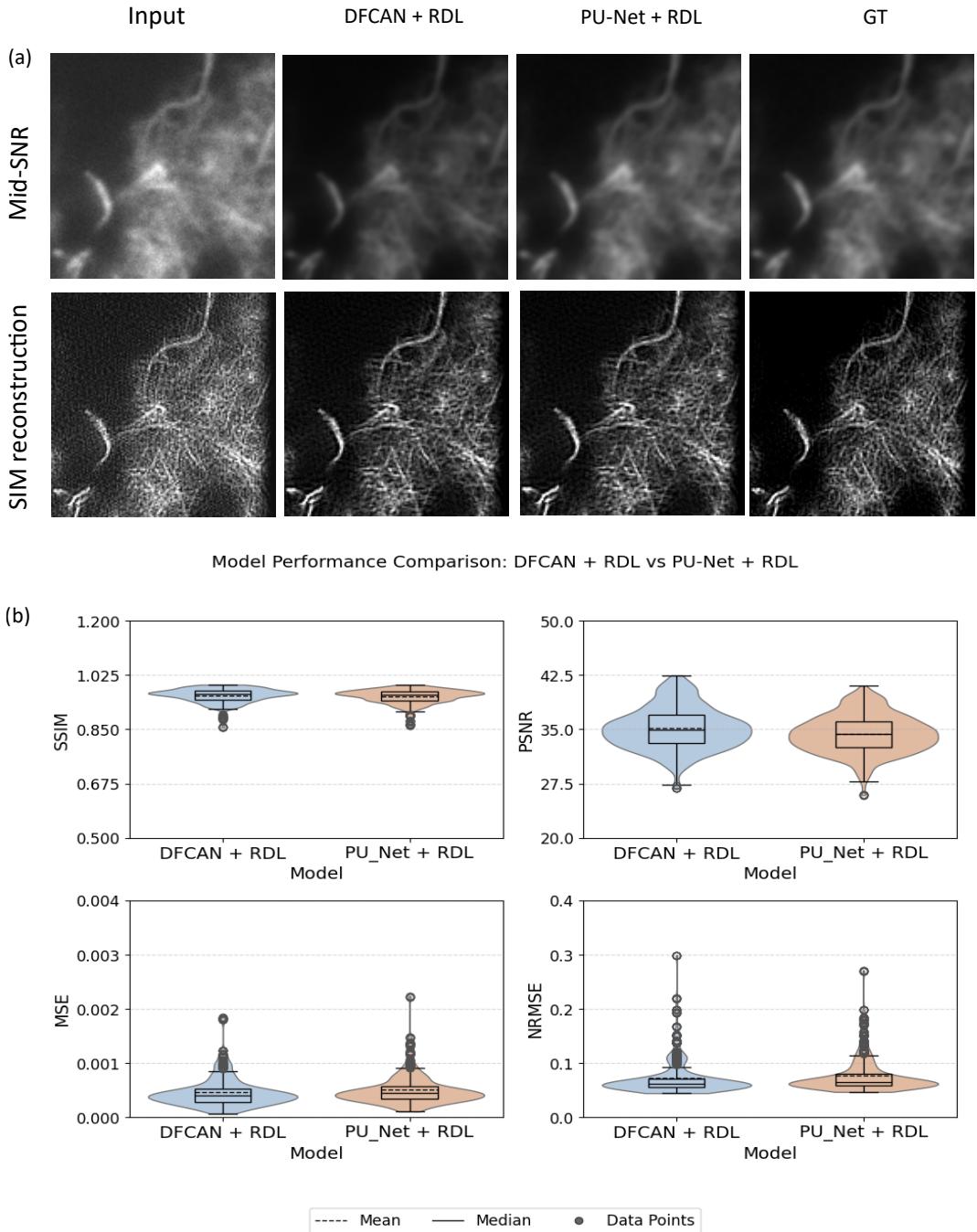


Figure 3.11: **Model comparison of low-SNR F-actin data.** (a) F-actin reconstructions from low-SNR raw data and SIM images for the DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL methods are shown. Despite fairSIM post-processing, both methods struggle to resolve fine filament details due to high density and noise. (b) Evaluation metrics indicate comparable performance for both approaches.

In the context of the F-actin mid-SNR levels, the denoised and the super-resolution SIM reconstructed images are significantly closer to the high-SNR ground truth images and the SIM reconstruction. The SIM reconstructions produced using the FairSIM plugin, for DFCAN + RDL, compared with PU-Net + RDL as shown in Figure 3.13 a, demonstrate a better reconstruction of fine details. This outcome is further supported by the qualitative analysis in Figure 3.13 a, where the NRMSE values are lower and exhibit a narrower distribution, indicating enhanced generalization performance at mid-SNR levels. Notably, the training loss plots for RDL-SIM with PU-Net highlight the comparative advantage of the PU-Net architecture and its SIM reconstruction superiority on dense structures, as evidenced by its consistently lower training loss relative to that observed with DFCAN combined with RDL-SIM, as shown in Figure 3.12. The impact of this can be observed in Figure 3.12 b, where the PSNR and NRMSE scores are slightly better than the DFCAN+RDL-SIM network.



**Figure 3.12: Training loss of mid-SNR F-actin data.** Comparison of training loss for F-actin using mid-SNR raw data, highlighting the impact of PU-Net in the loss curve.



**Figure 3.13: Model comparison of mid-SNR F-actin data.** (a) Image reconstruction of F-actin was done using mid-SNR noisy raw data and corresponding SIM reconstruction. The DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL methods yielded similar SIM results in the post-processing phase with fairSIM, indicating their equivalence. (b) Evaluation metrics show comparable performance for both approaches.

To further evaluate the generalization capabilities of the models, we conducted tests using our own F-actin dataset. As illustrated in (Figure 3.14) a, both approaches yield similar results on this dataset. However, the evaluation metrics in (Figure 3.14) b indicate bimodal distributions in the violin plots of NRMSE, MSE, and PSNR. This finding suggests potential biases within the training and validation datasets, highlighting the necessity for significantly increased data variability, a larger dataset, data augmentation strategies, and potential fine-tuning of the RDL component to enhance the robustness of the models. We have observed a similar bimodal distribution in the SIM reconstruction module, shown in Figure 3.7 b.

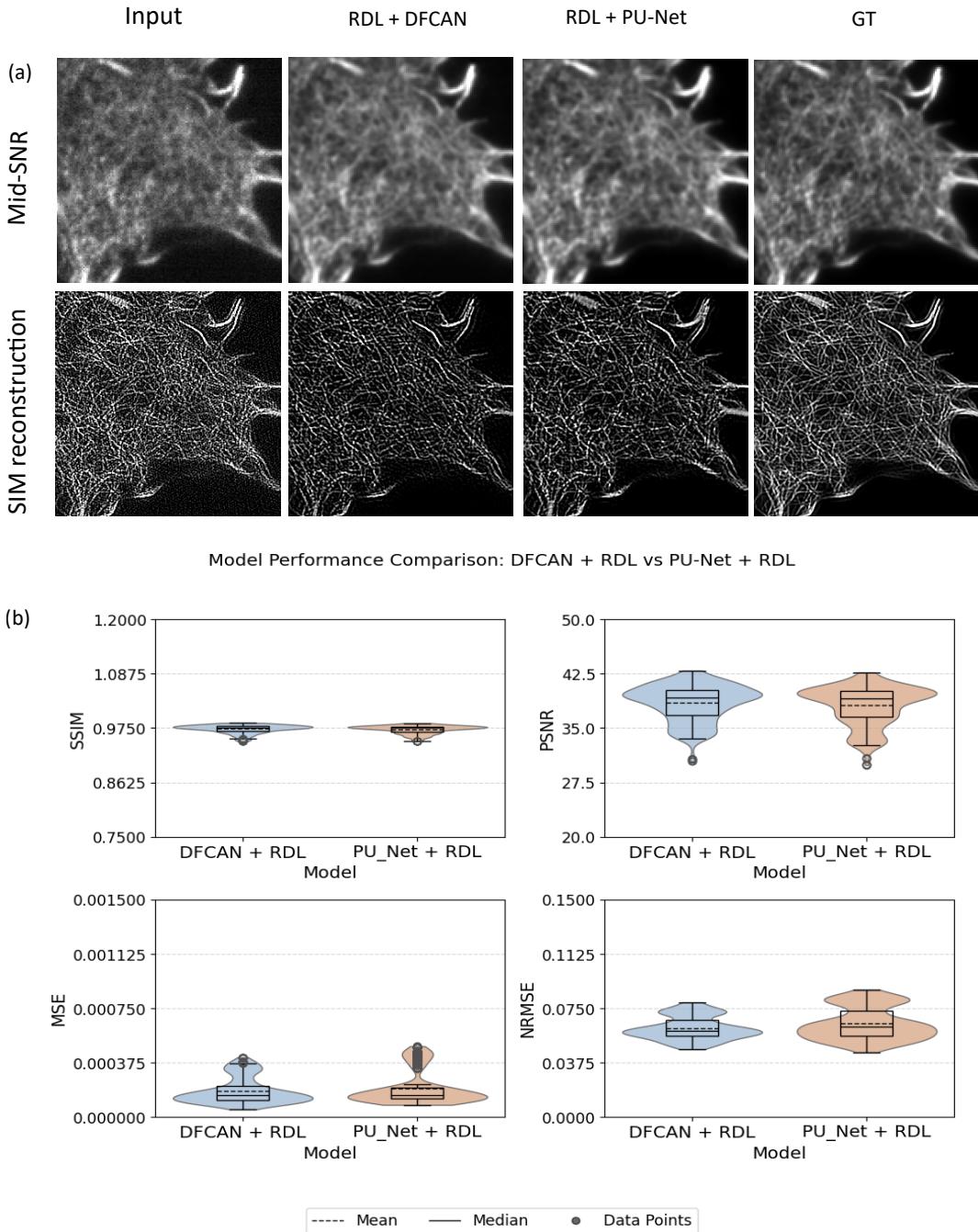


Figure 3.14: **Model comparison of our own low-SNR F-actin data.** (a) Image reconstruction of F-actin from noisy data, featuring DFCAN + RDL and PU-Net ( $9 \times 3 \times 3$ ) + RDL outputs. Post-processing with fairSIM shows both methods yield similar SIM reconstructions, demonstrating their equivalence. (b) Evaluation metrics reveal comparable performance between the two approaches.

## 4 Discussion and Outlook

In this study, we explored advanced deep learning architectures to enhance the image reconstruction of Structured Illumination Microscopy (SIM). Our focus was on denoising and resolution improvement in challenging low-SNR conditions, and we present the findings along with their limitations.

According to our third objective, benchmarking SIM reconstruction, we train Microtubules and F-actin both separately with 3 different variants of PU-Net and the comparison model DFCAN. As shown in Figure 3.1 a, for Microtubules and Figure 3.4 a, for structurally highly complex specimens F-actin where the filaments are as small as 7nm in diameter (see Section 1.3.3), we show that PU-Net has a clear advantage in comparison to DFCAN in reconstructing super-resolution SIM images. This striking gain is more visible in F-actin, where DFCAN fails to resolve the finer filaments, whereas the variant of PU-Net (7x5x5) and PU-Net(9x3x3) can reconstruct and recover these fine features. We can confirm this from the resolution analysis, where we can see a significant gain in resolution for both of the biological specimens with PU-Net.

Previous work by Qiao et al. (2021) demonstrated a pronounced trade-off between achieved resolution and reconstruction error when applying U-Net-based architectures to Microtubules and F-actin. In their study, the U-Net-based model ScU-Net with 2D kernels was unable to resolve the finest filamentous structures, whereas 2D Fourier channel attention mechanism-based approaches, such as DFCAN and its generative counterpart, DFGAN [9], successfully recovered sub-diffraction details. In contrast, our modified U-Net-based architecture, PU-Net ( $9 \times 3 \times 3$ ) variant, reliably captures and reconstructs these fine-grained features, outperforming the state-of-the-art method DFCAN by yielding notably higher resolution and marginally lower NRMSE. We attribute this improvement to PU-Net's true 3D kernels, which exploit

spatial correlations along channel dimensions(3 angle,3 phase) that are inherently limited in purely 2D kernel-based architectures.

Our overall findings for the super-resolution SIM reconstruction module emphasize that, as we hypothesize in our aims and objectives 1.5, since the imaging is consistent across the nine channels and is concatenated as image data, capturing spatial correlations along this channel dimension by treating it as a Z-axis adds information that is not present in 2D kernel-based models. The 2D kernel-based models typically capture global cross-correlations across the channels and may require computationally expensive, laborious engineering approaches, such as Fourier channel attention methods (like DFCAN), to extract the same missing information. In contrast, the approach taken by PU-Net is much simpler, faster, and is more aligned with machine learning practices, utilizing 3D kernels to capture local information at scale while achieving marginally superior performance levels comparable to DFCAN. From these findings, we chose PU-Net (9x3x3) for further implications of the next objectives.

Despite its strengths, our PU-Net pipeline has exhibited 'honeycomb' artifacts when applied to a very noisy and dense F-actin dataset. To address this issue, we plan to experiment with replacing the last 2D UpSampling layer with a 2D Transpose Convolution layer, followed by a 2D convolutional layer. This configuration enables the network to learn its upsampling kernels and has been shown to reduce checkerboard and honeycomb artifacts in other deep learning-based image reconstruction tasks.

To address our fourth and fifth objectives, PU-Net integration and determining SNR threshold, we have trained two separate models of PU-Net+RDL-SIM and DFCAN+RDL-SIM on each biological specimen for each low and mid signal-to-noise levels. The resulting reconstructions were then processed with the FairSIM plugin (Section 2.2.7) to identify the lowest SNR permitting artifact-free SIM reconstructions and to delineate potential application domains.

Our experiments on Microtubules show that both the DFCAN + RDL and PU-Net (9 $\times$ 3 $\times$ 3) + RDL approaches yield comparable image reconstruction quality for low-

SNR and mid-SNR datasets, as presented in Figure 3.8 and Figure 3.9. Additionally, the quantitative assessment through SIM reconstruction with the FairSIM plugin in ImageJ (Section Section 2.2.7) confirms the effectiveness of both methods in reconstructing microtubule structures from significantly noisy low SNR images.

For F-actin analysis, we evaluated performance at low signal-to-noise ratio (SNR). Although both DFCAN + RDL and PU-Net + RDL approaches produced similar denoised overall quality, the FairSIM plugin encountered challenges in parameter estimation due to unclear illumination patterns in the predicted image, thus resulting in artifact-prone super-resolution SIM reconstruction. On the other hand, at mid-SNR levels, the denoised reconstructed images closely resembled high-SNR ground truth images as presented in Figure 3.13 a. The network predicted denoised image demonstrated marginally comparable performance for PU-Net + RDL over DFCAN + RDL, supported by lower NRMSE values and narrower distributions. The training loss for RDL-SIM using PU-Net (as shown in Figure 3.12) consistently demonstrated an advantage, resulting in a overall lower loss divergence. This indicates the success of PU-Net within the RDL-SIM pipeline. Additionally, the SIM reconstruction from fairSIM provided improved, artifact-free image reconstruction compared to the original input image, highlighting its potential application for mid-SNR levels in the F-actin dataset.

All of our experiments confirm that F-actin’s densely interwoven, finely branched structure makes accurate recovery of its thinnest filaments far more difficult than in Microtubules. Importantly, we trained on nearly the same number of F-actin and Microtubule cells, yet as detailed in Section 1.3.3, F-actin’s substantially higher structural complexity. This implies that a much larger, more diverse training set is necessary to achieve comparable reconstruction fidelity.

To achieve our final objective, we train and validate both the SIM reconstruction module and the RDL-SIM image denosing module with our F-actin dataset. The above-mentioned data limitations become most pronounced as the ROIs are small ( $300 \times 300 \times 9$ ), dominated by background, and exhibit little morphological diversity. Coupled with single-level, low-SNR inputs, the SIM reconstruction module struggles to learn fine-scale features, resulting in the bimodal performance distributions seen

in our density plots for both the SIM reconstruction module(Figure 3.7 b) and the image denoising module(Figure 3.14 b). This example underscores a key principle of deep learning: as the structural complexity of the data increases, the need for larger and more varied training sets grows in tandem. Adequate data quantity and diversity are therefore essential for achieving reliable, high-fidelity reconstructions in the field of bioimaging.

We found that, when predicting SIM reconstructions from a single time frame, low-SNR inputs, FairSIM often produces unreliable pattern-parameter estimates. As Qiao et al. (2022c) have shown, averaging raw acquisitions across multiple time points can substantially improve parameter fitting compared to using only a single time frame. Such temporal integration is particularly important for RDL-SIM, which depends on accurate prior pattern estimates. Unfortunately, because the BioSR dataset contains only single-time-point raw images, we were unable to evaluate the benefits of this multi-frame averaging strategy.

Moreover, in our experiments, we found that the DFCAN + RDL pipeline relies on the modulation pattern extracted from high-SNR ground-truth images as priors during training. However, when using  $128 \times 128 \times 9$  patches, many regions lack uniform filament density or clear modulation fringes, resulting in poor parameter estimates in the fairSIM plugin. As a next step, we plan to train on larger volumetric patches (e.g.,  $256 \times 256 \times 9$ ) and expand our dataset. The greater field of view should capture more structural features and well-defined moiré patterns, leading to more reliable prior estimates and more robust reconstructions.

In conclusion, this study demonstrates that PU-Net significantly enhances the fidelity of super-resolution SIM image reconstruction, especially for specimens like F-actin, which are structurally complex and dense. The PU-Net offers a much simpler and more efficient design with notably higher resolution in comparison to DFCAN. Our hypothesis suggests that incorporating 3D kernels to capture spatial correlations is an effective and faster alternative to DFCAN for the current RDL-SIM workflow. Furthermore, the results underscore the pivotal role of training data quality and diversity in conquering the reconstruction challenges posed by intricate biological structures. Collectively, these findings fulfill the primary objectives of this project and establish a robust foundation for future advances in high-fidelity super-resolution SIM reconstruction.

## 5 Declaration of used AI tools

The purpose of this thesis is to investigate and develop potential AI tools for image denoising and Structured Illumination Microscopy (SIM) reconstruction. However, the methods discussed in this thesis are not the only AI tools that have been utilized. In accordance with the AI tools guidelines of the University of Osnabrück, I would like to briefly outline the AI tools I used throughout this project.

I used GitHub Copilot and ChatGPT to help debug Python code and ImageJ scripts, as well as to resolve issues with LaTeX syntax, which were essential for this project. Additionally, I employed Grammarly to identify grammatical errors and enhance my academic writing.

## 6 Bibliography

1. Wu, Y. & Shroff, H. Faster, sharper, and deeper: structured illumination microscopy for biological imaging. *Nature Methods* **15**, 1011–1019. doi:10.1038/s41592-018-0211-z (Nov. 2018).
2. Qiao, C. *et al.* Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nature Biotechnology* **41**, 367–377. doi:10.1038/s41587-022-01471-3 (Oct. 2022).
3. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie* **9**, 413–468. doi:10.1007/bf02956173 (Dec. 1873).
4. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy* **198**, 82–87. doi:10.1046/j.1365-2818.2000.00710.x (May 2000).
5. Ma, Y. *et al.* Recent advances in structured illumination microscopy. *Journal of Physics Photonics* **3**, 024009. doi:10.1088/2515-7647/abdb04 (Jan. 2021).
6. *Structured Illumination Microscopy and 3D SIM Imaging- Oxford Instruments* <https://andor.oxinst.com/learning/view/article/super-resolution-imaging-structured-illumination-microscopy>.
7. Heintzmann, R. & Huser, T. Super-Resolution Structured Illumination Microscopy. *Chemical Reviews* **117**, 13890–13908. doi:10.1021/acs.chemrev.7b00218 (Nov. 2017).
8. Winkelmann, H., Richter, C. P., Eising, J., Piehler, J. & Kurre, R. Correlative single-molecule and structured illumination microscopy of fast dynamics at the plasma membrane. *Nature Communications* **15**. doi:10.1038/s41467-024-49876-9 (July 2024).
9. Qiao, C. *et al.* Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* **18**, 194–202. doi:10.1038/s41592-020-01048-5 (Jan. 2021).

10. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**, 1090–1097. doi:10.1038/s41592-018-0216-7 (Nov. 2018).
11. Ronneberger, O., Fischer, P. & Brox, T. *U-NET: Convolutional Networks for Biomedical Image Segmentation* 234–241. doi:10.1007/978-3-319-24574-4\{\_}28 (Jan. 2015).
12. Jin, L. *et al.* Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nature Communications* **11**. doi:10.1038/s41467-020-15784-x (Apr. 2020).
13. Markwirth, A. *et al.* Video-rate multi-color structured illumination microscopy with simultaneous real-time reconstruction. *Nature Communications* **10**. doi:10.1038/s41467-019-12165-x (Sept. 2019).
14. Qiao, C. *BioSR+: Dataset Extension of biological images for super-resolution microscopy* version v1 (Zenodo, Sept. 2022). doi:10.5281/zenodo.7115540.
15. A.Dosovitskiy, J.T.Springenberg, M.Riedmiller & T.Brox. *Discriminative Unsupervised Feature Learning with Convolutional Neural Networks* in *Advances in Neural Information Processing Systems 27 (NIPS)* (2014).
16. Fardo, F. A., Conforto, V. H., C, D. O. F. & Rodrigues, P. S. A Formal Evaluation of PSNR as Quality Measurement Parameter for Image Segmentation Algorithms. *arXiv (Cornell University)*. doi:10.48550/arxiv.1605.07116 (Jan. 2016).
17. Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600–612. doi:10.1109/tip.2003.819861 (Apr. 2004).
18. Wang, Z., Simoncelli, E. & Bovik, A. *Multiscale structural similarity for image quality assessment* in *The Thirly-Seventh Asilomar Conference on Signals, Systems and Computers, 2003* **2** (2003), 1398–1402 Vol.2. doi:10.1109/ACSSC.2003.1292216.
19. Descloux, A., Grußmayer, K. S. & Radenovic, A. Parameter-free image resolution estimation based on decorrelation analysis. *Nature Methods* **16**, 918–924. doi:10.1038/s41592-019-0515-7 (Aug. 2019).

20. Müller, M., Mönkemöller, V., Hennig, S., Hübner, W. & Huser, T. Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ. *Nature Communications* **7**. doi:10.1038/ncomms10980 (Mar. 2016).
21. Heintzmann, R. & Cremer, C. G. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE*. doi:10.1117/12.336833 (Jan. 1999).