University of Memphis University of Memphis Digital Commons

Electronic Theses and Dissertations

8-24-2023

STRUCTURED ILLUMINATION MICROSCOPE IMAGE RECONSTRUCTION USING UNROLLED PHYSICS-INFORMED GENERATIVE ADVERSARIAL NETWORK (UPIGAN)

Seyedeh Parisa Dajkhosh

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation

Dajkhosh, Seyedeh Parisa, "STRUCTURED ILLUMINATION MICROSCOPE IMAGE RECONSTRUCTION USING UNROLLED PHYSICS-INFORMED GENERATIVE ADVERSARIAL NETWORK (UPIGAN)" (2023). *Electronic Theses and Dissertations*. 2992. https://digitalcommons.memphis.edu/etd/2992

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.

STRUCTURED ILLUMINATION MICROSCOPE IMAGE RECONSTRUCTION USING UNROLLED PHYSICS-INFORMED GENERATIVE ADVERSARIAL NETWORK (UPIGAN)

by

Seyedeh Parisa Dajkhosh

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Electrical & Computer Engineering

The University of Memphis

August 2023

Copyright© 2023 Seyedeh Parisa Dajkhosh All rights reserved

Abstract

In three-dimensional structured illumination microscopy (3D-SIM) where the images are taken from the object through the point spread function (PSF) of the imaging system, data acquisition can result in images taken under undesirable aberrations that contribute to a model mismatch. The inverse imaging problem in 3D-SIM has been solved using a variety of conventional model-based techniques that can be computationally intensive. Deep learning (DL) approaches, as opposed to traditional restoration methods, tackle the issue without access to the analytical model. This research aims to provide an unrolled physics-informed generative adversarial network (UPIGAN) for the reconstruction of 3D-SIM images utilizing data samples of mitochondria and lysosomes obtained from a 3D-SIM system. This design makes use of the benefits of physics knowledge in the unrolling step. Moreover, the GAN employs a Residual Channel Attention super-resolution deep neural network (DNN) in its generator architecture. The results indicate that the addition of both physics-informed unrolling and GAN incorporation yield improvements in reconstructed results compared to the regular DL approach.

TABLE OF	CONTENTS
----------	-----------------

Ab	ostrac	t		iii
Li	st of]	Fables		vi
List of Figures			vii	
1	Intr	oductio	n	1
	1.1	Overv	iew	1
	1.2	Structu	ared Illumination Microscopy (SIM)	1
	1.3	Image	Reconstruction	4
		1.3.1	Image-to-image translation	5
		1.3.2	Image Super-Resolution	6
			Optical image super-resolution	6
			Digital image super-resolution	6
	1.4	Challe	nges	8
	1.5	Motiva	ation	9
	1.6	Contri	bution	10
2	Bacl	kground	d and Related Work	11
	2.1	Datase	ets	11
		2.1.1	SIM datasets	12
		2.1.2	Simulated datasets	13
	2.2	Deep l	Learning Background	14
		2.2.1	Fully Connected Neural Networks (FCN)	16
		2.2.2	Convolutional Neural Networks (CNNs)	16
		2.2.3	Super-Resolution Convolutional Neural Network (SRCNN)	17
		2.2.4	UNET	18
		2.2.5	Residual Channel Attention Network (RCAN)	21
		2.2.6	Generative Adversarial Netowrks (GANs)	22
			Conditional GAN (CGAN)	23
			Super-Resolution GAN (SRGAN)	23
			PIX2PIX Channel Attention CAN (CACAN)	24
		227	Channel Attention GAN (CAGAN)	25
		2.2.7	Unrolling	28 20
		2.2.0		30 21
			Dhysics Informed Unrolling	22
		229	Conclusion	32
		2.2.9		55
3	Data Processing		35	
	3.1	Data P	reprocessing	35
		3.1.1	Crop to Patches	35
		3.1.2	Normalization	36
			Min-Max Norm	37

Percentile Norm	38
3.2 Datasets	38
Methodology and Implementation	41
4.1 Unrolling	43
4.2 Loss Function	44
4.2.1 Adversarial Loss Function	44
4.2.2 Generator Loss	45
4.2.3 Discriminator Loss	46
4.3 UPIGAN Design	46
Experimental Results	50
5.1 Ablation Study	50
5.1.1 USR	54
5.2 Results	55
5.2.1 FairSIM Training	55
5.2.2 FixedCell Training	58
Discussion and Conclusion	66
6.1 Discussion	66
6.2 Conclusion	70
Appendix	
Original caGAN WF Loss	74
3 N 4 4 E 5 E 6 6 6 A C	Percentile Norm 2 Datasets Methodology and Implementation 1 Unrolling 2 Loss Function 4.2.1 Adversarial Loss Function 4.2.2 Generator Loss 4.2.3 Discriminator Loss 3 UPIGAN Design Experimental Results 1 Ablation Study 5.1.1 USR 2 Results 5.2.1 FairSIM Training 5.2.2 FixedCell Training Discussion and Conclusion 1 Discussion 2 Conclusion Appendix Driginal caGAN WF Loss

LIST OF TABLES

Table		Page
1.	Training, Validation, Testing, split of FairSIM dataset including raw 3D-SIM images and the corresponding reconstructed images [1].	39
2.	Training, Validation, Testing, split In this study, we used 18 samples of raw 3D-SIM images and the corresponding reconstructed images of mitochondrial dynamics in fixed H9c2 rat cardiomyoblast cells [2], [3].	40
3.	RCAN ablation study using FairSIM dataset. the first column is the scenario number for each experiment. The following four columns correspond to the hyper-parameters adjusted in this project. The # sign corresponds to "the number of". The last four columns show the performance metrics as the result of training the network using the parameters in the first four columns. The blue values represent the performance of the parameters that are finalized for completing the training. "tr" and "Chs" also stand for the "training" and "Channels" respectfully.	51
4.	GAN ablation study using FairSIM dataset. "Gen" stands for "Generator"	52
5.	γ ablation study using FairSIM dataset.	54
6.	The fixed parameters used for training the FairSIM dataset. The character "#" stands for "the number of" and "lr" is the abbreviation for "learning rate".	56
7.	Metric performance for the architecture of RCAN super-resolution with or without GAN, with or without physics-informed unrolling, and with both GAN and physics-informed unrolling (UPIGAN) on the FairSIM dataset.	56
8.	The parameters used for training the Fixedcell dataset.	61
9.	Metric performance for the architecture of RCAN super-resolution with or without GAN, with or without physics-informed unrolling, and with both GAN and physics-informed unrolling (UPIGAN) on the Fixedcell dataset.	61

vi

LIST OF FIGURES

Figure

- 1. 3D-SIM imaging is used. (A and B) show the impact of the striped interference patterns on detection in frequency space using structured illumination (B) and regular WF illumination (A, respectively). (C) 3D-SIM allows the separation of displaced components and their restoration to their correct placement in frequency space by recording five phases of the sine wave pattern at each z point, extending the axial support while maintaining within the resolution limit. Additionally, the diffraction grating is rotated into three places that are 60 degrees apart from one another, producing support that is almost rotationally symmetric throughout a broader frequency space area. Three picture stacks are then captured. Compared to the original image, the 3D-SIM reconstruction shows more accurate image details. A rise in resolution is seen by the frequency space graphic on the right. [4]
- SIM super-resolution compared to Wide-Field Microscopy: a) 15 Raw SIM Input Images (5 phases, 3 orientation angles of SI pattern). b) SIM reconstruction using GWF: Frequency component decomposition and filtering in Fourier space. c) Wide-Field Image: Sum of 15 raw SIM images. d) SIM provides approximately 2x resolution improvement over wide-field microscopy [1].
- 3. Super-Resolution in 3D SIM. The passage describes the observable regions for various types of microscopes, including WF, SIM using two or three illumination beams, and their corresponding spatial-frequency components. It also highlights the highest possible spatial frequencies that can be generated by illuminating through the objective lens in the case of the three-beam structured illumination microscopy [5].
- 4. Digital Image Super-Resolution [6] 7 5. 14 Poisson noise's impact on sample simulated chromatin architectures [7]. 6. A simple feedback system: for the forward direction, "A" corresponds to the analytical transfer function indicating how outputs are driven from the given inputs, and "B" explains the analytical formula of how the output is observed by sensors, etc. then, returned as an addition to the input of the system [8]. 15 7. An example of a Fully Connected Deep Neural Network [9]. 16
- 8. A example of a typical Convolutional Neural Network (CNN) [10]. 17

3

5

7

9.	Convolutional Neural Networks architecture for single image super resolution [11].	18
10.	Original U-Net Architecture for biomedical image segmentation [12].	19
11.	MIMO U-Net architecture used for image deblurring [13].	19
12.	Deep Residual Fourier Transformation deblurring based on MIMO-U- Net [14].	20
13.	U-Net structure used for image segmentation and restoration on atom images [15].	20
14.	RCAN Architecture, consists of an array of residual groups following a convolution layer and followed by the up-sampling layer and another convolution layer. Each residual group itself includes an array of residual blocks in it followed by a convolutional layer [16].	21
15.	Residual Network [17].	22
16.	"From left to right: bicubic interpolation, deep residual network optimized for MSE, deep RGAN optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4x upscaling]" [18].	24
17.	Network architecture for the proposed caGAN [19].	25
18.	The input performance of the mitochondrial inner membrane using various forms of raw pictures is compared. (a) A single wide-field raw picture volume slice. (b) GT-SIM image, which is a 3D-SIM image that was created using a conventional algorithm from raw data with high SNR. First column, the caGAN-SIM-WF picture was created from a single wide-filed image volume. (15 raw SIM image volumes) were used to rebuild the caGAN-SIM picture in the second column. In third and forth columns the scU-net [20], and the caGAN-SIM image were compared respectively. The caGAN-SIM-RP was created using fewer phase pictures and comprised of three raw SIM image volumes plus a WF image volume [19].	26
19.	Recurrent physics-informed ML engine. Where $N = 1 - \alpha . H^T . H$, and alpha is a small step size [21]	29
20.	Cascaded physics-informed ML engine [21].	30

21.	Sample image resolution enhancement using cascaded physics-informed ML engine. From left, the first is the raw input image. Then, each image corresponds to the output of first, second, third and fourth layers of the	
	cascaded network [21].	31
22.	Algorithm unrolling [22]	32
23.	UODP	33
24.	Crop the input images of sizes of 512x512xvariable depths to 128x128x3 using the crop2patches function.	36
25.	3D data sample, [1].	39
26.	GAN Block Diagram. Respectively, (a) and (b) represent the discriminator and the generator design, respectively. Each block takes the required inputs and passes them to the discriminator/generator architecture. In this architecture, input <i>d</i> includes 15 raw SIM images from a single focal place of the object, <i>o</i> (corresponding to 3 angles and 5 phases of the SI pattern, Fig. 1). <i>o</i> represents the ground truth, and <i>h</i> is the physics-informed term (PSF). The generator architecture has a super-resolution network implemented in it. After completion of training, \hat{o}_n provides the final reconstructed image [23].	48
27.	UPIGAN Generator Block Diagram. This block magnifies the SR block from Fig. 26. kx for $k = 1, 2$ is the super-resolution magnification in the RCAN block. At first, the raw input's resolution is doubled using RCAN architecture. In the unrolling step, the resolution in the RCAN block stays the same, as we have already achieved double the resolution; but it only enhances the quality of the image.	49
28.	RCAN FairSIM ablation Study metrics performance presented including a) the MAE from the RCAN loss function during the training, b-d) MSE, PSNR, and SSIM on the validation dataset along the time. The horizontal axis in each figure corresponds to the number of epochs ranging from 0 to 800. The vertical axis shows the corresponding metric value.	52
29.	GAN FairSIM ablation Study metrics performance.	53
30.	γ FairSIM ablation Study metrics performance presented.	53
31.	Block Diagram of the proposed unrolled Super-Resolution (USR) architecture. The math is shown on the right side where \hat{y}_k is the super-resolved output in k^{th} unrolling iteration, and x is the raw SIM input image [24].	55

ix

32.	FairSIM quantitative results comparison on various architectures.	57
33.	2D FairSIM qualitative results of 3 different samples. Each row corresponds to one sample.	58
34.	FairSIM qualitative results trained with 3d samples cropped to a 3D volume with $z = 3$. The first row represents the first slice of the volume in the z-axis. followed by $z = 2$ and in the last row $z = 3$.	59
35.	Another FairSIM qualitative results trained with 3d samples cropped to a 3D volume with $z = 3$.	60
36.	Fixedcell quantitative results comparison on various architectures [24], [23].	62
37.	2D Fixedcell qualitative results of 2 different samples. Each row corresponds to one sample.	63
38.	Fixedcell qualitative results trained with 3d samples cropped to a 3D volume with $z = 3$.	63
39.	3D FixedCell visual results with $z = 3$.	64
40.	Another Fixedcell qualitative results trained with 3d samples cropped to a 3D volume with $z = 5$.	65

Chapter 1 Introduction

1.1 Overview

This thesis discusses the study, development, and implementation of computational reconstruction of Structured Illumination Microscopy (SIM) pictures using an unrolled physics-informed generative adversarial network (UPIGAN), as well as its SIM applications. SIM computational reconstruction is a procedure that recovers the five-dimensional information stored in a SIM computationally instead of optically. The data for the iterative reconstruction approach in this thesis were gathered by [2] and [1]. The main objective is to correlate the reconstruction findings with the analytical information from the SIM system in order to boost confidence, enhance image quality, and finally create a computational platform that is effective and reliable for microscopy and general medical image reconstruction.

1.2 Structured Illumination Microscopy (SIM)

Light microscopy enables the study of three-dimensional (3D) live cell structures. Specifically, 3D structured illumination microscopy (3D-SIM) is a technique that utilizes structured illumination to modulate the fluorescence emission, thereby enabling higher frequencies to be transmitted than in wide-field (WF) fluorescence microscopy, and consequently, higher spatial resolution can be achieved in restored images after data processing [25]. WF is a type of fluorescence microscopy that enables the imaging of a large field of view in a sample with high resolution and contrast. It involves the use of a broad excitation light source to illuminate the entire sample, which is then captured by a sensitive camera. This makes it ideal for applications such as live-cell imaging and high-throughput screening [26].

SIM is a powerful super-resolution imaging technique that enables the visualization of fine details in biological specimens beyond the diffraction limit of light that restricts the resolution to about half the wavelength of light (200-300 nm). SIM works by illuminating the

sample with a series of light patterns; in the case used in this study, five sinusoidal patterns at three pattern orientation angles, as shown in Fig. 1. The illumination pattern modulates high-frequencies in the sample enabling them to be transmitted by the imaging system and then extracted computationally to reconstruct a super-resolution image. Gustaffson first introduced 2D-SIM in 2000 [27] and has since been further developed and optimized by several researchers. The traditional computational processing involves shifting and averaging the multiple images taken at various illumination angles, while other model-based approaches are shown to provide more accurate image restoration to remove the unwanted high-frequency information [28], [29].

A three-dimensional (3D) image captured using a conventional WF microscope can be modeled by applying a 3D convolution to the underlying object, o(x, y, z), with the point spread function of the microscope, h(x, y, z); resulting in d(x, y, z) as the observed raw image standing for "data". This equation is known as the 3D WF forward imaging model (Eqn. 1.1), where (x, y, z) are the coordinates in the space domain:

$$d_{WF}(x, y, z) = o(x, y, z) \otimes_3 h(x, y, z)$$

$$(1.1)$$

Fluorescence intensity in a 3D-SIM image recorded using axial scanning can be modeled as follows:

$$d_{SIM}(x, y, z) = [o(x, y, z).j(x, y)] \otimes_3 h(x, y, z).i(z),$$
(1.2)

where i(z) and j(x, y) are the axial and lateral functions in the SI pattern, respectively [28]. The goal of reconstruction in this study is to solve Eqn. 1.2 to recover the information from the ground truth. The deconvolution task of the Eqn.1.1 is also utilized for the physicsinformed unrolling, explained later in methodologies. As shown in the previous equation the images and PSF are functions of x, y, and z. However, in what follows, we suppress the dependent variables for simplicity, without loss of generality.



Figure 1. 3D-SIM imaging is used. (A and B) show the impact of the striped interference patterns on detection in frequency space using structured illumination (B) and regular WF illumination (A, respectively). (C) 3D-SIM allows the separation of displaced components and their restoration to their correct placement in frequency space by recording five phases of the sine wave pattern at each z point, extending the axial support while maintaining within the resolution limit. Additionally, the diffraction grating is rotated into three places that are 60 degrees apart from one another, producing support that is almost rotationally symmetric throughout a broader frequency space area. Three picture stacks are then captured. Compared to the original image, the 3D-SIM reconstruction shows more accurate image details. A rise in resolution is seen by the frequency space graphic on the right. [4]

1.3 Image Reconstruction

Image reconstruction refers to the process of creating a high-quality image from a set of lower-quality or incomplete image data. This is typically done through the use of mathematical algorithms and techniques, such as filtering and interpolation, to enhance and fill in missing information in the image. One application of image reconstruction is in medical imaging, where it is used to create high-resolution images from low-resolution scans. For example, computed tomography (CT) and magnetic resonance imaging (MRI) scans use mathematical algorithms to reconstruct 3D images of the body based on a series of 2D images. Another example of image reconstruction is in digital photography, where algorithms are used to enhance and improve the quality of images captured by digital cameras. This can include techniques such as denoising, sharpening, and color correction [30]. Inverse imaging is a term used to describe recovering or reconstructing the ground truth image ofrom the captured image d. The term "inverse" comes from the fact that these methods involve inverting the imaging process as in 1.1 or 1.2

Several classical methods such as Generalized Wiener Filter (GWF) [25] and iterative model-based methods [28] are known to solve the ill-posed inverse imaging problems in 3D-SIM microscopy, by dealing with the listed challenges. Image reconstruction methods are aimed at enhancing the quality of the image by reducing noise, filling in missing information, and improving image resolution. The Wiener filter is a linear filter that aims to minimize the mean square error between the noisy input signal and the estimated signal. It achieves this by using a statistical approach that considers the statistical properties of the signal and noise to estimate the desired signal. The filter is designed using the power spectral density of the noisy signal and the signal of interest. The Wiener filter has been applied in various fields such as image denoising, speech enhancement, and channel equalization. It has been shown to be effective in reducing noise while preserving important features of the signal, such as edges and sharp transitions. The GWF is an extension of the classic Wiener filter that allows for non-linear, non-Gaussian noise and signal models. The

GWF uses a Bayesian approach to estimate the signal of interest by considering the prior knowledge of both the signal and noise models [31].



Figure 2. SIM super-resolution compared to Wide-Field Microscopy: a) 15 Raw SIM Input Images (5 phases, 3 orientation angles of SI pattern). b) SIM reconstruction using GWF: Frequency component decomposition and filtering in Fourier space. c) Wide-Field Image: Sum of 15 raw SIM images. d) SIM provides approximately 2x resolution improvement over wide-field microscopy [1].

1.3.1 Image-to-image translation

Image-to-image translation refers to the task of converting an input image from one domain to another while preserving relevant visual characteristics [32]. It involves mapping images from a source domain to a target domain, such as transforming images from grayscale to color or converting sketches to realistic images. Image-to-image translation has gained significant attention in the field of computer vision and has been addressed using various techniques.

One popular approach is the use of generative adversarial networks (GANs), which consist of a generator and a discriminator network. The generator generates transformed images in the target domain, while the discriminator tries to distinguish between the generated images and real images from the target domain. The generator and discriminator are trained together in an adversarial manner, where the generator aims to generate realistic images that fool the discriminator, and the discriminator aims to correctly classify real and generated images.

1.3.2 Image Super-Resolution

Optical image super-resolution and digital image super-resolution are two different image-to-image translation techniques used to enhance the resolution of images.

Optical image super-resolution

Optical image super-resolution refers to the use of physical optics to improve the resolution of an image captured by an optical system, such as a microscope or a telescope. This can be achieved by using techniques such as deconvolution or adaptive optics, which can reduce the effects of aberrations and improve the overall resolution of the system. Image super-resolution in SIM refers to the computational process of reconstructing a highresolution image from multiple lower-resolution images obtained through structured illumination [33]. The structured illumination patterns used in SIM produce a series of images that are then processed using algorithms to extract high-frequency information and generate a super-resolved image [25]. The resulting image has a resolution beyond the diffraction limit, allowing for the visualization of structures that would otherwise be invisible using conventional microscopy techniques [33].

Digital image super-resolution

Digital image super-resolution, on the other hand, refers to the use of computational algorithms to enhance the resolution of a digital image [34]. This technique is used in various fields, including medical imaging, satellite imaging, and surveillance imaging. The image super-resolution process involves the use of computational algorithms to generate a high-resolution image from a low-resolution image. One common approach is to use deep learning methods, such as convolutional neural networks (CNNs), to learn the mapping between low-resolution and high-resolution images [35]. Another approach involves the



Figure 3. Super-Resolution in 3D SIM. The passage describes the observable regions for various types of microscopes, including WF, SIM using two or three illumination beams, and their corresponding spatial-frequency components. It also highlights the highest possible spatial frequencies that can be generated by illuminating through the objective lens in the case of the three-beam structured illumination microscopy [5].

use of statistical methods, such as sparse coding and dictionary learning, to generate a highresolution image from a set of low-resolution images [34]. However, classical methods cannot always address all the issues at once with sufficient accuracy, due to unknown or not observable system parameters required by an analytical model-based computational approach, and some can be computationally intensive.



Figure 4. Digital Image Super-Resolution [6]

While both techniques aim to improve the resolution of an image, they operate at different stages of the imaging process and have distinct limitations and applications. Optical super-resolution is typically limited by the physical properties of the imaging system and is more suited for specialized applications such as microscopy or astronomy.

1.4 Challenges

There are different challenges in the area of SIM microscopy imaging. Finding a microscopic dataset for deep learning applications can be challenging for several reasons. First of all, capturing microscopic images is time and energy-consuming, and for a deep network, hundreds or thousands of samples of captured images are required. These samples also need to be paired with their ground truth because they are usually blurry, of low resolution, under low light conditions, etc. There is a possibility of using a non-microscopic dataset such as DIV2K (DIVerse 2K resolution high-quality images) [36] with 1000 samples of low-resolution and high-resolution images. The network can be trained on nonmicroscopic data and later tested on microscopic images. However, there are risks in not using microscopic data, and there are significant differences between non-microscopic and microscopic data. Thus, it will be better if a microscopic dataset is used for image restoration. If you train a deep neural network (DNN) for microscopy image reconstruction using a non-microscopy dataset, there are several risks that you should be aware of:

- Domain shift: The non-microscopy dataset may be very different from the microscopy dataset in terms of the visual characteristics of the images, such as resolution, texture, contrast, and illumination. As a result, the DNN may learn to recognize and exploit features that are specific to the non-microscopy dataset but not relevant to the microscopy dataset, leading to poor performance and generalization [6].
- Overfitting: If the DNN is trained on a non-microscopy dataset that is too small or too simple, it may memorize the dataset and fail to generalize to new, unseen microscopy images. This can lead to overfitting, where the DNN performs well on the training set but poorly on the test set [37].
- Biases and artifacts: The non-microscopy dataset may contain biases or artifacts that are not present in the microscopy dataset, which can lead to the DNN learning spurious correlations that do not generalize well to microscopy images. For example,

if the non-microscopy dataset contains images with a certain color balance or noise pattern, the DNN may learn to rely on these features instead of learning the true underlying structure of the microscopy images [38], [39].

• Ethical concerns: If the non-microscopy dataset contains sensitive or personal information, such as medical records or faces of individuals, there may be ethical concerns about using the dataset for training a DNN without proper consent and privacy safeguards.

Another challenge in SIM is the low signal-to-noise ratio, which can limit the resolution and quality of the reconstructed image. Several algorithms have been developed to improve the signal-to-noise ratio, such as the Maximum-Likelihood SIM (ML-SIM) algorithm, which uses statistical methods to enhance the quality of the reconstructed image [40].

1.5 Motivation

To overcome these issues, it is necessary to come up with new algorithms to reconstruct the SIM images accurately. Here are a few reasons why it is important to develop new algorithms for SIM image reconstruction:

- Improved resolution: The main advantage of SIM is its ability to enhance the resolution of the images beyond the diffraction limit. However, the reconstruction algorithm used plays a critical role in achieving this goal. New algorithms that can better suppress artifacts, noise, and other factors can lead to even better resolution and better visualization of biological structures.
- Better accuracy: Accurate image reconstruction is essential for quantitative analysis of biological structures. With better reconstruction algorithms, we can achieve higher accuracy in measuring biological parameters, such as the size, shape, and orientation of cellular structures.
- Reduced acquisition time: SIM imaging typically requires multiple image acquisitions, which can be time-consuming. By developing new algorithms that require

fewer acquisitions, we can reduce the overall acquisition time and improve the throughput of the imaging process.

 Compatibility with different biological samples: Different biological samples require different imaging parameters, such as the illumination pattern and frequency. New algorithms that are adaptable to different samples and imaging parameters can improve the versatility and applicability of SIM imaging.

1.6 Contribution

When working with general datasets, which include images of nature and objects, it is possible that there is no consistent source of low-resolution images in all instances. Due to this variability, deep learning (DL) has emerged as the most effective method for image reconstruction [41]. On the other hand, by utilizing a dataset generated from a SIM system, we can obtain a reasonable approximation of the point spread function in the SIM model. However, it is important to note that the analytical model is still an estimation rather than an exact function. Additionally, classical deconvolution methods have limitations, and there may be noise present in the data. Therefore, it is advisable to employ both the known parameters in the model and deep learning techniques to address the challenges arising from poor estimation, partially known or unknown parameters, and noise.

In this thesis project, it is expected to achieve visually and computationally enhanced resolution by converting a regular deep super-resolution algorithm to a generative adversarial super-resolution algorithm. Another expectation is to receive even more enhanced images by unrolling the network using physics-informed terminology. Furthermore, we want to see how training on two-dimensional images versus three-dimensional volumes is different. We plan to answer which one is preferred in the case of 3D imaging. Achieving more efficient training is another purpose of this project by improving the physics-guided aspect of the algorithm.

Chapter 2

Background and Related Work

The development of physics-guided deep learning has revolutionized the field of microscopy super-resolution. Physics-guided deep learning incorporates physical principles into deep learning algorithms to achieve better accuracy and generalization. One crucial aspect of physics-guided deep learning in microscopy super-resolution is the availability and quality of datasets. High-quality big datasets are essential for training deep learning models effectively. In recent years, several datasets have been developed for this purpose. The use of these datasets in physics-guided deep learning has led to significant improvements in microscopy super-resolution [21]. As follows, several studies will be reviewed in case of finding the required datasets and possible deep neural network (DNN) architectures to train and evaluate the datasets with them. For the current thesis project, we have investigated multiple different datasets and deep learning models. Here we present some of those works related to our task. In this chapter, we first discuss our findings of datasets, what various types of datasets can be more beneficial for our project, and what the final datasets selected to complete this project are. The next section, explains the background knowledge and related work using deep neural networks for image reconstruction and specifically, digital image super-resolution.

2.1 Datasets

Considering the data capturing challenges discussed in the previous chapter, it is generally recommended to use a microscopy dataset for training a DNN for microscopy image reconstruction, or at least to fine-tune a pre-trained DNN on a non-microscopy dataset using transfer learning techniques to mitigate the risks mentioned above. Different datasets for this project have been reviewed. Three types of microscopes were selected as the potential datasets more appropriate for the purpose of this project. IDR [42] includes several microscopic datasets. Providing four studies regarding wide-field microscopy images, 44 studies for confocal microscopy images, and seven studies for structured illumination microscopy. The biggest number of studies are related to confocal microscopy. However, regarding the fact that confocal microscopes take images using optical sectioning [43], their output seems to be pretty clear. Meaning that their images are of high resolution. Therefore, there will be not much need for image restoration. Comparing wide-field and SIM datasets, restoring SIM images are of higher value for us as the extra data results in doubling the spatial resolution in comparison with WF microscopy [44] [5]. Moreover, SIM will give us a higher range of spatial frequencies. In this way, the optical transfer function can cover a greater region. In contrast with confocal microscopic images, SIM images still need reconstruction. As a result, it will be best if we stick to the SIM datasets.

2.1.1 SIM datasets

One of the available SIM datasets is FairSIM [1] with five 3D-SIM volumes from a biophotonic group. The biggest problem regarding this dataset is that they are too few. For deep learning training, we need much more image samples than only five. A possible solution for this issue is to use data augmentation such as cropping, flipping, rotating, etc. By doing so, it is possible to increase the number of samples. In SIM reconstruction, however, one should be cautious not to lose critical information by applying augmentation techniques to the original images.

Another dataset has recently been published by Opstad et al. [2] from this point on refereed to as "Fixedcell". The article describes the methods and procedures used to obtain high-resolution 3D images of mitochondrial dynamics and lysosomal function in live and fixed H9c2 rat cardiomyoblast cells. The images were acquired using a DeltaVision OMX V4 Blaze imaging system and reconstructed using the manufacturer-supplied softWoRx program. The data is available in the DataverseNO repository in the UiT Open Research Data collection. The 3DSIM data were of high quality and suitable for use in the development of SIM reconstruction algorithms. Considering the publication date and the variety of the dataset samples, it is a perfect candidate for the purpose of our study. For simplicity, we are using 18 samples of the fixed cell mitochondria. Luhong's et al. [20] provides an overall of 558 sample images that can be sufficient for training. These images are 2D slices of a 3D-SIM. They include images of the samples under high and low exposure to light. The ground truth is obtained by reconstructing them using a generalized wiener filter (GWF). As one of the goals of this project is to compare 3D and 2D results, even though this dataset has been used in our previous studies, it is not being used for this thesis project, and we will only focus on images that offer 3D volumes for training.

2.1.2 Simulated datasets

Another way to achieve the required number of samples is to simulate data. The use of simulated data for microscopy image super-resolution poses several challenges that may affect the generalization of the method to real data. Some of these challenges are: 1. Lack of variability in the data: Simulated data may not fully capture the variability and complexity of real microscopy data, which could lead to overfitting of the model to the simulated data and poor generalization to real data. 2. Lack of realistic noise and artifacts: Simulated data may lack the noise and artifacts present in real microscopy data, which can affect the ability of the model to generalize to real data. Modeling the realistic artifacts is a possibility, but aggravates the complexity of the problem. 3. Mismatch between the simulated and real data distributions: The distribution of the simulated data may not perfectly match the distribution of real microscopy data, leading to poor generalization and performance. 4. Difficulty in accurately simulating complex biological structures: The simulation of complex biological structures, such as organelles and subcellular structures, can be challenging and may not fully capture their true complexity. 5. Difficulty in simulating realistic imaging conditions: Simulating realistic imaging conditions, such as different wavelengths, photobleaching, and photo-toxicity effects, can be challenging and may not fully represent the true imaging conditions [45], [46].

One 3D-SIM star-like object is accessible by Van and Preza [28]. Boland et al. [7] also provide two 3d-SIM simulation studies with their open-access codes. One study builds

spherical point clouds, and the other one prepares 24 withheld chromatin images. Their study also uses a generalized wiener filter to get the reconstructed images as their ground truth. By varying the number of points in the simulation, it is possible to generate multiple samples with different levels of light. Furthermore, the Poisson noise level can be changed Boland et al. used their method to train a Residual Channel Attention Network (RCAN, will be further explained in the following section); the simulated SIM image stacks were used to train the model, while an up-sampled confocal picture served as the ground truth [7].



Figure 5. Poisson noise's impact on sample simulated chromatin architectures [7].

2.2 Deep Learning Background

To have a better understanding of Deep Learning (DL) in the first place, we are going to review the definition of a system in general. A system is composed of interconnected components that take inputs and produce one or more outputs. Traditionally, these systems were described using analytical or physical equations that explained the relationship between the components and the inputs/outputs. An example of a feedback system is shown in Fig. 6. where analytical solutions need to be found for components A and B. Knowing the analytical model of a system can help understand, explain, and control it. However, in some cases, finding the analytical model can be difficult, inaccurate, or impossible. This is where DL come in - they can find the analytics themselves, making them useful in situations where traditional analytical methods fall short [47].



Figure 6. A simple feedback system: for the forward direction, "A" corresponds to the analytical transfer function indicating how outputs are driven from the given inputs, and "B" explains the analytical formula of how the output is observed by sensors, etc. then, returned as an addition to the input of the system [8].

DL has revolutionized the world of modeling. Its purpose is to consider the system as a black box. By feeding the system with tones of various data samples as input, then capturing their outputs, it tries to learn the analytical relation between inputs and outputs using back-propagation. For each iteration, the current output of the improvised model is compared with the expected output of the system known as ground truth. Ground truth has previously been labeled for its specific given input. The backpropagation process then updates the model weights to minimize the difference between the current system's output and the expected output known as the loss function. Finally, after testing the model with unseen data samples, the analytic is ready to be used as the physical description of the model. As follows, we will discuss different Deep Learning algorithms [48]. These algorithms not only review the potential methods to apply image reconstruction, but they also give a background of the networks that help us further understand the finalized UPIGAN, the DL architecture introduced in this project. Image reconstruction is a crucial task in various fields, and there are several methods available to solve it. As follows, we explore a few state-of-the-art reconstruction methods using DL.

2.2.1 Fully Connected Neural Networks (FCN)

In Fig. 7. a fully connected neural network (FCN) is seen. An FCN, also known as a multi-layer perceptron (MLP), is a type of artificial neural network where all neurons in one layer are connected to all neurons in the next layer. This allows the network to learn complex non-linear relationships between inputs and outputs. The architecture of a fully connected neural network typically consists of an input layer, one or more hidden layers, and an output layer. Each layer contains a set of neurons that apply an activation function to the weighted sum of inputs from the previous layer. During training, the network adjusts the weights and biases of each neuron to minimize a loss function, such as mean squared error or cross-entropy, using backpropagation [49].



Figure 7. An example of a Fully Connected Deep Neural Network [9].

2.2.2 Convolutional Neural Networks (CNNs)

Another commonly used method for image reconstruction is the use of convolutional neural networks (CNNs). CNNs have been shown to be effective in solving various image reconstruction problems, such as super-resolution, and denoising. For example, in medical imaging, CNN-based methods have been used for brain MRI reconstruction [50]. Convolutional neural networks (CNNs) are a type of neural network that are designed to process data with a grid-like topology, such as images or videos. They use convolutional layers, which apply a set of filters to the input data to extract features at different spatial locations.

Pooling layers are also used to reduce the spatial dimensions of the feature maps while preserving the most important information.

CNNs have become the state-of-the-art method for various computer vision tasks. They are more common than fully connected neural networks because they can exploit the spatial structure of the input data and capture local patterns and correlations between neighboring pixels. This makes them more efficient and effective at learning complex features from high-dimensional inputs, while reducing the number of parameters and computation required [49] [37].



Figure 8. A example of a typical Convolutional Neural Network (CNN) [10].

2.2.3 Super-Resolution Convolutional Neural Network (SRCNN)

The Super-Resolution Convolutional Neural Network (SRCNN) is a deep learning method for single image super-resolution, which aims to recover a high-resolution image from a low-resolution input. SRCNN was introduced in the paper "Image Super-Resolution Using Deep Convolutional Networks" by Dong et al. in 2014.

SRCNN uses a deep neural network with three convolutional layers to learn a mapping between low-resolution and high-resolution image patches. The network takes as input a low-resolution image patch and outputs a high-resolution patch of the same size. During training, the network is trained to minimize the mean squared error between the predicted high-resolution patches and the ground truth high-resolution patches.

SRCNN has shown superior performance compared to traditional image super-resolution methods, such as interpolation-based and sparsity-based methods. The learned filters in the

convolutional layers capture complex image features, such as edges and textures, and are able to generate visually pleasing high-resolution images [35].



Figure 9. Convolutional Neural Networks architecture for single image super resolution [11].

However, CNNs are simple, and a complicated network might look more attractive and accurate at first sight, they are capable of building complicated and strong networks. Label2Label [51] is a new architecture introduced to restore the cellular structure in fluorescence microscopy. This study claims that they can assess the performance of a Cycle-GAN [52].

2.2.4 UNET

The U-Net structure was first introduced to apply image segmentation to biomedical images. Its main application is still to perform segmentation. [12] However, its structure helps training more accurately. Therefore, it can be used for more general applications, making the network popular in various fields. By looking at Fig. 10. the significance of the network is shown. In the beginning, the network is down-sampling the input image to extract features. Then, while up-sampling the features to achieve the desired output, some layers are concatenated with the up-sampling layers from the beginning stages. This helps to bring direct information from the input image to ease the reconstruction. U-Net is also capable of training using fewer training samples than other networks. [15]

The explained structure recently is being used for applications beyond just segmentation by choosing the corresponding ground truth and making some enhancements in the network design. Several studies are using U-Net for image restoration applications. MIMO-U-Net [13] is one of the enhanced architectures for image deblurring using U-Net to deconvolve



Figure 10. Original U-Net Architecture for biomedical image segmentation [12].

blurred images. This architecture is later enhanced to have even better performance on deblurring with deep residual Fourier transformation [14].



Figure 11. MIMO U-Net architecture used for image deblurring [13].

These data were tested on general blurry datasets such as GoPro [53] with more than 3000 blurred images and their reconstructions. But, they are not trained and tested on microscopic data. Among the studies concentrating on microscopy image restoration, Luhong et al. [20] propose four different U-Net-based architectures, including U-Net-SIM15, U-



Figure 12. Deep Residual Fourier Transformation deblurring based on MIMO-U-Net [14].

Net-SIM3, scU-Net, and U-Net-SRRF5 for image super-resolution under low light conditions. This architecture is trained on 2-dimensional images from structured illumination microscopy. Other than that, using a simple U-Net structure, Lin et al. [15] claim that they are able to perform segmentation, localization, super-resolution, denoising, and deblurring all using the same architecture given in Fig. 13. on their atom data from atomic-resolution scanning transmission electron microscopy (STEM).



Figure 13. U-Net structure used for image segmentation and restoration on atom images [15].

2.2.5 Residual Channel Attention Network (RCAN)

As Residual Channel Attention Network (RCAN) [16] has been utilized in our introduced architecture further discussed in this thesis, it can be a good idea to have a short survey of what it has to offer. RCAN applies single image super-resolution (in contrast with imaging communities, in the deep learning community super-resolution means achieving double the spatial resolution). Fig. 15. shows its main parts. Shallow feature extraction happens in a simple convolutional layer. Its output is followed by a "residual in residual" deep feature extraction module, including multiple residual groups. Finally, the output of the deep feature extraction module gets up-sampled to provide a high-resolution result. Each residual group consists of multiple residual channel attention blocks concatenated together. This architecture is known for super-resolution applications. However, originally, it was not applied to any microscopy or medical data, it has later been used by several microscopy papers.



Figure 14. RCAN Architecture, consists of an array of residual groups following a convolution layer and followed by the up-sampling layer and another convolution layer. Each residual group itself includes an array of residual blocks in it followed by a convolutional layer [16].

Soft attention, according to their study, occurs when the context vector is calculated as a weighted sum of the encoder's hidden states. Channel attention, in this case, is more flexible than SR-CNN approaches for the real data [16].

Residual networks are the ones that connect the output of one layer not only to its next layer's input but also to another layer's input. Skipping some connections helps deal with the gradient vanishing problem while learning the network's weights. It also helps to avoid saturation of the accuracy [17]. Besides, a channel attention network is a deep learning model that focuses on individual channels with soft attention [54].



Figure 15. Residual Network [17].

The RCAN method is later utilized by various studies concentrating on microscopy image restoration. For example, it is used in image sharpening and denoising of fluorescence microscopy [55]. Boland et al. in [7] present a way to rebuild 3D SIM image stacks with twice the axial resolution possible through standard SIM reconstructions by utilizing current developments in image up-scaling through an RCAN deep learning model. They additionally show that their approach is noise-resistant and test it against two-point situations and axial gratings. Eventually, they talk about how the approach may be modified more to boost resolution [7].

2.2.6 Generative Adversarial Netowrks (GANs)

So far, the networks examined to optimize the pixel difference between predicted and output HR pictures. Although this measure works well, it is not perfect; humans discriminate images based on perceptual quality rather than pixel difference. Generative models (or GANs) attempt to optimize perceptual quality in order to create pictures that are pleasing to the human eye [56]. GANs have become an important tool for image super-resolution due to their ability to generate high-quality and realistic images. Specifically, GAN-based methods such as SRGAN have shown significant improvements in terms of visual quality and perceptual similarity to the ground truth image compared to traditional methods [18]. In addition, GAN-based methods are also more flexible and can be easily adapted to different types of images and resolutions. Such as "different upscaling factors, e.g., from 2× to 8×, and various image formats including gray-scale, RGB, and hyperspectral images" [16]. Therefore, studying GANs for image super-resolution is important for advancing the field of image processing and enabling a wide range of applications in various domains [57].

Conditional GAN (CGAN)

CGAN (Conditional Generative Adversarial Network) is a type of GAN that allows for the generation of samples conditioned on additional information, such as class labels or input images. The generator takes both a noise vector and the additional information as input, while the discriminator is trained to distinguish between real and fake samples based on both the generated image and the additional information. GAN architectures discussed in this thesis can be considered as CGANs since the ground truth is present in all methods [58].

Super-Resolution GAN (SRGAN)

To produce higher resolution images for super-resolution, GAN prepares a deep network in conjunction with an adversary network. Fig. 16. shows that compared to a similar design without GAN, SRGAN (Super-Resolution Generative Adversarial Networks) is more appealing to the human eye with more details. SRGANs use a generator network to upscale the low-resolution image and a discriminator network to distinguish between the generated high-resolution images and real high-resolution images.

The SRGAN architecture was proposed by Ledig et al. in 2017 and showed significant improvements over traditional single-image super-resolution methods. The authors introduced a new loss function, called perceptual loss, which incorporates both a content loss and an adversarial loss. The content loss ensures that the generated image is semantically similar to the ground truth image, while the adversarial loss encourages the generator to produce visually realistic images [18].



Figure 16. "From left to right: bicubic interpolation, deep residual network optimized for MSE, deep RGAN optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4x upscaling]" [18].

Pix2Pix

The pix2pix algorithm is a deep learning model used for image-to-image translation tasks. It was proposed by Isola et al. in their paper "Image-to-Image Translation with Conditional Adversarial Networks" [32]. The algorithm employs conditional generative adversarial networks (CGANs) to learn the mapping between an input image and an output image using a modified U-Net. It leverages a generator network that generates the output image from the input and a discriminator network that tries to distinguish between the generated output and the real output.

The pix2pix algorithm uses a loss function that combines an adversarial loss, which encourages the generated output to be realistic, and a pixel-wise loss, which ensures the similarity between the generated output and the real output. This combination allows the model to produce high-quality and visually coherent output images.

The pix2pix algorithm has been widely adopted in various image translation tasks, such as image colorization, image segmentation, and style transfer. It has shown impressive results in generating realistic and visually appealing output images that preserve the structural characteristics of the input.

Channel Attention GAN (CAGAN)

Bringing up the idea of using channel attention networks for image restoration, besides using GANs for the same application, another study proposes to combine the two to reach out to an even stronger network [19]. In their study, a conditional GAN is being used and trained by combining pixel-wise loss and discriminative loss. It claims that conditioning the losses of both the discriminator and the generator will result in higher performance in specifically, biological image restoration.

Using channel attention structure in the GAN's generator, Qiao et al. introduce the architecture in Fig. 17. They use this network to super-resolve 3D-SIM images using 3D processing. Their work is also the major inspiration for the current thesis project [19].



Figure 17. Network architecture for the proposed caGAN [19].

caGAN-SIM uses 15 times less signal intensity than the standard technique while yet achieving equivalent or superior reconstruction quality. Using the example of dynamic interactions between microtubules and lysosomes in live cells, better effectiveness of caGAN-
SIM is presented for diverse subcellular structures and its capability in long-term multicolor 3D super-resolution imaging [19].



Figure 18. The input performance of the mitochondrial inner membrane using various forms of raw pictures is compared. (a) A single wide-field raw picture volume slice. (b) GT-SIM image, which is a 3D-SIM image that was created using a conventional algorithm from raw data with high SNR. First column, the caGAN-SIM-WF picture was created from a single wide-field image volume. (15 raw SIM image volumes) were used to rebuild the caGAN-SIM picture in the second column. In third and forth columns the scU-net [20], and the caGAN-SIM image were compared respectively. The caGAN-SIM-RP was created using fewer phase pictures and comprised of three raw SIM image volumes plus a WF image volume [19].

The task of RCAN architecture in this project is to play the role of the generator, in which the super-resolution results are generated. What happens in the generator, is pretty much similar to the procedure of the study in [7]. The difference is the usage of a discriminator in addition to the physics-guided section. The discriminator is a simple convolutional classifier. Its task is to classify the output of the generator as real ground truth or generated ones. If the generator succeeds to trick the discriminator to make the wrong decision in half of the situations, this means that the discriminator is performing randomly and the

generator is doing a good job in generating super-resolved images. The loss function of this network includes three terms as described in Eq. 2.1

$$L_G | D(\hat{Y}, Y) = L_{sr}(\hat{Y}, Y) + \alpha L_{WF}(\hat{Y}, Y_{WF}) + \beta L_D(\hat{Y})$$
(2.1)

Where G and D stand for the generator and discriminator respectively, Y is the ground truth and \hat{Y} is the reconstructed image, and $L_{WF} = ||Y_{WF} - \hat{Y} * PSF||_2$. In the later equation, mentioned as the physics-guided term, Y_{WF} stands for the true input WF image which is the average of SIM frames along all phases and angles for each z slice. α and β are the weights of the losses beside them. The physics-guided terms take into account the effect of the wide-field point spread function on the reconstructed image in comparison with the original WF input.

The Residual Channel Attention Network (RCAN) was originally designed for 3-color channels and made use of the relationship between distinct color channels for super-resolution. 3D-SIM datasets consist of images acquired using different phases/angles of the structured illumination pattern, which can be thought of as different channels. Thus, RCAN sounds like a suitable candidate for SIM reconstruction. The loss function in the caGAN architecture includes the known widefield point spread function (PSF) of the optical system to compare the simulated outputs from the restored images with the input images. However, in the proposed UPIGAN, we replace the loss function with the common loss function in-troduced initially by the GAN paper as the "min-max" loss function [57].

The other significant difference between the proposed UPIGAN and the caGAN is the unrolling step; the network is also unrolled by the optical transfer function information. To clarify, caGAN adds a physics-informed term to the loss function mentioned previously as $L_{WF}(\hat{Y}, Y_{WF})$. In the next section, the physics-informed deep learning is going to be introduced. The proposed Unrolled Physics-Informed Generative Adversarial Network (UPIGAN) replaces the term $L_{WF}(\hat{Y}, Y_{WF})$ with a physics-informed unrolling architecture that will be deeply discussed in the next chapter. The basic reason for this replacement

is that there is an inconsistency in the implementation of the $L_{WF}(\hat{Y}, Y_{WF})$ in the original Python program provided by the authors with the equation provided in the paper. Initially, the inconsistency mentioned was not readily apparent or easily discernible. The fact that the paper in question was published in a renowned and respected academic journal further contributed to the lack of doubt regarding the accuracy of the implementation. Given that deep learning, as a field, is still not entirely explainable or transparent, the results presented by the authors seemed satisfactory and yielded acceptable outputs. However, upon closer examination, it became evident that the implementation diverged from the mathematical propositions and theories put forth in the paper. To delve into the specifics of this inconsistency, a comprehensive and detailed explanation can be found in the Appendix chapter, which provides a thorough analysis of the deviation between the stated methodology and the actual implementation.

2.2.7 Physics Informed DL

Recently, DL methodologies have been proposed to solve the inverse imaging problem without knowing the analytical model, and hence they can be biased toward the trained data. Although Deep Learning models, if trained properly, are reliable enough and even more accurate than the classical model-based method, they have their own downsides. DNNs can learn unknown parameters to describe a model. However, they require high memory, high computational usage, and a vast amount of data. Providing these three are time, energy, and money-consuming. In order to learn the relation of inputs and outputs, a big amount of input data has to be fed to the model, there are overwhelming challenges to gathering too much data on many specific applications. After gathering all the required data, they need to be kept in memory. Furthermore, depending on the complexity of the network, it also consumes memory. All the model calculations have to be performed by a high-performance computer and the process usually takes a considerable amount of time. As a result, there are different approaches presented in this article that prefer to combine a DNN with an analytical model. The reason is that in many situations, only a few parameters of the model

are unknown. Thus, it is a good idea to only learn the unknown parameters, and keep the known parts to be calculated classically. Or on the other hand, if the model includes only a few known parameters, the known parts of the DNN can be replaced with the known analytic [22].



Figure 19. Recurrent physics-informed ML engine. Where $N = 1 - \alpha . H^T . H$, and alpha is a small step size [21]

One way is to follow the physical model by a recurrent network. The input is the raw intensity measurement. α indicates the step size. In the cascaded method (Fig. 20.), the recurrent architecture is sequenced in m infinite number of steps. The difference here is that N does not return as current network input, but it is passed to the next step. H^T in both figures 19. and 20. corresponds to the transpose of the linear physical model. In the case of non-linearity, the inverse needs to be approximated. This method is also known as the "Deep unfolding" algorithm [22]. This algorithm is commonly used in knowledge-based methods that are combined with deep learning.



Figure 20. Cascaded physics-informed ML engine [21].

Presented a sample result from 20. in Fig. 21. the input improvement can be easily seen in the result of passing the under-sampled input to 5 sequential cascaded networks. As time passes, one more level of enhancement in the resolution is presented.

2.2.8 Unrolling

The unrolling algorithm involves expanding the iterative computation of a given algorithm over multiple time steps or iterations, effectively creating a longer but fixed computation graph that can be more easily optimized using traditional gradient descent methods. This approach can be used to improve the convergence rate and accuracy of iterative algo-



Figure 21. Sample image resolution enhancement using cascaded physics-informed ML engine. From left, the first is the raw input image. Then, each image corresponds to the output of first, second, third and fourth layers of the cascaded network [21].

rithms for a wide range of signal and image processing applications, including compressed sensing, image deblurring, and image denoising [41].

In the context of deep neural networks (DNNs) and generative adversarial networks (GANs), "unrolling" refers to the process of unfolding or expanding the iterative computation of the network over multiple time steps or layers. This approach allows for the efficient computation of gradients during backpropagation, which is used to optimize the parameters of the network. Unrolling is commonly used in optimization methods for training DNNs and GANs. The basic idea is to unroll the iterative computation of the network, creating a longer but fixed computation graph that can be more easily optimized using traditional gradient descent algorithms. In addition to improving optimization efficiency, unrolling can also be used to improve the stability and convergence of training algorithms for DNNs and GANs. By unrolling the computation graph over multiple time steps or layers, the network can more effectively capture long-term dependencies and better approximate the true data distribution [59].

Unrolled GAN

The "Unrolled Generative Adversarial Networks" paper proposes a modification to the training algorithm for GANs that involves "unrolling" the iterative computation of the dis-

criminator network. The authors argue that this modification can improve the stability and convergence of the GAN training process, by making it easier to optimize the discriminator network. The unrolled GAN algorithm involves computing multiple steps of the discriminator network for each step of the generator network, effectively creating a longer but fixed computation graph that can be optimized using traditional gradient descent methods. The authors demonstrate that this modification can lead to improved performance on a variety of image generation tasks, including generating high-resolution images from low-resolution inputs. Overall, the paper provides a novel approach to improving the training stability and quality of GANs, and has been cited extensively in subsequent research on GANs and related generative modeling techniques [59].

Physics Informed Unrolling

Model-Based Deep Learning involves incorporating prior knowledge about the underlying data model into the training process. The model-based DNN algorithm involves formulating a mathematical model for the data generation process, which can then be used to design an appropriate architecture for the DNN. By incorporating prior knowledge into the training process, it is possible to improve the accuracy and generalization performance of DNNs, particularly in cases where the available training data is limited or noisy [41].



Figure 22. Algorithm unrolling [22]

The unrolling method in the current thesis project utilizes deep neural networks as prior knowledge to guide the restoration process using the unrolled optimization with deep priors algorithm (UODP) [60]. The UODP formulates an optimization problem as a series of iteratively-refined neural network architectures, with each iteration serving to refine the prior knowledge that is used to guide the optimization process. This approach can lead to faster and more accurate optimization, particularly in cases where the optimization problem is ill-posed or noisy. The major reason to select this algorithm for this project is that the physics-informed block used in this algorithm is a very similar approach to the classical wiener filter algorithm for image reconstruction.



Figure 23. UODP

2.2.9 Conclusion

The background provided can help us better articulate the general steps required to prepare and implement the UPIGAN architecture listed below:

- 1. Prepare the dataset: The first step is to prepare the structured illumination microscope dataset for super-resolution. This involves collecting, pre-processing, and partition-ing the data into training, validation, and testing sets.
- 2. Build the generator: The generator is responsible for up-sampling low-resolution images to high-resolution images. In this case, we'll use an unrolled residual channel attention GAN. This involves building a neural network that can learn how to generate high-quality images.
- 3. Build the discriminator: The discriminator is responsible for distinguishing between real and generated images. It's used to train the generator by providing feedback on

how well the generator is doing. We'll use a convolutional neural network for the discriminator.

- 4. Train the GAN: We'll train the GAN by alternating between training the generator and discriminator. This involves feeding low-resolution images into the generator, up-sampling them, and then passing them through the discriminator to see how well they match the high-resolution images in the training set. We'll use loss functions to measure the difference between the generated and real images.
- 5. Test the model: Once the model is trained, we'll use the testing set to evaluate its performance. We'll measure the PSNR and SSIM of the generated images and compare them to the ground truth images.
- 6. Fine-tune the model: If the model doesn't perform well on the testing set, we can finetune it by adjusting the hyper-parameters, changing the architecture of the generator and discriminator, or modifying the training process.

The mentioned steps will be further explained in the upcoming chapters of this thesis document. The results inform us how classical methods can be combined with or replaced by deep learning modeling. They also provide information about the dependency of the relevance of the images to the training procedure.

Chapter 3

Data Processing

3.1 Data Preprocessing

The major problem regarding most datasets including the introduced datasets above is that they have too few samples. For deep learning training, much more image samples are needed. A possible solution for this issue is to use data augmentation such as cropping, flipping, rotating, etc. By doing so, it is possible to increase the number of samples. In SIM reconstruction, however, one should avoid losing critical information by applying augmentation techniques to the original images. In this section, it is explained how these samples have been cropped into smaller 2D and 3D patches to apply data augmentation and reduce volatile memory usage. However, one should keep in mind not to crop the images into such small patches that it will be hard to determine the details of the images anymore.

3.1.1 Crop to Patches

Regarding the selected patch size, the patches are cropped from the original image with the lowest possible overlap. Thus, the number of cropped sections would be equal to:

$$quotient(\frac{original_pixel_size}{patch_size}) + 1$$
(3.1)

The overlap depends on the division ratio. Fig. 24. shows how the patches are cropped. The SIM images are stored in 3D volumes of $(x, y, angles \times z \times phases)$. In this program, it is essential to keep in mind that the cropping in the z-axis should only crop the z-planes and keep all the angles and phases of each z-plane separately. Thus, the 3D volumes are converted to 5D volumes of (x, y, z, angle, phase). After completing the crop, the images are reordered to the standard 3D volumes of $(x, y, angles \times z \times phases)$ and stored for the main project.

To accomplish the purpose of this project, utilizing a microscopic dataset for image restoration is essential, owing to the significant disparities between non-microscopic and microscopic data. [38, 39]. One of the available SIM datasets is FairSIM [1], with five



Figure 24. Crop the input images of sizes of 512x512xvariable depths to 128x128x3 using the crop2patches function.

3D-SIM volumes from a biophotonic group. The 3D data available from "FairSIM" comes from different SIM systems. The ones selected for this project come from a 3D SIM system with a wavelength of 525nm making the number of samples even smaller (three samples). One sample is shown in Fig. 25.

Furthermore, it's crucial to avoid combining all of these photos and then randomly dividing them since, in certain situations, they overlap or are clipped samples of the same item, which might lead to a strong training but a poor generalization.

The raw SIM samples with various input dimensions are cropped to $128 \times 128 \times z \times 3 \times 5$ where 3 is the number of phases and 5 is the number of angles (also mentioned as orientations). The ground truth which is the reconstruction of the raw SIM images using a conventional (discussed in [28] and references therein) method is also cropped to $256 \times 256 \times z$. The values of z are limited to 1, 3, and 5 due to the memory constraints of our computing machine. Our project aims to evaluate the effect of training on three-dimensional volumes.

3.1.2 Normalization

Normalization is performed to ensure that the input data used for training a deep learning model has consistent scales and distributions across different features or variables. By normalizing the data, we bring it into a range where the values are more comparable and easier for the model to process. Normalization helps in improving convergence, avoiding vanishing or exploding gradients, and achieving better generalization [37]. Normalizing data in deep learning is crucial for several reasons [61] including:

- 1. Gradient-based Optimization: Normalization helps in preventing gradient explosions or vanishing gradients during the training process, making it easier for the model to optimize and converge.
- 2. Balanced Learning: Normalization ensures that all features contribute more evenly to the learning process, preventing dominance by features with larger scales.
- 3. Regularization: Normalization acts as a form of regularization by imposing constraints on the model's weights, promoting better generalization, and reducing overfitting.
- Faster Convergence: Normalized data can lead to faster convergence during training, reducing the number of iterations required for the model to reach optimal performance.
- 5. Increased Model Stability: Normalization reduces the sensitivity of the model to the scale of input features, making it more stable and robust to variations in the input data .

Min-Max Norm

Min-Max normalization scales the data to a fixed range, typically between 0 and 1. It linearly transforms each feature value based on the minimum and maximum values observed in the data. The formula for Min-Max normalization is:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}$$
(3.2)

Here, X is the original feature value, X_{min} is the minimum value in the dataset, and X_{max} is the maximum value in the dataset.

Percentile Norm

Percentile normalization is a data normalization technique that aims to transform the values of a dataset to a common scale based on their relative position within the data distribution. It ensures that each data point is mapped to a percentile value, representing its position within the distribution. This normalization method is particularly useful when dealing with skewed or heavy-tailed datasets where extreme values can heavily impact the analysis.

The percentile normalization process involves the following steps: 1. Sorting: First, the dataset is sorted in ascending order. 2. Assigning Percentiles: Each data point is then assigned a percentile value based on its rank within the sorted dataset. The percentile value represents the percentage of data points that are equal to or below a particular data point. 3. Rescaling: Finally, the dataset is rescaled using the calculated percentiles. The original values are replaced with their corresponding percentile values. This mapping ensures that the transformed values represent the position of each data point within the data distribution.

Percentile normalization allows for better comparison and analysis of data across different scales or when dealing with outliers. By mapping the data to percentiles, it mitigates the impact of extreme values and ensures that data points are represented in a relative manner.

It is worth noting that percentile normalization does not guarantee that the resulting distribution will have a specific statistical property (such as zero mean and unit variance). Instead, it focuses on preserving the relative positions of data points within the distribution.

In this project, all images are normalized using percentile norm, then passed to a minmax norm to ensure the pixel values range between 0 and 1 to provide the full accepted range of image processing in the Python programming language.

3.2 Datasets

Selected samples from the FairSIM dataset include LSEC Actin data with a depth of 7 and U2OS Actin data with a depth of 53 for training, and one U2OS Tubulin image



Figure 25. 3D data sample, [1].

with 8 z-planes as the test data shown in Table 1. This set of samples has been chosen for training and testing since there has been a limit in the number of available samples from a SIM system of the wavelength of 525nm. It was important to keep U2OS samples in both training and validating to make a better generalization. However, as there is no Actin sample in the validation, that might result in an inappropriate training result. It is better to look at the images altogether and then decide to choose the validation and training sets.

Table 1. Training, Validation, Testing, split of FairSIM dataset including raw 3D-SIM images and the corresponding reconstructed images [1].

	Original	2d(z = 1)	z = 3	z = 5
Training	2	960	336	208
Validation & Testing	1	126	48	32
Total	3	1086	384	240

Opstad, Ida S., et al. [2] has recently published another dataset. The article describes the methods and procedures for obtaining high-resolution 3D images of mitochondrial dynamics and lysosomal function in liver and fixed H9c2 rat cardiomyoblast cells. The images were acquired using a DeltaVision OMX V4 Blaze imaging system and reconstructed using the manufacturer-supplied softWoRx program. The 3D-SIM data are of high quality and suitable for developing SIM DL reconstruction algorithms. Considering the publication date and the variety of the dataset samples, it is a perfect candidate for our study. In this project, we are using 18 samples of fixed-cell mitochondria.

It should be understood that while the SIM images are generally of good quality, they

may still exhibit reconstruction artifacts. These artifacts can be of lower intensity compared to the actual biological features of interest and can be removed by linearly adjusting the image brightness. Negative intensity values, resulting from the reconstruction algorithm and considered non-physical, can usually be disregarded without losing the biological context [2].

Table 2. Training, Validation, Testing, split In this study, we used 18 samples of raw 3D-SIM images and the corresponding reconstructed images of mitochondrial dynamics in fixed H9c2 rat cardiomyoblast cells [2], [3].

	Original	2d(z = 1)	z = 3	z = 8
Training	13	6032	2011	754
Validation	3	1616	539	22
Testing	2	928	310	116
Total	18	8576	2860	892

Chapter 4

Methodology and Implementation

Introducing UPIGAN (Unrolled Physics Informed Conditional Generative Adversarial Network), an advancement in the field of generative models. UPIGAN represents a fusion of two powerful concepts: conditional generative adversarial networks (CGANs) and physics-informed learning. By combining the ability of CGANs to reconstruct raw images conditioned on the ground truth with the incorporation of analytical laws and constraints, UPIGAN pushes the boundaries of DL specifically designed for SIM image superresolution. In this paper, we present a comprehensive exploration of UPIGAN's architecture and training methodology.

Previously introduced physics-informed caGAN architecture [19] utilizes the advantages of an analytical model and an adversarial generative network combined with a deep residual channel attention network (RCAN) [16] and provides a powerful means specifically for microscopy image restoration. The loss function in the caGAN architecture includes the known wide-field point spread function (PSF) of the optical system to compare the simulated outputs from the restored images with the input images. The neural network is trained in various training conditions, including training with both 2D and 3D data with two different datasets introduced earlier in this chapter.

Training a deep learning model with 2D or 3D patches of a 3D dataset involves certain trade-offs that need to be considered. Here are some key trade-offs:

- Memory and Computational Requirements: Working with 3D datasets requires handling volumetric data, which significantly increases memory and computational requirements compared to 2D datasets. Processing and training models on 3D patches can be more computationally intensive and memory-demanding, potentially limiting the scale of the dataset or model complexity that can be utilized.
- 2. Data Representation: 2D patches capture spatial information from individual slices of the 3D volume, while 3D patches capture both spatial and contextual information

across neighboring slices. By using 2D patches, the model may lose some valuable contextual information present in the 3D data. On the other hand, 3D patches can retain more comprehensive spatial relationships but may increase the risk of overfitting due to larger input sizes and higher model complexity.

- 3. Training Efficiency and Convergence: Training with 2D patches typically converges faster compared to training with 3D patches due to smaller input sizes. 3D patches introduce additional challenges, such as increased sample correlation, which can lead to slower convergence and longer training times. Balancing training efficiency and convergence is crucial to achieving optimal model performance.
- 4. Generalization and Robustness: Training with 3D patches can enhance the model's ability to generalize and robustly handle unseen 3D data, as it learns from the inherent volumetric patterns and structures. However, if the available dataset is limited or lacks diversity, training a 3D model may be more susceptible to overfitting. In such cases, 2D patches might provide a more stable training process and generalization capability.
- 5. Interpretability and Visualization: 2D patches offer a simpler and more intuitive visual representation, making it easier to interpret and visualize the learned features or model outputs. Conversely, 3D patches may provide more complex and detailed representations, but visualizing and interpreting them can be challenging due to the inherent volumetric nature of the data.

Ultimately, the choice between using 2D or 3D patches for training a deep learning model on a 3D dataset depends on various factors, including available computational resources, dataset size and diversity, training efficiency requirements, desired generalization capabilities, and the specific characteristics of the problem at hand.

4.1 Unrolling

To clarify the unrolling process, unrolling the generator involves inputting raw sim images into the first layer of the generator, which has the same dimensionality as the input images. However, after unrolling, there is a dimensionality mismatch when feeding the output back into the input. To address this issue, there are several possible solutions available.

The first solution involves down-sampling the output, but this is not a correct approach because the goal of the project is to up-sample the input images. Down-sampling the upsampled images is inappropriate and defeats the purpose of the project.

The second solution is to up-sample the raw sim images first and then use the residual channel attention network (RCAN) to increase the image quality. While this solution is effective in solving the dimensionality mismatch in the z dimension, it does not address the dimensionality mismatch in the number of phases and angles, which is another issue between the input and the output.

The third and final solution involves incorporating two distinct super-resolution architectures in the generator design. In the first part of the generator, a super-resolution RCAN architecture is used to produce an output that is equivalent to the ultimate output of the system and the ground truth. This output is then fed into another unrolled network that contains physics-informed terminology. This network can use the same RCAN structure without an up-scaling layer, or any other image reconstruction network. For this project, the network used in the unrolling step is RCAN with a scaling factor of 1. By incorporating two distinct super-resolution architectures, the dimensionality mismatch can be effectively addressed, leading to high-quality image reconstruction.

In the first layer of unrolling the generator, the input has the same size as the raw sim images. After unrolling, we need to feed the output of the network to its input, which may result in a dimensionality mismatch. Therefore, we need to develop a robust mechanism to handle the dimensionality changes during the unrolling process effectively. The proposed unrolled Physics Guided Channel Attention Network (UPICAN) is illustrated in the block diagram 26. The generator architecture is depicted on the right side of the diagram and is designed to process raw simulated images. The generator acts as a regular super-resolution RCAN architecture, which takes both the input and the ground truth and minimizes the loss function in a DNN training iterative process to enhance the visual quality of the output, making it look like the ground truth with double the resolution of the raw input. The discriminator is responsible for distinguishing between real and generated images. It trains the generator by providing feedback on how well the generator performs. A convolutional neural network classifier has been used for the discriminator. The GAN is trained by alternating between training the generator and discriminator. This involves feeding low-resolution images into the generator, up-sampling them, and passing them through the discriminator to see how well it can be tricked by the generator to confuse the high-resolution generated images and the ground truth images.

Algorithm 1 Pseudocode of the physics-informed unrolling process. **Require:** $n \ge 0$ ▷ number of unrolling iterations **Require:** $\gamma_0 \ge 0$ 1: $N \leftarrow 1$ 2: $SR \leftarrow RCAN_{2x}(d(x, y, z))$ 3: $OTF_{WF}(u, v, w) \leftarrow FT[PSF_{WF}^{T}(x, y, z)]$ ▷ FT is the Fourier Transform 4: $\hat{o}_0(x, y, z) \leftarrow SR(x, y, z)$ 5: while $N \neq n$ do $\hat{O}_N(u, v, w) \leftarrow \frac{FT[SR \otimes_3 PSF_{WF}^T(x, y, z) + SR + RCAN_{1x}(\hat{o}_{N-1}(x, y, z))]}{\gamma_{N-1} |OTF_{WF}(u, v, w)|^2 + 1}$ 6: $\hat{o}_N(x, y, z) \leftarrow IFT[\hat{O}_N(u, v, w)]$ ▷ IFT is the Inverse Fourier Transform 7: $\gamma_N \leftarrow \gamma_{N-1}/2$ 8: $N \leftarrow N - 1$ 9:

4.2 Loss Function

4.2.1 Adversarial Loss Function

Adversarial loss is a popular technique used in deep learning for generative models. It involves training two models simultaneously: a generator model that produces synthetic samples and a discriminator model that distinguishes between the synthetic and real samples. The generator model is trained to produce samples that fool the discriminator, while the discriminator model is trained to accurately distinguish between the two types of samples [37]. The adversarial loss is defined as the objective function used to train the GAN, which involves minimizing the cross-entropy loss between the discriminator's predictions and the true labels of the samples. The generator is trained to maximize this loss, while the discriminator is trained to minimize it. This is called the mini-max loss. The min-max loss, also known as the adversarial loss, can be represented as:

$$L_{min-max}(G,D) = E_o[log(D(o))] + E_d[log(1 - D(G(d)))]$$
(4.1)

Here, G and D represent generator and discriminator network respectively, d denotes the raw input, and o represents the ground truth. The loss consists of two terms: the expectation over the real data samples and the expectation over the generated samples. The first term aims to maximize the probability of the discriminator correctly classifying the real data samples (o) given the corresponding raw input (d). The generator tries to minimize this term to improve the realism of its generated samples. The second term aims to maximize the probability of the discriminator correctly classifying the generated samples (G(d)) as fake, given the raw input (d). The discriminator aims to minimize this term by correctly distinguishing the generated samples from real ones. The min-max loss formulation in GANs represents the adversarial nature of the training process, where the generator and discriminator networks play a game against each other to improve their respective performances.

4.2.2 Generator Loss

As follows, we are introducing one loss function for the generator design and another for the discriminator to reach the goal of the mini-max game. The loss function of the generator includes two distinct terms as described in Eqn. 4.2

$$L_G(\hat{o}, o, d) = L_{SR}(o, \hat{o}) + \alpha L_{adv}(D(\hat{o}), 1)$$
(4.2)

Here G and D stand for generator and discriminator respectively, o is the ground truth object image and \hat{o} is the reconstructed image, and d represents the raw SIM images. α is the selected weight for the mini-max game, ranging between 0 and 1 inclusive. In a word, the generator loss considers the effect of loss from the super-resolution architecture itself, besides a weighted adversarial loss from the generator's perspective. The superresolution loss in this study is selected as the super-resolution used in the original RCAN architecture [16], which is the Mean Absolute Error defined as $L_{SR}(o, \hat{o}) = ||o - \hat{o}||$. The adversarial loss used in the generator (L_{adv}) is defined as follows (Eqn. 4.3):

$$L_{adv}(o, \hat{o}) = BinaryCrossEntropy(o, \hat{o})$$

= -(olog(\hlocolog) + (1 - o)log(1 - \hlocolog)) (4.3)

4.2.3 Discriminator Loss

The loss function used for the discriminator is also a Binary CrossEntropy loss (Eqn. 4.3).

$$L_D(o, \hat{o}) =$$

$$BinaryCrossEntropy(1, D(o))$$

$$+BinaryCrossEntropy(0, D(\hat{o}))$$

(4.4)

4.3 UPIGAN Design

The physics-informed caGAN architecture utilizes the advantages of an analytical model and an adversarial generative network combined with a deep residual channel attention network (RCAN) [16] and provides a powerful means specifically for microscopy image restoration [19]. To do so, the residual channel attention generative adversarial network is adopted and trained under various conditions with multiple datasets obtained from [2]. The loss function in the caGAN architecture includes the known WF point spread function (PSF_{WF}) of the optical system to compare the simulated outputs from the restored images with the input images. The results will show us in the following chapter how deep learning modeling can combine with or replace classical model-based methods. The unrolling step in this project includes exploiting the benefits of the Unrolled Physics Informed Channel Attention Network (UPICAN) by unrolling it to incorporate knowledge about the Point Spread Function (PSF) of the WF Microscope. We intend to improve the adversarial loss function by incorporating the mini-max loss to challenge the generator further to perform better against the discriminator. As depicted in Fig. 26. one discriminator compares all the discrimination from the ground truth with an array of ones. The other discriminator is responsible for the faked generated results compared to zeros.

The role of the first RCAN is to generate the super-resolved image by twice the resolution of the input raw SIM. The condition checks if the number of unrolled iterations has been completed and outputs the final reconstruction from the super-resolution section if so. Otherwise, it keeps unrolling through the RCAN architecture with a resolution mapping of one-to-one until the completion of unrolling iterations. The physics-informed section in this block is shown in the pseudocode presented in Alg. 1 mimics the behavior of a simplified Wiener Filter. The Wiener Filter is a deconvolution method used to recover a signal convolved with a known system response. It is based on the assumption that the signal and the noise in the convolved signal are both random processes statistically independent of each other [62].



Figure 26. GAN Block Diagram. Respectively, (a) and (b) represent the discriminator and the generator design, respectively. Each block takes the required inputs and passes them to the discriminator/generator architecture. In this architecture, input *d* includes 15 raw SIM images from a single focal place of the object, *o* (corresponding to 3 angles and 5 phases of the SI pattern, Fig. 1). *o* represents the ground truth, and *h* is the physicsinformed term (PSF). The generator architecture has a super-resolution network implemented in it. After completion of training, \hat{o}_n provides the final reconstructed image [23].



Figure 27. UPIGAN Generator Block Diagram. This block magnifies the SR block from Fig. 26. kx for k = 1, 2 is the super-resolution magnification in the RCAN block. At first, the raw input's resolution is doubled using RCAN architecture. In the unrolling step, the resolution in the RCAN block stays the same, as we have already achieved double the resolution; but it only enhances the quality of the image.

Chapter 5

Experimental Results

In this chapter, we present the ablation study, our results, and the performance of the deep learning model on our image data.

5.1 Ablation Study

An ablation study is a systematic analysis method that involves selectively removing or disabling specific components, modules, or factors in a system or model to evaluate their individual contributions and assess their impact on overall learning performance. It helps in understanding the importance and relevance of each component, aiding in model optimization, and providing insights into system behavior [63]. The primary purpose of conducting an ablation study is to analyze and evaluate the impact of individual components or factors on the overall performance of the UPIGAN learning process. It helps in identifying the most critical components, optimizing the system, and gaining insights into the functioning and behavior of the system [64]. A typical ablation study involves the following steps: 1) Identifying the components or factors of interest, 2) Systematically removing or disabling these components, 3) Evaluating the performance of the modified system, and 4) Comparing the results with the baseline system to determine the contribution and impact of each component [65]. Ablation studies are widely used in various domains, including deep learning, machine learning, and computer vision, to understand the behavior and significance of different components in complex systems.

Training with each dataset requires a separate ablation study. In this thesis project, the DNN architecture includes various sections that require tuning of separate hyper-parameters. There is an RCAN super-resolution algorithm whose parameters need to be tuned. These parameters include batch size, number of residual channel attention groups, number of residual channel attention blocks (RCABs), and number of channels in the channel attention attention architecture. One of the limitations of the current study is the random access memory available. A light enough architecture is needed to provide enough space for the 3D

unrolling of a DNN. At the same time, the performance of the system should not drop significantly by making the network lighter. As a result, we start the training process by choosing the parameters used in the original architecture of RCAN [16]. Then replace them with smaller values and compare the effect of this change on the performance of the super-resolution.

The ablation study on the FairSIM dataset is explained in detail. As noted in Table 3. not even a drop in the number of residual groups and the number of RCABs has a negative impact on the training, but also improves the performance in most cases. It can be due to the complexity of the system that might lead to overfitting. Decreasing the batch size can have a slightly higher error or lower PSNR, however, it still increases the SSIM value. Therefore, it is guaranteed that we can minimize the said three parameters. On the other hand, using fewer channels in the architecture drops the performance significantly. As a result, the number of channels will remain unchanged. The table is followed by the performance metrics along the training in Fig. 28. The scenario numbers in the table are color-coded in the figure.

Table 3. RCAN ablation study using FairSIM dataset. the first column is the scenario number for each experiment. The following four columns correspond to the hyper-parameters adjusted in this project. The # sign corresponds to "the number of". The last four columns show the performance metrics as the result of training the network using the parameters in the first four columns. The blue values represent the performance of the parameters that are finalized for completing the training. "tr" and "Chs" also stand for the "training" and "Channels" respectfully.

	Batch Size	# Groups	# RCABS	# Chs	tr MAE	MSE	PSNR	SSIM
1		3	5	16	0.00490	0.01765	18.33	0.3181
2	8	5	5	64	0.00467	0.00722	22.55	0.4833
3	0			16	0.00495	0.02536	16.75	0.3047
4		2	3	64	0.00471	0.00752	22.13	0.4980
5	2	-		04	0.00479	0.00872	21.54	0.5308

After finalizing the parameters for the RCAN, the network will be promoted to a conditional GAN architecture by adjusting the number of iterations for the generator (RCAN)



Figure 28. RCAN FairSIM ablation Study metrics performance presented including a) the MAE from the RCAN loss function during the training, b-d) MSE, PSNR, and SSIM on the validation dataset along the time. The horizontal axis in each figure corresponds to the number of epochs ranging from 0 to 800. The vertical axis shows the corresponding metric value.

here while the discriminator is trained once. Finding the most effective number of iterations for the generator, we start adjusting the α value responsible for the GAN loss. Table 4. and Fig. 29. represent the background study to select number of iterations for the generator and α . Adjusting the main parameters for the RCAN used as a Generator in a GAN, leaves us with finding the best hyper-parameter to unroll the network. To find the optimum γ value Table 5.

	# Gen	# Disc		Con MAE	MSE	DCND	SCIM	
	Iteration	Iteration	α	Gen MAE	NISE	FSINK	391101	
1	1			0.05644	0.00655	23.20	0.4510	
2	3		0.1	0.05303	0.00552	23.84	0.4692	
3	5		0.1	0.1	0.05639	0.00578	24.25	0.4720
4		1		0.04894	0.00408	26.27	0.4629	
5	10		0	0.05033	0.00389	26.06	0.4451	
6	10		0.5	0.05024	0.00411	26.00	0.4417	
7			0.25	0.04862	0.00387	26.27	0.4495	

Table 4. GAN ablation study using FairSIM dataset. "Gen" stands for "Generator"



Figure 29. GAN FairSIM ablation Study metrics performance.



Figure 30. γ FairSIM ablation Study metrics performance presented.

	There et a later strang a sing i another dataset								
γ	RCAN Training MAE	MSE	PSNR	SSIM					
0	0.05916	0.01269	23.82	0.5540					
0.05	0.05462	0.00576	24.29	0.6362					
0.1	0.05477	0.00750	22.70	0.6033					
0.3	0.05524	0.00639	23.36	0.6516					
0.5	0.05600	0.00785	22.82	0.6233					
0.8	0.05543	0.00534	24.60	0.5845					
1	0.05380	0.00503	24.83	0.6742					
10	0.05292	0.00597	23.75	0.6305					

Table 5. γ ablation study using FairSIM dataset.

The major part of the current project ablation study is the presence of generative adversarial networks. As a part of this ablation study on the Fixedcell dataset, unrolled super resolution (USR) has been introduced and submitted for publication [24].

5.1.1 USR

The proposed USR is illustrated in the block diagram in Fig. 31. The generator architecture is depicted on the left side of the diagram, and it is designed to process raw simulated images. The generator acts as a regular super-resolution RCAN architecture, which takes in both the input and the ground truth and minimizes the loss function in a DNN training iterative process to enhance the visual quality of the output, making it look like the ground truth with double the resolution of the raw input.



Figure 31. Block Diagram of the proposed unrolled Super-Resolution (USR) architecture. The math is shown on the right side where \hat{y}_k is the super-resolved output in k^{th} unrolling iteration, and x is the raw SIM input image [24].

5.2 Results

5.2.1 FairSIM Training

The training process for the proposed method was carried out using the finalized parameters as outlined in Table 6. These parameters were carefully selected based on prior experimentation and expert knowledge to ensure optimal performance. The convergence and effectiveness of the training procedure were evaluated through quantitative validation metrics, depicted in Fig. 32. as plotted values over the validation iterations. Additionally, Table 7. provides a comprehensive overview of the metric values obtained for the final iteration of the validation process.

To further evaluate the performance of the proposed method, qualitative validation samples are presented in Fig. 33. showcasing the reconstructed images. And in 3D training, Fig. 34. and Fig. 35. offer representative examples of the obtained results.

These visualizations and quantitative metrics serve as indicators of the effectiveness and reliability of the developed approach. They provide insights into the capability of the 3D training of FairSIM data to enhance image resolution and accurately capture the underlying details of the samples. The results demonstrate the potential of the proposed method in advancing the field of 3D imaging and its application in various domains such as medical imaging and microscopy.

Table 6. The fixed parameters used for training the FairSIM dataset. The character "#" stands for "the number of" and "lr" is the abbreviation for "learning rate".

# gen	batch	gen	Disc	lr decay	# groups	# PCABs	0	~
iteration	size	start lr	start lr	rate	# groups	# ICADS	α	Ŷ
10	2	10e-4	10e-6	0.5	2	3	0.25	1

Table 7. Metric performance for the architecture of RCAN super-resolution with or without GAN, with or without physics-informed unrolling, and with both GAN and physicsinformed unrolling (UPIGAN) on the FairSIM dataset.

Network	MSE	PSNR	SSIM
RCAN	0.008212	21.64	0.5321
CAGAN	0.003590	26.72	0.7091
USR	0.005056	24.81	0.6819
UPIGAN	0.003584	26.83	0.7043



(b) PSNR



(c) SSIM



Figure 32. FairSIM quantitative results comparison on various architectures.



Figure 33. 2D FairSIM qualitative results of 3 different samples. Each row corresponds to one sample.

5.2.2 FixedCell Training

The hyper-parameters for the Fixedcell dataset were determined using a similar approach. The training process was carried out with the specific parameters listed in Table 8. To evaluate the performance of the trained model under these conditions, the results are presented both qualitatively and quantitatively.

The qualitative assessment of the trained model's performance is demonstrated through visual comparisons of the validation result images. Figures 37. 38. 39. and 40. showcase the visual comparisons, allowing a detailed analysis of the reconstructed images.

In addition to the visual assessment, quantitative measures were employed to further evaluate the model's performance. The MSE, PSNR, and SSIM metrics were utilized to provide objective insights into the quality of the reconstructed images. The results of these metrics on the validation samples are presented in Figure 36. and Table 9.

The combination of qualitative visual comparisons and quantitative metrics provides a comprehensive evaluation of the trained model's performance on the Fixedcell dataset.



Figure 34. FairSIM qualitative results trained with 3d samples cropped to a 3D volume with z = 3. The first row represents the first slice of the volume in the z-axis. followed by z = 2 and in the last row z = 3.



Figure 35. Another FairSIM qualitative results trained with 3d samples cropped to a 3D volume with z = 3.

These results serve to illustrate the effectiveness and accuracy of the developed approach, providing valuable insights for further analysis and discussion in the subsequent sections.

# gen iteration	batch size	gen start lr	Disc start lr	lr decay rate	# groups	# RCABs	α	γ
3	2	10e-4	10e-6	0.5	2	3	0.1	0.5

Table 8. The parameters used for training the Fixedcell dataset.

Table 9. Metric performance for the architecture of RCAN super-resolution with or without GAN, with or without physics-informed unrolling, and with both GAN and physicsinformed unrolling (UPIGAN) on the Fixedcell dataset.

Network	MSE	PSNR	SSIM
RCAN	0.011340	20.3	0.3379
CAGAN	0.009153	21.77	0.4971
USR	0.010210	20.97	0.3893
UPIGAN	0.009163	21.92	0.5809


(b) PSNR







Figure 36. Fixedcell quantitative results comparison on various architectures [24], [23].



Figure 37. 2D Fixedcell qualitative results of 2 different samples. Each row corresponds to one sample.



Figure 38. Fixedcell qualitative results trained with 3d samples cropped to a 3D volume with z = 3.



Figure 39. 3D FixedCell visual results with z = 3.



Figure 40. Another Fixedcell qualitative results trained with 3d samples cropped to a 3D volume with z = 5.

Chapter 6

Discussion and Conclusion

6.1 Discussion

In the discussion section, we delve deeper into the obtained results and their implications. Our findings reveal several important aspects regarding the impact of unrolling the RCAN (Residual Channel Attention Network) and incorporating model-based widefield (WF) point spread function (PSF) information in the UPIGAN (Unrolled Physics-Informed Channel Attention Generative Adversarial Network) framework.

The visualizations of the results aim to demonstrate the efficacy of the proposed method in capturing important details and improving the resolution and quality of the reconstructed images. By visually comparing the reconstructed images against ground truth, it is possible to observe the enhancements achieved in terms of sharpness, texture, and overall image fidelity. The qualitative samples presented showcase the successful reconstruction of challenging image features, such as fine structures, edges, and intricate patterns. Moreover, the presented images demonstrate the ability of the proposed method to effectively address common issues encountered in 3D-SIM reconstruction, including noise reduction, artifact suppression, and contrast enhancement. These visual and quantitative validation results provide strong evidence of the effectiveness and robustness of the proposed method. The achieved improvements in image quality and resolution highlight the potential of the developed approach in advancing 3D-SIM image reconstruction and its applicability in various microscopy and medical imaging applications.

Firstly, the results clearly demonstrate that unrolling the RCAN network alone has a positive impact on the overall performance of the image reconstruction task. The unrolling process takes advantage of physics-informed terminology, which aids in capturing the underlying physical characteristics and improves the fidelity of the reconstructed images. This highlights the effectiveness of integrating physics knowledge into the deep learning framework.

Moreover, the incorporation of model-based WF PSF information into the UPIGAN framework further enhances the performance, contributing to even higher-quality reconstructions. By leveraging the knowledge of the imaging system's point spread function, UPIGAN can effectively compensate for low-light conditions, electrooptical noise, and other undesirable aberrations that are inherent in the acquired data. This signifies the importance of utilizing prior information about the imaging system in solving inverse imaging problems.

Additionally, the results shed light on the effectiveness of converting the RCAN network to a CGAN (Conditional Generative Adversarial Network) architecture. The CGAN approach allows for conditional image generation, where the generator learns to produce higher-resolution outputs conditioned on the low-resolution inputs. This conversion demonstrates promising improvements in performance compared to the RCAN model alone, emphasizing the potential of generative adversarial networks in the context of super-resolution imaging.

However, it is noteworthy to mention that while the transition from RCAN to UPI-CAN or from RCAN to CAGAN leads to significant performance enhancements, the same level of improvement is not as prominent when comparing RCAN to UPIGAN in terms of quantitative results. The quantitative metrics might not reflect the full extent of the performance improvement achieved by UPIGAN. On the other hand, the qualitative results visually exhibit the notable enhancements achieved by UPIGAN, showcasing the efficacy of the physics-guided unrolling process and the integration of the CGAN architecture.

In addition to the aforementioned findings, it is important to address an aspect that requires caution when interpreting the results: the presence or absence of fine details in the reconstructed images. Although the images are converted into numerical matrices, it is observed that certain fine details exist within the reconstructed images. These details possess non-zero values in comparison to the black (zero) background. However, their values are incredibly small, to the extent that they are beyond the visual perception capabilities of the human eye.

This phenomenon highlights the need for a careful interpretation of the reconstructed images. While the quantitative evaluation metrics and visual comparisons demonstrate improvements in image resolution, it is crucial to acknowledge that these subtle details might not be visually discernible due to their minute magnitudes. The limitations of human perception, particularly when it comes to distinguishing fine details, should be taken into consideration while assessing the quality and completeness of the reconstructed images.

Moreover, it is worth exploring future research avenues to investigate techniques that can potentially enhance the visibility of these subtle details. This could involve employing advanced visualization methods or incorporating human perceptual models to better represent and display the reconstructed images, allowing researchers and medical professionals to discern even the finest structural information.

It is important to maintain a cautious approach when evaluating the reconstructed images, recognizing that the absence of visually apparent fine details does not necessarily indicate a failure in the reconstruction process. The primary focus should remain on the overall improvement in resolution and the successful restoration of essential image features rather than solely relying on the visual detection of every minuscule detail.

In summary, our discussion highlights the positive impact of unrolling the RCAN network, incorporating model-based WF PSF information, and converting the RCAN to a CGAN in the UPIGAN framework. While quantitative metrics might not fully capture the performance improvement achieved by UPIGAN, the qualitative results substantiate its efficacy in enhancing image reconstruction quality. These findings underscore the potential of combining physics knowledge with deep learning approaches, offering a promising avenue for more efficient and accurate 3D-SIM image reconstructions.

The comparison between 2D and 3D reconstruction approaches unveils distinct benefits and considerations for each methodology. However, the results obtained from our study do not indicate a significant difference between the two. In the context of our research, two alternative approaches were explored.

The first approach involved training a 2D network capable of handling each z-slice of a 3D test volume independently. This approach allowed us to apply super-resolution techniques to individual slices of the volume. The advantage of this method lies in its simplicity and feasibility, as it eliminates the need to handle the full 3D volume during training and testing. By training the network on 2D slices, we aimed to assess the effectiveness of a simplified approach for achieving enhanced resolution in 3D-SIM reconstructions.

The second approach entailed training the network using complete 3D volumes and subsequently feeding the entire test volume into the network for reconstruction. In this case, the network had the capability to directly process 3D data, encompassing the full volume. The advantage of this approach lies in its potential to capture more comprehensive spatial dependencies within the 3D volume. It was anticipated that incorporating the 3D point spread function (PSF) in the unrolling process would lead to superior reconstruction results. However, as mentioned in the introduction chapter, training with 3D data introduces its own set of challenges, such as increased complexity and resource requirements.

Interestingly, our findings did not reveal a significant discrepancy in terms of reconstruction quality between the 2D and 3D approaches. The results suggest that, for the specific dataset and experimental conditions considered in this study, the benefits gained from 3D training did not substantially surpass those achieved with 2D training. This observation implies that while incorporating the full 3D volume during training has its theoretical advantages, the practical gains in resolution enhancement may not be substantial enough to justify the additional complexities and resource demands associated with 3D training.

It is important to note that the potential for achieving significantly better resolution from 3D data could exist, particularly for larger volumes as proved by conventional methods such as in [28]. However, such gains might require a substantial investment of time, computational resources, and expertise to overcome the inherent challenges associated with 3D training. The decision to pursue 2D or 3D training should be based on a careful evaluation of the specific requirements and constraints of the imaging application at hand.

Further investigations and refinements are necessary to explore the full potential of 3D training and its implications in 3D-SIM image reconstruction. The complexities and tradeoffs associated with 2D and 3D training should be carefully considered, taking into account the specific goals and limitations of each application scenario.

6.2 Conclusion

In conclusion, this research project focused on addressing the challenges faced in reconstructing 3D-SIM images through the development of an unrolled physics-informed generative adversarial network (UPIGAN). By leveraging the advantages of physics knowledge in both the unrolling step and the loss function, and incorporating a Residual Channel Attention super-resolution deep neural network (DNN) in the generator architecture, the goal was to achieve visually and computationally enhanced resolution.

Traditional approaches to inverse imaging problems in 3D-SIM often fall short due to model mismatches and computational limitations. Deep learning approaches, such as the UPIGAN proposed in this study, offer a promising alternative by learning directly from data without relying on an analytical model. The integration of physics-informed terminology and the use of a generative adversarial framework aimed to provide improved reconstruction results.

The research objectives were multi-fold. Firstly, the project aimed to convert a regular deep super-resolution algorithm into a generative adversarial super-resolution algorithm to achieve enhanced resolution. Additionally, the investigation aimed to assess the differences between training on two-dimensional images and three-dimensional volumes, determining the preferred approach for 3D imaging. Furthermore, the project sought to improve training efficiency by enhancing the physics-guided aspect of the algorithm.

Through the implementation and evaluation of the unrolled physics-informed channel attention GAN network, utilizing data samples from a 3D-SIM system, the research aimed to contribute to the advancement of image reconstruction techniques in microscopy imaging applications. The combination of deep learning, physics-informed design, and generative adversarial networks holds the potential to overcome limitations encountered by traditional methods, leading to more accurate and efficient image reconstruction in the context of 3D-SIM.

Overall, this thesis project explored novel avenues for improving the reconstruction of 3D-SIM images, leveraging the power of deep learning and physics-informed approaches. The outcomes of this research have the potential to enhance image quality and computational efficiency in the field of medical and microscopy imaging, contributing to advancements in various scientific and diagnostic applications.

This project serves as the foundation for further advancements and there are several avenues for future work that can be explored. The following are some potential suggestions for expanding upon this project:

• To address the issue of inadequate visibility of fine details in the reconstructed image, there are a couple of suggested approaches. Firstly, one option is to employ a more advanced normalization algorithm to enhance the presentation of the results. By utilizing a superior normalization technique, the reconstructed image can be visually optimized, allowing for better perception and analysis of fine details.

Alternatively, instead of solely focusing on reconstructing images for direct human examination, an alternative strategy involves storing the reconstructed images and employing them in other applications. For instance, integrating the super-resolution approach with an AI algorithm designed for diagnosis can yield promising results. Although the fine details might not be discernible to the human eye, they could still be perceivable to the AI algorithm. By incorporating the reconstructed images into an AI-based diagnostic system, it becomes possible to leverage the AI's enhanced perception to detect specific features of interest that may be crucial for disease diagnosis or other analysis purposes. This synergistic approach allows for a comprehensive utilization of the reconstructed images, leveraging both human and artificial intelligence capabilities. While the human eye may not capture fine details, the AI algorithm can make up for this limitation and identify significant features that might otherwise be overlooked. By combining the strengths of super-resolution techniques, image reconstruction, and AI algorithms, a more comprehensive and accurate analysis can be achieved, enhancing the potential for improved disease diagnosis or other applications that rely on image-based information.

- Application of Transfer Learning: Consider leveraging the power of transfer learning by applying pre-trained networks to this problem domain. Utilizing a well-trained network as a starting point can accelerate the learning process and potentially enhance the performance of the model.
- Integration of Stronger Physics-Informed Techniques: Explore the incorporation of more robust physics-informed techniques in the unrolling process. For instance, investigate the utilization of techniques proposed in [66] to further enhance the physicsaware aspects of the model. This could potentially improve the accuracy and reliability of the predictions.
- Designing Specific Performance Metrics for 3D Comparison: Focus on developing performance metrics specifically tailored for 3D comparison tasks. Rather than averaging the metric for all z slices, the aim is to define appropriate evaluation measures that enable control over each individual z slice independently. This would provide more fine-grained insights and control over the analysis of the results.
- Data Cleanup Prior to Training: Prioritize data cleaning procedures to address dysfunctional samples. It is essential to preprocess the data by identifying and removing any problematic or anomalous samples that could negatively impact the training pro-

cess or lead to misleading results. This ensures the model's training is based on reliable and accurate data.

- Incorporation of Unrolled GAN as a Separate Unrolling Step: Explore the addition of an unrolled GAN, as described in the background chapter [59], as a distinct unrolling step within the framework. By introducing this technique, the overall performance and effectiveness of the GAN itself can be further improved, potentially leading to more realistic and higher-quality output generation.
- Incorporate Real-Time Data Augmentation Other than cropping: Explore the use of data augmentation techniques to increase the diversity and quantity of the training data. Techniques such as rotation, scaling, flipping, and adding noise can help improve the model's generalization and robustness to variations in input data.
- Investigate Model Architecture Variations: Experiment with different architectural variations of the network, such as exploring deeper or wider models, introducing skip connections, or incorporating residual blocks. These variations may enhance the model's capacity to capture complex patterns and improve its overall performance.
- Scale up to Large Datasets: Evaluate the scalability of your model to larger datasets. Assess how well it performs when trained on a substantial amount of data, which can involve handling data storage and processing challenges, as well as optimizing the training process for efficiency.

By delving into these future work suggestions, this project can progress towards addressing more complex challenges, refining the model's performance, and expanding its applicability in various domains.

Appendix A

Original caGAN WF Loss

Previously, it was mentioned that the WF loss introduced by Cagan et al. has not been utilized in this project due to certain technical reasons. Now, we will delve into a detailed explanation of these technical reasons. We will outline how the function was defined in Python, its purpose, and the specific issue that arises.

def create_psf_loss(psf): def loss_wf(y_true, y_pred): # Wide field loss x_wf = K.conv3d(y_pred, psf, padding='same') x_wf = K.pool3d(x_wf, pool_size=(2, 2, 1), strides=(2, 2, 1), x_min = K.min(x_wf) x_wf = (x_wf - x_min) / (K.max(x_wf) - x_min) wf_loss = K.mean(K.square(y_true - x_wf))

return wf_loss

return loss_wf

The given function, create_psf_loss, is a higher-order function that returns a custom loss function for use in Keras models. It takes a parameter, PSF, which represents the point spread function. Inside the create_psf_loss function, there is another function defined called loss_wf. This inner function serves as the custom loss function for the wide-field loss calculation. It takes two arguments, y_true and y_pred, which are the ground truth and predicted values, respectively. The loss_wf function performs the following steps:

- Convolution: It applies a 3D convolution operation on y_pred with the given psf as the kernel. This simulates the effect of the point spread function on the predicted output to get the estimated raw wf input.
- 2. Pooling: The resulting convolved output, x_wf, is then down-sampled using 3D pool-

ing with a pool size of (2, 2, 1), strides of (2, 2, 1), and an average pooling mode. This down-sampling backward the super-resolution.

- Normalization: Next, the min-max normalization is applied to ensure that x_wf is within [0, 1].
- 4. Loss Calculation: Finally, the wide field loss is calculated by the mean squared difference between y_true and the normalized x_wf.

The loss_wf function returns the wide field loss, wf_loss. When calling create_psf_loss(psf), it returns the loss_wf function, which can then be used as a custom loss function in a Keras model. The returned loss function incorporates the provided PSF and calculates the wide field loss based on the ground truth and predicted values during model training.

Based on the scientific explanation, to get the correct results, the estimated wide-field input (calculated by convolving the predicted reconstruction to the PSF) should be then compared with the actual wf raw input. However, based on the explanation given above, the function in the original caGAN code compares the widefield predicted input, with the ground truth, rather than the widefield raw input as stated in the paper. In another word, the network tends to make the raw input look like the ground truth which is definitely not what the purpose of this study is. This deviation from the original approach is a technical mistake that, while it may have a positive impact on training, should be approached cautiously. It is crucial to avoid any scientific inaccuracies or misconceptions.

Given these considerations, the decision was made to exclude the use of this particular loss function from the project. This choice was made to ensure scientific rigor and to mitigate any potential misinterpretation or unintended consequences.

By being mindful of the technical details and potential implications, it was determined that omitting the utilization of this loss function is the most prudent course of action in order to maintain the integrity and accuracy of the project.

REFERENCES

- M. Müller, V. Mönkemöller, S. Hennig, W. Hübner, and T. Huser, "Open-source image reconstruction of super-resolution structured illumination microscopy data in imagej," *Nature Communications 2016 7:1*, vol. 7, pp. 1–6, Mar. 2016. [Online]. Available: https://www.nature.com/articles/ncomms10980
- [2] I. S. Opstad, G. Godtliebsen, F. Ströhl, T. Myrmel, B. S. Ahluwalia, K. Agarwal, and Åsa Birna Birgisdottir, "Three-dimensional structured illumination microscopy data of mitochondria and lysosomes in cardiomyoblasts under normal and galactose-adapted conditions," *Scientific Data 2022 9:1*, vol. 9, pp. 1–7, 3 2022. [Online]. Available: https://www.nature.com/articles/s41597-022-01207-7
- [3] A. Abraham, P. Qaladize, S. D. A. Solis, C. VAN, and C. Preza, "Automatic data preprocessing for training neural networks for image restoration," manuscript submitted to Quaesitum 2023.
- [4] L. Schermelleh, P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, H. Leonhardt, and J. W. Sedat, "Subdiffraction multicolor imaging of the nuclear periphery with 3d structured illumination microscopy," *Science*, vol. 320, no. 5881, pp. 1332–1336, June 2008. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1156947
- [5] E. H. Rego and L. Shao, "Practical structured illumination microscopy," *Methods in Molecular Biology*, vol. 1251, pp. 175–192, 01 2014.
- [6] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *arXiv:2104.07636*, 2021.
- [7] M. A. Boland, E. A. Cohen, S. R. Flaxman, and M. A. Neil, "Improving axial resolution in structured illumination microscopy using deep learning," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 379, June 2021. [Online]. Available: /pmc/articles/PMC8072200//pmc/articles/PMC8072200/?report= abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8072200/
- [8] Control system wikipedia. [Accessed: Nov 1, 2021]. [Online]. Available: https: //en.wikipedia.org/wiki/Control%5Fsystem
- [9] Artificial neural networks wikipedia. [Accessed: Nov 1, 2021]. [Online]. Available: https://en.wikipedia.org/wiki/Artificial%5Fneural%5Fnetwork
- [10] Convolutional neural networks wikipedia. [Accessed: Nov 1, 2021]. [Online]. Available: https://en.wikipedia.org/wiki/Convolutional%5Fneura%5Fneura%5Fne
- [11] R. V. Vijay, "Image deblurring using convolutional neural network," vol. 11, pp. 7–12, 2016.
 [Online]. Available: www.iosrjournals.org
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, May 2015. [Online]. Available: https://arxiv.org/abs/1505.04597v1

- [13] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-tofine approach in single image deblurring," Aug. 2021. [Online]. Available: https: //arxiv.org/abs/2108.05054v2
- [14] X. Mao, Y. Liu, W. Shen, Q. Li, and Y. Wang, "Deep residual fourier transformation for single image deblurring," 2021.
- [15] R. Lin, R. Zhang, C. Wang, X. Q. Yang, and H. L. Xin, "Temimagenet training library and atomsegnet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images," *Scientific Reports 2021 11:1*, vol. 11, pp. 1–15, Mar. 2021. [Online]. Available: https://www.nature.com/articles/ s41598-021-84499-w
- [16] Y. Zhang, "Image super-resolution using very deep residual channel attention networks," pp. 286–301, 2018.
- [17] "Residual neural network wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/ Residual_neural_network
- [18] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. S. Twitter, "Photo-realistic single image super-resolution using a generative adversarial network," pp. 4681–4690, 2017.
- [19] C. Qiao, X. Chen, S. Zhang, D. Li, Y. Guo, Q. Dai, and D. Li, "3d structured illumination microscopy via channel attention generative adversarial network," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 4, pp. 1–11, July 2021.
- [20] L. Jin, B. Liu, F. Zhao, S. Hahn, B. Dong, R. Song, T. C. Elston, Y. Xu, and K. M. Hahn, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," *Nature Communications 2020 11:1*, vol. 11, pp. 1–7, Apr. 2020. [Online]. Available: https://www.nature.com/articles/s41467-020-15784-x
- [21] A. Ozcan, G. Barbastathis, and G. Situ, "On the use of deep learning for computational imaging," *Optica, Vol. 6, Issue 8, pp. 921-943*, vol. 6, pp. 921–943, Aug. 2019, [Accessed: Nov 1, 2021]. [Online]. Available: https://www.osapublishing.org/optica/abstract.cfm?uri= optica-6-8-921
- [22] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning: Key approaches and design guidelines," in 2021 IEEE Data Science and Learning Workshop (DSLW). IEEE, June 2021, pp. 1–6, [Accessed: Nov 1, 2021]. [Online]. Available: https://ieeexplore.ieee.org/document/9523403/
- [23] S. P. Dajkhosh, M. Hossain, and C. Preza, "Structured illumination microscope image reconstruction using unrolled physics-informed generative adversarial network (upigan)," ser. Computational Imaging VII. SPIE, 2023. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12523/2663268/ Structured-illumination-microscope-image-reconstruction-using-unrolled-physics-informed-generative/ 10.1117/12.2663268.full
- [24] —, "Usr: Unrolled super-resolution with deep priors for structured illumination microscopy," manuscript submitted for Optica Imaging Congress 2023.

- [25] M. G. Gustafsson, L. Shao, P. M. Carlton, C. J. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophysical journal*, vol. 94, pp. 4957–4970, June 2008. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18326650/
- [26] Z. Liu, L. D. Lavis, and E. Betzig, "Imaging live-cell dynamics and structure at the single-molecule level," *Molecular Cell*, vol. 58, no. 4, pp. 644–659, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1097276515001653
- [27] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, 2000. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2818.2000.00710.x
- [28] C. T. S. Van and C. Preza, "Improved resolution in 3d structured illumination microscopy using 3d model-based restoration with positivity-constraint," *Biomed. Opt. Express*, vol. 12, no. 12, pp. 7717–7731, Dec. 2021. [Online]. Available: http: //opg.optica.org/boe/abstract.cfm?URI=boe-12-12-7717
- [29] L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," *Journal of Cell Biology*, vol. 190, no. 2, pp. 165–175, 07 2010. [Online]. Available: https://doi.org/10.1083/jcb.201002018
- [30] M. S. Avinash C. Kak, Principles of Computerized Tomographic Imaging, ch. 7. Algebraic Reconstruction Algorithms, pp. 275–296. [Online]. Available: https://epubs.siam.org/doi/abs/ 10.1137/1.9780898719277.ch7
- [31] D. Taubman, "Generalized wiener reconstruction of images from colour sensor data using a scale invariant prior," in *Proceedings 2000 International Conference on Image Processing* (*Cat. No.00CH37101*), vol. 3, 2000, pp. 801–804 vol.3.
- [32] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [33] M. G. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 13 081–13 086, 9 2005. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.0406877102
- [34] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [35] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 2790–2797.
- [36] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www. deeplearningbook.org.

- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770– 778.
- [39] H.-G. Nguyen, R. Fablet, and J.-M. Boucher, "Visual textures as realizations of multivariate log-gaussian cox processes," in *CVPR 2011*, June 2011, pp. 2945–2952.
- [40] J. Liu, X. Chen, Z. Wang, and W. Yin, "Alista: Analytic weights are as good as learned weights in lista," Sept. 2018.
- [41] Y. Li, M. Tofighi, J. Geng, V. Monga, and Y. C. Eldar, "Efficient and interpretable deep blind image deblurring via algorithm unrolling," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 666–681, Jan. 2020, [Accessed: Nov 1, 2021]. [Online]. Available: http://arxiv.org/abs/1902.03493
- [42] "Idr: Image data resource." [Online]. Available: https://idr.openmicroscopy.org/search/ ?query=Imaging%20Method:wid
- [43] J.-A. Conchello and J. W. Lichtman, "Optical sectioning microscopy," 2005. [Online]. Available: http://www.nature.com/naturemethods
- [44] M. Lahrberg, M. Singh, K. Khare, and B. S. Ahluwalia, "Accurate estimation of the illumination pattern's orientation and wavelength in sinusoidal structured illumination microscopy," *Appl. Opt.*, vol. 57, no. 5, pp. 1019–1025, Feb. 2018. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-57-5-1019
- [45] C. Richardson and J. W. Lichtman, "Deep learning-based super-resolution in biological imaging," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 162–168, 2015.
- [46] C. Schmied, W. Steinchen, S. Jünger, and W. Hübner, "Simulating biological images for superresolution microscopy," *Journal of microscopy*, vol. 270, no. 2, pp. 141–154, 2018.
- [47] "System wikipedia," [Accessed: Nov 1, 2021]. [Online]. Available: https://en.wikipedia. org/wiki/System
- [48] Deep learning wikipedia. [Accessed: Nov 1, 2021]. [Online]. Available: https://en.wikipedia.org/wiki/Deep%5Flearning
- [49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [50] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: https: //proceedings.neurips.cc/paper/2016/file/0ed9422357395a0d4879191c66f4faa2-Paper.pdf
- [51] L. S. Kölln, O. Salem, J. Valli, C. G. Hansen, and G. McConnell, "Label2label: Training a neural network to selectively restore cellular structures in fluorescence microscopy," *Journal* of Cell Science, vol. 135, Feb. 2022. [Online]. Available: https://journals.biologists.com/jcs/ article/135/3/jcs258994/274322/Label2label-training-a-neural-network-to

- [52] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference* on Computer Vision (ICCV), 2017.
- [53] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in CVPR, July 2017.
- [54] M. Venkatachalam, "Attention in neural networks. some variations of attention by mahendran venkatachalam — towards data science." [Online]. Available: https: //towardsdatascience.com/attention-in-neural-networks-e66920838742
- [55] J. Chen, H. Sasaki, H. Lai, Y. Su, J. Liu, Y. Wu, A. Zhovmer, C. A. Combs, I. Rey-Suarez, H. Y. Chang, C. C. Huang, X. Li, M. Guo, S. Nizambad, A. Upadhyaya, S. J. J. Lee, L. A. Lucas, and H. Shroff, "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes," *Nature Methods 2021 18:6*, vol. 18, pp. 678–687, May 2021. [Online]. Available: https://www.nature.com/articles/s41592-021-01155-x
- [56] "A review of image super-resolution paperspace blog." [Online]. Available: https: //blog.paperspace.com/image-super-resolution/
- [57] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014. [Online]. Available: https://arxiv.org/abs/1406.2661
- [58] M. Nayak, "An introduction to conditional gans (cgans)." [Online]. Available: https://medium.datadriveninvestor.com/an-introduction-to-conditional-gans-cgans-727d1f5bb011
- [59] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks," arXiv preprint arXiv:1611.02163, 2016.
- [60] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled optimization with deep priors," 2018.
- [61] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning (ICML)*, 2015.
- [62] N. Wiener, "Extrapolation, interpolation, and smoothing of stationary time series," *John Wiley* & *Sons*, 1949.
- [63] J. Guo, T. Wang, and L. Zhang, "Deep learning model ablation: Understanding the role of architectural components," in *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, 2020.
- [64] R. S. Michalski, "Model-based machine learning," *Machine learning*, vol. 95, no. 2, pp. 161–233, 2014.
- [65] L. Zhou, Y. Liu, and G. Neubig, "Evaluating the impact of model components on neural machine translation quality," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.
- [66] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov, "Sgd-net: Efficient model-based deep learning with theoretical guarantees," *IEEE Transactions* on Computational Imaging, vol. 7, pp. 598–610, 2021. [Online]. Available: https: //doi.org/10.1109%2Ftci.2021.3085534