EXPLORING DEEP LEARNING METHODS FOR LOW NUMERICAL APERTURE TO HIGH NUMERICAL APERTURE RESOLUTION ENHANCEMENT IN CONFOCAL MICROSCOPY

Shashwat Patra

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

Recommended Citation
Patra, Shashwat, "EXPLORING DEEP LEARNING METHODS FOR LOW NUMERICAL APERTURE TO HIGH NUMERICAL APERTURE RESOLUTION ENHANCEMENT IN CONFOCAL MICROSCOPY" (2023). Electronic Theses and Dissertations. 3092.
https://digitalcommons.memphis.edu/etd/3092

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.
EXPLORING DEEP LEARNING METHODS FOR LOW NUMERICAL APERTURE TO HIGH NUMERICAL APERTURE RESOLUTION ENHANCEMENT IN CONFOCAL MICROSCOPY

By

Shashwat Ranjan Patra

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Electrical and Computer Engineering

Advisor: Dr. Ana Doblas

The University of Memphis

May 2023
Acknowledgments

I would like to thank Dr. Ana Doblas for her guidance, support, and, most of all, patience throughout my time as a Master’s student. Her insights have been invaluable to my growth as a student, researcher, and person. I would like to thank my father, my first teacher, for his love, advice, and strength. He is always the first person I go to whenever I face challenges. I would like to thank Dr. Omar Skalli and Dr. Madhusudhanan Balasubramanian for taking their time to serve on my committee.
Abstract

Confocal microscopy is a widely used tool that provides valuable morphological and functional information within cells and tissues. A major advantage of confocal microscopy is its ability to record multi-color and optically sectioned images. A major drawback to confocal microscopy is its diffraction-limited spatial resolution. Though techniques have been developed that break this limit in confocal microscopy, they require additional hardware or accurate estimates of the system’s impulse response (e.g., point spread function). Here we investigate two deep learning-based models, the cGAN and cycleGAN, trained with low-resolution (LR) and high-resolution (HR) confocal images to improve spatial resolution in confocal microscopy. Our findings conclude that the cGAN can accurately produce HR images if the training set contains images with a high signal-to-noise ratio. We have also found that the cycleGAN model has the potential to perform as the cGAN model but without the requirement of using paired inputs.
# Table of Contents

**List of Figures**................................................................................................................. v

**Abbreviations**.................................................................................................................. vii

**Chapter I: Introduction**........................................................................................................ 1

1.1: Background ....................................................................................................................... 1

1.2: Super Resolution ............................................................................................................... 2

1.3: Super Resolution Confocal Microscopy ........................................................................... 5

1.3.1: Hardware Approaches .................................................................................................. 8

1.3.2 Software ....................................................................................................................... 9

1.4: Overview .......................................................................................................................... 11

**Chapter II: Proposed Methodology**..................................................................................... 13

2.1: Confocal Scanning Microscopy ....................................................................................... 13

2.1.1: Principle ..................................................................................................................... 13

2.1.2: Simulated Confocal Image Dataset ............................................................................ 16

2.1.3 Experimental Dataset .................................................................................................. 18

2.2: Conditional Generative Adversarial Networks and CycleGAN ....................................... 19

2.2.1: cGAN Framework ...................................................................................................... 23

2.2.2: cycleGANs .................................................................................................................. 28

**Chapter III: Results and Numerical Validation**................................................................. 31

3.1 cGAN Results ................................................................................................................... 31

3.2 cycleGAN Results .......................................................................................................... 39

3.2.1 cycleGAN Simulated Dataset Results ........................................................................ 39

3.2.2 cycleGAN Experimental Dataset Results .................................................................... 42

**Chapter IV: Conclusion**...................................................................................................... 46

**References**....................................................................................................................... 48

**Appendix**.......................................................................................................................... 54
List of Figures

Figure 2.1. Diagram of a Confocal Scanning Microscope. .......................................................... 14
Figure 2.2. Simulated Airy disk PSFs of the excitation, emission, and confocal system coinciding with objective lenses of NA 0.4 and 1.4 ........................................................................ 16
Figure 2.3. Examples of images in our simulated dataset. The leftmost column contains the image before convolution and noise degradation. Second and third columns are the low- and high-resolution images, respectively. Except for the first image in each, the rows correspond to different noise degradation levels.................................................................................. 18
Figure 2.4. Example of recorded LR (before and after the preprocessing) and HR images on the experimental dataset........................................................................................................... 19
Figure 2.5. A. Illustration of a fully-connected ANN. As highlighted, the first layer is the input layer, the layers between the first and last are the hidden layers, and the last layer is the output layer. Panel B shows the individual process that each neuron in a layer executes. Nodes, \( x_1, x_2, \ldots, x_n \) refers to the data output by neurons in the previous layer and “b” is the added bias. The node sums these values and applies activation function \( f(\cdot) \). Panel C illustrates the activation functions that will be used in this document. The leftmost is the LeakyReLU function, the middle is the ReLU function, and the rightmost is the sigmoid function.................................................. 20
Figure 2.6. A) Illustration of the convolution operation used between layers. Green block is the three-dimensional (3D) input layer, the pink, blue, and yellow blocks are 3D filters. Each filter produces one 2D output which is the stacked to produce a 3D output. Panel B shows convolution with a single kernel. The kernel is slid across the input layer. An inner product between the data in the kernel and the data in the input layer it overlaps is taken. The single value is then placed on the output layer in the location corresponding to where the kernel is in the input layer........... 22
Figure 2.7. Architecture and dataflow diagram of the cGAN model used. ................................. 26
Figure 2.8. Architecture and dataflow diagram for the cycleGAN model used. ........................... 29
Figure 3.1. Performance of the cGAN model based on the convergence metrics (e.g., PSNR, SSIM and standard deviation of the background) between the generated high-resolution image and the ground truth. The blue and orange curves represent the behavior of the training and validation set, respectively. .................................................................................................................. 32
Figure 3.2. Three examples of the performance of the cGAN model from the 45dB SNR validation dataset showing the low-resolution, high-resolution ground truth, and cGAN predicted high-resolution images. The last column shows the line profiles along cellular structures ........................................................................................................................................ 33
Figure 3.3. Performance of the cGAN model at different SNR levels using the validation dataset. In the line profiles plots, the orange line corresponds to the ground-truth HR images, and the green line corresponds to the model-predicted HR images. ................................. 34
Figure 3.4. Evaluation of the resolution improvement achieved by the cGAN model using a single FOV from the 45dB SNR dataset. The Fourier transforms contain a circle which surround the area in which most of the signal’s energy is contained....................................................... 35
Figure 3.5. Line profile from the center of the Fourier transform image to the upper right corner along with its exponential fits. The red line corresponds to the line profile with noise and the blue line corresponds to the exponential fit. The blue shaded area ends at the cutoff frequency defined as the location where the curve is 1% of the max value of the line profile. ......................... 36
Figure 3.6. Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 45 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the
Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair........................37

**Figure 3.7.** Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 30 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair........................37

**Figure 3.8** Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 15 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair........................38

**Figure 3.9.** Performance of the cycleGAN model based on the convergence metrics (e.g., PSNR, SSIM and standard deviation of the background) between the generated high-resolution image and the ground truth. The blue and orange curves represent the behavior of the training and validation set, respectively. ..........................................................39

**Figure 3.10.** Three examples of the performance of the cycleGAN model from the 45dB SNR validation dataset showing the low-resolution, high-resolution ground truth, and cycleGAN predicted high-resolution images. The last column shows the line profiles along cellular structures. .............................................................................................................40

**Figure 3.11.** Low-resolution and predicted high-resolution Fourier transform image pairs from the 45 dB SNR dataset for the cycleGAN model. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the LR and predicted HR images are shown below each pair. ........................................41

**Figure 3.12.** Experimental performance of the cycleGAN model to generate high-resolution images from low-resolution confocal images. .............................................................................................................42

**Figure 3.13.** Comparison of the Fourier Transforms of the low-resolution and predicted high-resolution images to show the increase of the spatial frequency content. .................................................................44

**Figure 3.14.** Low-resolution and cycleGAN predicted high-resolution Fourier transform image pairs from the experimental dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair..............44
Abbreviations

ANN: Artificial Neural Network

AU/a.u.: Arbitrary Units

cGAN: Conditional Generative Adversarial Network

CNN: Convolutional Neural Network

CSM: Confocal Scanning Microscopy

dB: Decibels

DL: Deep Learning

FOV: Field of View

FWHM: Full Width Half Maximum

GAN: Generative Adversarial Network

HR: High Resolution

LR: Low Resolution

NA: Numerical Aperture

OPRA: Optical Photon Reassignment Microscopy

OS: Optical Sectioning

PALM: Photoactivated Localization Microscopy
PSF: Point Spread Function

PSNR: Peak Signal-to-Noise Ratio

SIM: Structured Illumination Microscopy

SMLM: Single Molecule Localization Microscopy

SNR: Signal to Noise Ratio

S-SIM: Saturated Structured Illumination Microscopy

SSIM: Structural Similarity Index Metric

STED: Stimulated Emission Depletion

STORM: Stochastic Optical Reconstruction Microscopy
Chapter I: Introduction

1.1: Background

Biological research relies heavily on cell imaging to evaluate cells' physiological status and behavior. Therefore, advances in imaging systems are essential to develop our understanding of cellular mechanisms further and apply them to new diagnostic methods and disease treatment. Traditional widefield fluorescence microscopy is one of the most basic microscopy techniques. The entire specimen of interest is exposed to the light source in widefield fluorescence microscopy, resulting in an image being viewed by the observer using eyepieces or recorded onto a camera’s sensor. The major limitations of widefield fluorescence microscopy are its limited resolution power and lack of optical sectioning capability. The spatial resolution is the minimum distance between two objects ($d$) at which they produce separate distinguishable images. In widefield fluorescence microscopy, the resolution is limited by diffraction being equal to $d = n_{\text{im}}\lambda/(2\text{NA})$ where $n_{\text{im}}$ is the refractive index of the immersion medium used for the microscope objective lens ($n_{\text{im}} = 1$ for dry microscope objective lenses), $\lambda$ is the emission wavelength, and NA is the numerical aperture of the microscope objective lens. For example, the theoretical minimum resolvable distance is 256 nm for an emission wavelength of 515 nm and an oil immersion objective lens with NA = 1.4 and $n_{\text{im}} = 1.51$. The resolution parameter is a key feature in a microscope since it informs us of the minimum resolvable detail distinguished in the recorded image. A low-resolution microscope system produces blurry images, restricting its use for quantitative analysis of biological systems. The second most critical limitation of widefield fluorescence microscopy arises from the fact that there is no axial discrimination throughout the entire specimen volume. Optical sectioning (OS) capability refers to the microscope’s ability to provide transverse slices within the sample’s volume without any out-of-focus information. Due
to the lack of OS capability, widefield fluorescence images contain a high background signal level that obscures specimen detail and dramatically reduces contrast.

Confocal microscopy, in particular, has helped biological researchers understand cell motility, behavior, and regulation in three-dimensional environments that mimic the specific organization of organs (Matsumoto, 2003). The hallmark of confocal microscopes over other imaging modalities is their capability to produce optically sectioned images thanks to the point-based illumination and detection configuration. Optically-sectioned images are two-dimensional (2D) images from transverse sections of a three-dimensional (3D) sample in which the out-of-focus information (e.g., blurred areas) has been optically removed, providing clear imaging of the in-focus sections of the sample at each axial plane. Although this is the key feature of confocal microscopy compared to conventional fluorescence microscopy, another advantage of confocal microscopes is their ability to produce multi-color imaging in which different sections of the biological specimen are stained using multiple dyes, enabling colocalization studies in biological systems. These features make confocal microscopy a widely used tool to provide valuable morphological and functional information within cells and tissues. Confocal microscopy has been used in live-cell imaging, analysis of cells and tissues, drug discovery and delivery studies, and cancer pathologies.

1.2: Super Resolution

One of the major drawbacks of confocal microscopy is its spatial resolution which is limited by diffraction. Diffraction refers to the spreading of light as it passes through an aperture in an optical imaging system which causes blurring to the resultant image, limiting the detected resolution. In 1873, the German Physicist Ernst Abbe developed his diffraction limit theory reporting that the lateral spatial resolution of an optical system depends on the light source
wavelength and the microscope objective lens by its NA and the refractive index of its immersion medium. Assuming an air-based objective lens (i.e., $n_{im} = 1$), the lateral spatial resolution is proportional to the source’s wavelength and inversely proportional to the numerical aperture of the microscopic objective lens (Silfies et al., n.d.). The lateral spatial resolution is a barrier parameter in optical microscopy. Optical super-resolution microscopic techniques refer to those methods that exceed this seemingly unbreakable limit (Schermelleh et al., 2019). The most common super-resolution techniques used are structured illumination microscopy (SIM), stochastic optical reconstruction microscopy (STORM), and stimulated emission depletion microscopy (STED).

Structured illumination microscopy (SIM), developed by Mats Gustafsson in 2000, is a super-resolution technique that improves lateral spatial resolution by illuminating the sample with a structured spatial pattern. In traditional SIM systems, the sample is illuminated by a sinusoidal pattern generated by the coherent superposition between two beams coming from the ±1 terms of a diffraction grating. The sinusoidal illumination enables the formation of Moiré fringes (the product of two multiplicatively superimposed fine patterns) onto the camera’s sensor. The high-frequency information of the sample that is not usually transferred through the imaging process of a classical widefield microscope can now be encoded with the generation of Moiré fringes. After combining at least three images with the Moiré fringes properly laterally shifted, a final super-resolved image can be reconstructed. The resolution limit in SIM depends on the lateral frequency of the sinusoidal pattern, being double when the frequency of the pattern coincides with the cutoff frequency (e.g., diffraction limit) of the microscopic imaging system (Gustafsson, 2000). This limit can be further exceeded by saturating the fluorophores thus creating a nonlinear relationship between the illumination intensity and the emission rate of the
sample. This system, also developed by Gustafsson, is called saturated structured illumination microscopy (S-SIM). The nonlinearity introduces an infinite amount of higher frequency harmonics to the Moiré fringes, theoretically allowing us to produce images with unlimited resolution. However, in practice, only a small set of these harmonics have enough energy to generate super-resolution images above the SIM limit with a high signal-to-noise ratio (SNR) (Gustafsson, 2005). SIM and S-SIM systems also provide optically-sectioned images of 3D samples and multi-color imaging by staining different sections of the sample with several staining dyes. However, the quality of the super-resolved images may be hampered by the reconstruction algorithm, which can produce unwanted artifacts (Christensen et al., 2021).

Another common super-resolution microscopic technique is stochastic optical reconstruction microscopy (STORM). Developed by Zhuang et al. in 2006, STORM is based on single-molecule localization microscopy (SMLM). SMLM refers to reconstructing the image of the sample from multiple images where a changing subset of the fluorophores are activated and recorded. The set of fluorophores activated at one time are spaced greater than Abbe’s resolution limit, so their images do not overlap, improving localization to sub-diffraction accuracy. STORM systems generally use organic dyes that stochastically activate and deactivate based on the illumination wavelength (Almada et al., 2015). STORM can produce images with a spatial resolution of under 20 nm (Rust et al., 2006). Photoactivated localization microscopy (PALM) is also an SMLM technique developed by Harald Hess and Eric Betzig in 2005 (Hess & Betzig, 2005). Unlike STORM, PALM uses labels whose fluorescence emission state can be directly controlled (Almada et al., 2015). PALM also uses another high-powered laser to photobleach already imaged fluorophores to reduce the probability of activating in subsequent imaging cycles. Experimental PALM systems have shown 10 nm resolution (Zhong, 2010). Though both
have nanometer lateral resolution, PALM and STORM rely on using specific staining dyes with uncommon fluorophore properties. Other limitations of PALM and STORM systems are their high probability of achieving photobleaching, inability to provide OS capability, and reduced imaging time, which is considerably slower than SIM (Almada et al., 2015).

Stimulated emission depletion microscopy (STED), developed in 1994 to improve the spatial resolution of confocal scanning microscopy, has gained substantial usage and development in the last ten years (Vicidomini et al., 2018). In fact, S.W. Hell received the 2014 Nobel Prize in Chemistry for his STED contribution to super-resolution microscopy. STED is a scanning-based microscopy imaging technique like confocal microscopy that breaks the diffraction limit by de-exciting fluorophores surrounding the illuminated point. This is achieved by encompassing the illumination beam with a STED beam, a beam with a doughnut-shaped spot distribution with a zero-intensity center. The STED beam de-excites the surrounding fluorophores with stimulated emission allowing only the fluorophore in the center to be activated by the illumination beam (Vicidomini et al., 2018). Like S-SIM, resolution in STED is only limited by SNR. Experimental images with a lateral resolution below 25 nm have been generated in STED imaging (Harke et al., 2008). Like confocal microscopy, STED also boasts optical sectioning capability. However, like PALM and STORM, STED relies on staining dyes with specific fluorophore properties and has a higher risk of photobleaching the sample (Vicidomini et al., 2018).

1.3: Super Resolution Confocal Microscopy

Despite the success of these super-resolution microscopic techniques in achieving impressive resolution power, they present two major drawbacks compared to confocal microscopy: (1) their high price and (2) lack of accessibility among the biological sciences community. Super-
resolution in confocal microscopy can be achieved by adding hardware components or with purely computational methods. The following section contains super-resolution methods used that remain in the confocal microscopy modality. Table 1 consolidates papers with the different confocal super-resolution methods mentioned below.

**Table 1: Consolidation of super-resolution confocal microscopy techniques discussed in the following sections.**

<table>
<thead>
<tr>
<th>Year</th>
<th>Type</th>
<th>Methods Discussed</th>
<th>Resolution Improved</th>
<th>Data Source</th>
<th>Objective Lens</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>Hardware Detection System</td>
<td>Image Scanning Microscopy, Pixel Reassignment Methods</td>
<td>1.18x</td>
<td>Experimental</td>
<td>UPLXAP O 60X NA 1.42</td>
<td>(Chen et al., 2016; Müller &amp; Enderlein, 2010)</td>
</tr>
<tr>
<td>2013</td>
<td>Hardware Detection System</td>
<td>Optical Photon Reassignment Microscopy (OPRM)</td>
<td>1.3x</td>
<td>Experimental</td>
<td>Plan-Apochrom at 63× NA .7 oil</td>
<td>(Roth et al., 2013)</td>
</tr>
<tr>
<td>2015</td>
<td>Hardware Detection System</td>
<td>OPRM and Spinning Disk</td>
<td>1.37x</td>
<td>Experimental</td>
<td>UPLSAP O100XO NA 1.4 oil</td>
<td>(Azuma &amp; Kei, 2015)</td>
</tr>
<tr>
<td>Year</td>
<td>Type</td>
<td>Methods Discussed</td>
<td>Resolution Improved</td>
<td>Data Source</td>
<td>Objective Lens</td>
<td>Ref</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------</td>
<td>----------------------------------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>----------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>2021</td>
<td>Hardware Detection System</td>
<td>Spatial Mode Sorter/Richardson-Lucy Deconvolution</td>
<td>1.3x</td>
<td>Simulation</td>
<td>Not Reported</td>
<td>(Bearne et al., 2021)</td>
</tr>
<tr>
<td>2017</td>
<td>Hardware Illumination System</td>
<td>Bessel-Gauss Beams</td>
<td>1.3x</td>
<td>Experimental</td>
<td>UPlanSApo 60X NA 1.2 water</td>
<td>(Thibon et al., 2017)</td>
</tr>
<tr>
<td>2018</td>
<td>Hardware Illumination System</td>
<td>Phase Masks to generate special PSFs</td>
<td>1.78x</td>
<td>Simulation</td>
<td>NA 1.41</td>
<td>(Le et al., 2018)</td>
</tr>
<tr>
<td>2001</td>
<td>Computation_al</td>
<td>Richard-Lucy Deconvolution, Bayesian Inference, Wavelet Denoising</td>
<td>1.2-1.5x depending on noise</td>
<td>Experimental</td>
<td>Zeiss Achroplan 40x NA0.75 water</td>
<td>(de Monvel et al., 2001)</td>
</tr>
<tr>
<td>2019</td>
<td>Computation_al/Deep Learning</td>
<td>Conditional Generative Adversarial Network</td>
<td>2.64x</td>
<td>Experimental</td>
<td>Leica HC PL APO 100x/1.40x-NA Oil</td>
<td>(W. Wang et al., 2021)</td>
</tr>
</tbody>
</table>
1.3.1: Hardware Approaches

Detection Based Enhancement

Traditional confocal microscope systems use a single pinhole and photomultiplier detector to record the illuminated point (Elliott, 2020). Rather than recording the detection point spread function (PSF) distribution, photomultipliers integrate the intensities recorded over the sensor, condensing the PSF distribution to a single value. Replacing the single pinhole and photomultiplier sensor with more complex systems enables the recording of more information. Pixel-reassignment is a technique that replaces the single pinhole with a pinhole array to sample the detected PSF. An algorithm then reassigns each-recorded intensity from the pinhole array based on probabilities of their actual location. This method requires the knowledge of the effective PSF of the confocal microscope system, theoretically improving resolution to a factor of $\sqrt{2}$ (Chen et al., 2016). Combining image scanning microscopy with the pixel reassignment method improved the resolution capability by decreasing the minimum resolvable distance by 18% of a traditional confocal microscope system (Chen et al., 2016; Müller & Enderlein, 2010). This method, however, increases image acquisition time. Optical photon reassignment microscopy (OPRA) applies the pixel reassignment concept but optically rather than computationally, which significantly reduces image acquisition time (Roth et al., 2013). A proposed system applied OPRA on a spinning disk confocal microscope system, reporting an improved resolution by a factor of 1.37 times (Azuma & Kei, 2015). Another optical pixel reassignment method was developed, called rescan confocal microscopy, which decreased the full-width half maximum (FWHM) of the experimental PSF by 30% (Chen et al., 2016; de Luca et al., 2013). Another work replaced the photomultiplier detector with a Zernike spatial mode sorter. The Zernike spatial mode sorter records components of the Zernike polynomial, enabling
the reconstruction of super-resolved images. Results of a simulated confocal microscope with a
Zernike spatial mode sorter showed a resolution improvement of 1.3 times (Bearne et al., 2021).

_Illumination Based Enhancement_

Confocal microscopes generally use lasers focused onto single points and scanned throughout the
sample. Conventional confocal microscope illumination systems use Gaussian beam lasers,
whose focus spot follows a Gaussian distribution. Bessel-Gauss beams as the illumination source
have been proposed (Thibon et al., 2017). Bessel-Gauss beams have a smaller focus spot
compared to Gaussian beams, leading to increased resolution. However, the side lobes in Bessel-
Gauss beams create unwanted artifacts and pull energy from the focal spot, thus reducing SNR.
In 2017, Thibon et al. adjusted the pinhole size to reduce side lobes' effects, achieving a
resolution improvement of 1.3 times compared to traditional confocal microscope systems
(Thibon et al., 2017). Another method to reduce the size of the point spread function used a
phase mask with a left-handed azimuthal polarization (Le et al., 2018). Similarly to the previous
method, energy is shifted to sidelobes which create artifacts in the recorded images.
Nevertheless, this method produced a PSF which resulted in a 1.78 times improvement in
resolution compared to traditional confocal microscope systems (Le et al., 2018).

1.3.2 Software

_Deconvolution_

The most common computational approach for resolution enhancement is deconvolution
algorithms. Since confocal microscopes are linear systems, the imaging process can be modeled
by a convolution of the object's fluorescence distribution and the system’s response (e.g., PSF).
Deconvolution algorithms try to reverse the optical imaging process by estimating the spatial
frequencies of the samples’ fluorescence distribution. Deconvolution algorithms are more
difficult to use in confocal imaging rather than widefield microscopy due to its lower SNR and
lack of precise ways to estimate the PSF (de Monvel et al., 2001). Nevertheless, deconvolution
algorithms have shown to produce desirable results. Monvel et al. have shown that the
Richardson-Lucy algorithm and Bayesian inference paired with wavelet denoising provide super-
resolved confocal imaging (de Monvel et al., 2001). For an image with SNR of 6.2 dB, known
exact PSF, and wavelet denoising, de Movel et al. demonstrated an improvement of $1.42\times$ and
$1.43\times$ with Bayesian inference and the Richardson-Lucy algorithm, respectively. For the same
case, but using an approximate PSF, the resolution improvement was reduced to 1.26 times for
the Bayesian inference and 1.27 times for the Richardson-Lucy algorithm. The results depends
significantly on the SNR. For example, knowing the exact PSF and increasing the SNR of the
image from 6.2 dB to 13.8 dB, the resolution has been improved by a factor 1.52 using the
Richardson-Lucy algorithm. Nonetheless, if one uses an approximate PSF, the improvement of
the resolution is not as great, being $1.26\times$ instead of $1.52\times$ (de Monvel et al., 2001).

Deconvolution methods can be used in conjunction with hardware-based approaches to further
increase their resolution enhancement capability.

Deep Learning

Deep learning (DL) has revolutionized many tasks in microscopy, such as object segmentation,
object classification, and image denoising. Different studies have also proved an enhancement of
spatial resolution using DL models in other imaging modalities, including STORM, SMLM,
widefield, SIM, and PALM (Yang et al., 2021). In particular, we would like to highlight Prof.
Ozcan's work at UCLA (Wang et al., 2019). His research group developed a DL model to
achieve super-resolution from low-resolution to high-resolution conventional fluorescent images
with a resolution improvement of 1.875 times. In that same work, the authors investigate a learning-based model for cross-modality super-resolution imaging from confocal to STED images, achieving a resolution improvement of 2.64 times by applying conditional generative adversarial networks (cGANs) (Wang et al., 2019). More recently, in 2021, Wang et al. investigated a U-Net model in confocal scanning microscopy to provide a resolution improvement of 1.3× (Wang et al., 2021).

1.4: Overview

Confocal microscopy is a powerful tool used by researchers in the biological sciences, but its resolution is diffraction limited. Super-resolution techniques have been developed that break this limit; however, they are expensive and not widely accessible to the biological community. Although hardware- and computational-based methods have been developed to break the diffraction limit in confocal imaging, they require either additional hardware or precise knowledge of the system’s point spread function. Deep learning has been shown to successfully improve resolution in confocal microscopy, achieving a 2.64 times improvement of resolution, turning confocal microscopes into virtual STED imaging platforms (Wang et al., 2019).

We propose implementing, training, and validating two learning-based models (a cGAN and a cycleGAN) for resolution enhancement in confocal microscopy. We have trained our model using simulated data to translate low-resolution (LR) confocal images to high-resolution (HR) confocal images. Because the output is still a confocal microscope image, the resolution limit of confocal microscopes is not broken; however, we allow low NA microscope objective lenses to achieve similar results to high NA objective lenses. Confocal systems with higher NA objectives have longer image acquisition times and a lower field of views (FOV). Since we are producing HR images from a low NA objective system, the confocal microscope system will
have a lower image acquisition time and higher FOV when compared to a traditional system using the high NA objective. Since high NA objective lenses are generally more expensive, our proposed method will also lower the cost of the system.

A limitation of the cGAN for image-to-image translation tasks is that the data used to train the model must be paired. This means that our LR and HR images must have the same magnification and orientation. Creating a confocal image dataset with paired images is difficult as one has to find the same cell with different microscope objectives and FOVs, matching the FOVs during the pre-processing stage. The cycleGAN model was developed for image-to-image translation without needing a paired training dataset. We propose to train a cycleGAN model for resolution enhancement in confocal microscope images. We first train a model with our simulated dataset to get a quantitative evaluation of our model. Then, we train another cycleGAN on experimental data for evaluating the framework on real confocal images.
Chapter II: Proposed Methodology

We investigate two DL algorithms for resolution enhancement in confocal microscopy: a conditional generative adversarial network (cGAN) and the cycleGAN. We have trained and validated both models using a simulated confocal dataset created by convolving images with PSFs of confocal microscopes with two different numerical apertures (NA). For the cycleGAN model, we have also trained the model on an experimental dataset recorded at the Integrated Microscopy Center, University of Memphis. This chapter discusses the principle of confocal imaging, the creation of the simulated dataset, the theory behind the learning-based models used in this Master thesis, and the framework for the models.

2.1: Confocal Scanning Microscopy

2.1.1: Principle

A confocal scanning microscope (CSM) comprises two identical subsystems: the illumination and collection systems. Figure 2.1 shows the operation principle of a CSM system. The monochromatic light emitting from a laser source is collimated, expanded, and focused onto the three-dimensional sample with a high-NA objective. The tightly focused beam illuminates all parts of the sample within the illumination cone. The small region surrounding the beam's focus receives very high illumination density. Assuming that the illumination wavelength is within the excitation band of the fluorescent dye, the sample emits fluorescent light whose spatial intensity distribution is proportional to the illumination intensity. Since the collection system is adjusted so that the focus of the objective lens is conjugated with a pinhole, the light emerging from such point is collected with the maximum intensity.
One of the main advantages of CSM systems is the OS capability (i.e., the optical removal of the out-of-focus information). Parts of the sample outside the in-focus plane receive a smaller illumination density, reducing the probability of fluorescent emission. Since these planes are not conjugated with the pinhole (e.g., image plane), the light is collected with very low efficiency. Therefore, the probability of collecting light from out-of-focus planes is much smaller than that one from in-focus planes. The point-based detection using the pinhole is responsible for providing optical sectioning capability.

Figure 2.1. Diagram of a Confocal Scanning Microscope.
On the other hand, CSM systems are not widefield systems, meaning that confocal images are obtained after a synchronized transverse and axial scanning of the sample, generating the 3D matrix corresponding to the intensities acquired at each sample’s volume position.

Assuming that the illumination and collection systems are telecentric-afocal systems, the intensity distribution of the image is given by

\[ I_{\text{conf}}(x, y, z) = O(x, y, z) \otimes_3 \text{PSF}_{\text{conf}}(x, y, z), \]

where \( \otimes_3 \) refers to the 3D convolution operator, \( O(x, y, z) \) is the function that accounts for the concentration of fluorescent emitters, and \( \text{PSF}_{\text{conf}} \) is the point spread function of the confocal system (Martínez-Corral & Saavedra, 2009). The \( \text{PSF}_{\text{conf}} \) function can be decomposed into the product of the intensity PSFs of the illumination and detection systems

\[ \text{PSF}_{\text{conf}}(x, y, z) = \text{PSF}_{\text{ill}}(x, y, z) \times \text{PSF}_{\text{det}}(\varepsilon x, \varepsilon y, \varepsilon z), \]

where \( \varepsilon = \lambda_{\text{ill}}/\lambda_{\text{det}} \) takes account of the difference in wavelength between the illumination (\( \lambda_{\text{ill}} \)) and detection (\( \lambda_{\text{det}} \)) systems (Martínez-Corral & Saavedra, 2009). For CSM systems that have identical pupil apertures, the illumination and detection PSFs are identical, following the PSF distribution of a widefield microscope given by square modulus of

\[ h(r, z) = -i \frac{2\pi}{\lambda} \exp \left\{ i\pi \frac{z}{\sin^2 \alpha} \right\} \int_0^\alpha p(\theta) J_0 \left( 2\pi r \frac{\sin \theta}{\sin \alpha} \right) \times \exp \left\{ -i2\pi z \frac{\sin^2 \theta}{2} \right\} \sin \theta \, d\theta, \]

where \( J_0 \) is Bessel function (Martínez-Corral & Saavedra, 2009). Note that Equation (2.3) is expressed on cylindrical coordinates (e.g., \( r \) and \( z \)). Since confocal microscopy is generally used on fluorescent samples, the only differences between the calculation of the illumination and
detection PSFs will be the wavelength and wave number used. The excitation, emission, and resultant confocal PSF are shown in Figure 2.2.

**Figure 2.2.** Simulated Airy disk PSFs of the excitation, emission, and confocal system coinciding with objective lenses of NA 0.4 and 1.4

### 2.1.2: Simulated Confocal Image Dataset

Using Equations 2.2 and 2.3, we created two PSFs corresponding to confocal systems with microscope objective lenses of NA 0.4 and 1.4. For both PSFs, the illumination source wavelength was 0.488 µm while the emission wavelength was 0.520 µm. The two PSFs are shown in Figure 2.2. For object fluorescence distributions, we used images from the Kinome Atlas from the Cell Imaging Library (Zhang et al., n.d.). The Kinome Atlas consists of images of kinases expressed in HeLa cells stained with DAPI and other fluorophores and imaged by a confocal microscope. We chose this dataset as it represents samples imaged by confocal
microscopes. Using Equation 2.1, we convolved both PSFs with each of the Kinome Atlas images used to create pairs of low- and high-resolution images. Realistic confocal images recorded by an optical imaging system are degraded by noise. Since the number of photons emitted by an excited fluorophore follows a Poisson distribution, the noise in confocal microscopy images is Poisson. Using the MATLAB \texttt{poissrnd} function, we degraded our high- and low- resolution image pairs with Poisson noise to create realistic simulated confocal microscope image datasets. We created three datasets of 907 LR and HR images, corresponding to SNR levels of 15 dB, 30 dB, and 45 dB, to test the performance of the cGAN learning-based model at different noise levels. Figure 2.3 shows three images in our simulated dataset for the three noise level. For training the cGAN model, each dataset had an 80-20 training/validation split (e.g., 726 training images and 181 validation images).
Figure 2.3. Examples of images in our simulated dataset. The leftmost column contains the image before convolution and noise degradation. Second and third columns are the low- and high-resolution images, respectively. Except for the first image in each, the rows correspond to different noise degradation levels.

2.1.3 Experimental Dataset

We have used the Nikon Ti-E A1rSi CSM system at the University of Memphis Integrated Microscopy Center to record our images of the red and green channels of bovine pulmonary artery endothelial cells stained with DAPI and other fluorophores. We recorded 101 HR images with a 60×/1.4 NA microscope objective lens with a field of view of 512×512 pixels². We also
recorded 10 LR images with a 10×/0.3 NA microscope objective lens with a field of view of 2048×2048 pixels\(^2\). To match the FOV and magnification between the image sets, we used MATLAB to magnify the LR images 6 times. MATLAB using bilinear interpolation, increasing the image size to 12288×12288 pixels\(^2\). Then, we cropped 2048×2048 windows of the magnified images to match the FOV of the HR images and resized them to 512×512 to match their sizes. We manually chose 101 images where the cellular structures were most visible. Figure 2.4 shows an example of recorded LR (before and after the preprocessing) and HR images on the experimental dataset. We then performed dataset augmentation on both the LR and HR image sets by rotating our images by 0, 90, 180, and 270 degrees and flipping them vertically and horizontally. Finally, we have a dataset of 606 LR and HR images, respectively. We again used an 80-20 training/validation split resulting in 485 training images, and 121 validation images.

![Image of recorded LR (before and after the preprocessing) and HR images on the experimental dataset.](image)

**Figure 2.4.** Example of recorded LR (before and after the preprocessing) and HR images on the experimental dataset.

### 2.2: Conditional Generative Adversarial Networks and CycleGAN

Artificial neural networks (ANN) form a subset of machine learning models that are based on biological neural networks. ANNs pass information through a series of layers, as shown in
Figure 2.5A. The first layer is called the input layer, the last layer is called the output layer, and the layers in between are called hidden layers. Each layer consists of neurons that input data from neurons in previous layers (from the user in the case of the input layer).

**Figure 2.5A Illustration of a fully-connected ANN.** As highlighted, the first layer is the input layer, the layers between the first and last are the hidden layers, and the last layer is the output layer. **Panel B** shows the individual process that each neuron in a layer executes. Nodes, \(x_1, x_2, \ldots, x_n\) refers to the data outputted by neurons in the previous layer and “\(b\)” is the added bias. The node sums these values and applies activation function \(f(\cdot)\). **Panel C** illustrates the activation functions that will be used in this document. The leftmost is the LeakyReLU function, the middle is the ReLU function, and the rightmost is the sigmoid function.
As shown in Figure 2.5B, the neurons then perform a weighted sum of the data, adds a bias to that value, and passes the value through an activation function (O’Shea & Nash, 2015a). Activation functions (examples shown in Figure 2.5C) add nonlinearity to the network, improving the model's generality. Finally, the neuron passes the data to neurons in subsequent layers or displays it if it is the output layer. An ANN can produce a variety of linear and nonlinear mappings between the input and output layers based on the neurons' weights, biases, and activation functions. Outside of the activation functions, which are user-defined, ANNs can self-tune the weights, biasing the network by training the network on prior paired input and output data. The most common method to train ANNs is called backpropagation. Backpropagation passes input data through the network, compares the network output to the paired output with a loss function, computes the gradient of the loss function based on the current weights and biases, and finally changes the weights and biases based on the gradients to lower the loss function (O’Shea & Nash, 2015b). The loss function will eventually converge with a large enough dataset, resulting in a trained ANN. ANNs can be trained in batches rather than one example at a time to reduce training time. This strategy also allows the network to take advantage of batch normalization, which further improves network efficiency by normalizing the inputs to each layer in each batch (Ioffe & Szegedy, 2015).

Convolutional Neural Networks (CNNs) were developed to use ANN concepts efficiently for image data. Each layer in a CNN is structured as a multi-dimensional set of neurons. Rather than a weighted sum, the layers are connected by the convolution operation with trainable kernels. The convolution operation, shown in Figure 2.6B, incrementally slides a kernel laterally across the layer. At each increment, an elementwise multiplication is taken between the kernel and the datapoints in the layer it overlaps, and then the values are summed. The output of the
convolution is a two-dimensional matrix whose values are the results of the summation placed corresponding to the position of the kernel.

![Figure 2.6](image)

**Figure 2.6.** A) Illustration of the convolution operation used between layers. Green block is the three-dimensional (3D) input layer, the pink, blue, and yellow blocks are 3D filters. Each filter produces one 2D output which is the stacked to produce a 3D output. Panel B shows convolution with a single kernel. The kernel is slid across the input layer. An inner product between the data in the kernel and the data in the input layer it overlaps is taken. The single value is then placed on the output layer in the location corresponding to where the kernel is in the input layer.

Each convolutional operation has three parameters: filters, stride, and padding. The filters parameter refers to the number of kernels used between each convolutional layer. As shown in Figure 2.6A, the outputs of the convolutions with each kernel are then stacked channel-wise to form the next layer. This gives the filters parameter control over the depth of the next layer. The stride operation refers to how the kernel increments as it slides through the layer. A stride of 1
would mean that the kernel will move one neuron at a time. The stride parameter controls the
spatial dimension of the next layer. The padding parameter refers to spatially surrounding the
layer with zeros which further controls the spatial dimensions of the next layer (O’Shea & Nash,
2015b). To further reduce network parameters and complexity, CNNs also can have pooling
layers. Pooling divides the layer into smaller windows, and then combines each window with a
certain operation. The most common pooling operation is a 2x2 max-pooling which takes the
maximum value in every nonoverlapping 2 by 2 window in the layer (O’Shea & Nash, 2015b).
Like ANNs, CNNs also use backpropagation to train the network.

2.2.1: cGAN Framework
Deep generative models try to create new data based on learning patterns found in the data,
training the model using these patterns. To reduce the complexity of this task, Goodfellow et al.
developed the generative adversarial network, a framework in which a generative model
competes against a discriminator model (Goodfellow et al., 2014a). The generative model creates
new data from an input noise vector, while the discriminator model determines whether the data
is generated or found in the training set. The discriminator is trained to maximize the likelihood
of correctly labeling the source of the data, while the generator is simultaneously trained to
minimize it. This min-max competition can be described by the following value function

\[
\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{x \sim p_{z}(x)}[\log (1 - D(G(z)))],
\]

where \( D(x) \) is the probability of correctly determining the data source and \( G(z) \) is the generated
data from the generator model (Goodfellow et al., 2014b). Theoretically, GANs are considered
trained when they achieve Nash equilibrium meaning that neither the generator nor discriminator
make significant improvement in output (Lucic et al., 2017). Nonetheless, in practice, this Nash equilibrium is difficult to achieve (Kodali et al., 2017).

GANs can be extended to a conditional model if an additional vector of prior information is inputted to the generator and discriminator. These models, termed conditional generative adversarial networks (cGAN), change the GAN value function to

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x|y)] + \mathbb{E}_{x \sim p_x(z)} [\log (1 - D(G(z|y)))]$$

where $y$ is the conditional data (Mirza & Osindero, 2014).

Isola et al. proposed a GAN model for the general image-to-image translation tasks coined pix2pix (Isola et al., 2016). Image-to-image translation refers to mapping images of one class to another. In this case, the GAN is a conditional model as the input image contains the prior information the model uses to generate the output image. Pix2pix employed a U-Net as the generator, and a CNN called PatchGAN as the discriminator. A U-Net is an encoder-decoder convolutional neural network that was first developed for biological image segmentation. Encoder-decoder networks can be decomposed into a contracting path (encoding) and an expanding path (decoding). The contracting layers downsample the data in the previous layer and double the number of features extracted till a bottleneck layer. Then, the expansion layers upsample the data in the previous layer and halves the number of features. The contraction and expansion layers of the same dimensions also share skip connections, passing information from the contracting layer to the expanding layer (Ronneberger et al., 2015). These connections exist because, in many image translation tasks, there is significant information shared between the inputs and outputs (Isola et al., 2016).
The PatchGAN discriminator is a CNN that determines if smaller $N \times N$ neighborhoods of the image rather than the whole image itself are real or fake. Each $N \times N$ patch can be described as a Markov random field, meaning that the pixels within the patch are independent of those outside. This effectively treats the PatchGAN component of the loss function as adding texture information (Isola et al., 2016). It has been shown that traditional loss functions such as L1 and L2 accurately pass low frequency information in image translation tasks. Therefore, the generator loss function consists of both a traditional loss function as well as the PatchGAN loss. This supplements the PatchGAN by having it only model the high frequency information. The resultant generator loss function becomes (Isola et al., 2016)

$$G^* = \arg \min_G \max_D \mathcal{L}_{\text{PatchGAN}}(G, D) + \lambda \mathcal{L}_{\text{Traditional}}(G).$$

We propose to train a pix2pix based cGAN for our task of lateral resolution enhancement. The U-net consists of 7 encoding layers that apply the convolution operation with kernels of size $4 \times 4$ and a stride of $2 \times 2$. The number of filters for the encoding blocks in order are 64, 128, 256, 512, 512, 512, 512. Each encoding layer uses the LeakyReLU function. The bottleneck layer also applies the convolution operation with kernels of size of $4 \times 4$ and stride of $2 \times 2$. It involves 512 filters and applies the classic ReLU activation function. The 7 decoding layers mirror the encoding layers, but use the convolution transpose function (deconvolution) with the same parameters as its encoding counterparts. The activation function used is the classic ReLU. A final convolutional layer is used after the decoding block with the same kernel size and stride. The number of filters applied corresponds to the number of channels the images in the dataset contain, and the tanh activation function is used. The patchGAN discriminator uses 5 convolutional layers with a kernel size of $4 \times 4$ with a stride of $2 \times 2$ and the LeakyReLU activation function. The number of filters used in order are 64, 128, 256, 512, 512. A final convolutional
layer is used with a kernel size of $4 \times 4$ a stride of $1 \times 1$, 1 filter, and the sigmoid activation function. This network architecture corresponds to a patch size of $70 \times 70$. The traditional loss function is mean absolute error, $|G(x) - y|$, also called L1 with $\lambda$ set to 100. The model will be trained on a simulated dataset that contains pairs of low-resolution and high-resolution confocal microscope images. Figure 2.7 shows the architecture and data flow diagram of the cGAN model used.

![Architecture and data flow diagram of the cGAN model used.](image)

**Figure 2.7. Architecture and data flow diagram of the cGAN model used.**

As mentioned previously, it is difficult for GANs to reach Nash equilibrium, therefore we have used other metrics to determine the quality and convergence of our model. The metrics we have chosen are peak signal-to-noise ratio (PSNR), structured similarity index measure (SSIM), and the background standard deviation of our images. PSNR is based on pixel value differences between an image and its ground truth. The PSNR value between images $f$ and $g$ is given by

$$PSNR(f, g) = 10 \log_{10} \left( \frac{\text{Max}^2}{\text{MSE}(f, g)} \right),$$

2.7
where Max is the maximum pixel value of the image and MSE is the mean squared error (Hore & Ziou, 2010)

\[
\text{MSE}(f, g) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (f_{ij} - g_{ij})^2,
\]

where \( M \) and \( N \) are the pixel dimensions of the image. We chose this metric to quantify the pixel value accuracy of our model’s outputs.

SSIM is a metric developed to compare images based on human visual system. The SSIM value between images \( f \) and \( g \) is given by

\[
\text{SSIM}(f, g) = \frac{(2\mu_f\mu_g + C_1)(2\sigma_{fg} + C_2)}{\left(\mu_f^2 + \mu_g^2 + C_1\right)\left(\sigma_f^2 + \sigma_g^2 + C_2\right)},
\]

where \( \mu \) is pixel value mean, \( \sigma_{fg} \) is the correlation coefficient between \( f \) and \( g \), \( \sigma_f \) and \( \sigma_g \) are the standard deviations of the images, and \( C_1 \) and \( C_2 \) are constants to avoid instability when the pixel value means are close to 0 (Z. Wang et al., 2004). \( C_1 \) and \( C_2 \) are

\[
C_i = (K_iL)^2,
\]

where \( K_i \) is a user-defined constant \(<1\) and \( L \) is the maximum value of pixel intensities. Rather than applying this equation on the whole image, SSIM is recommended to be used on local, circular, gaussian weighted windows of the image and then averaged (Wang et al., 2004). We used SSIM to quantify the structural accuracy of our model’s outputs. We used SSIM with a window size of \( 11 \times 11 \), gaussian weights with a standard deviation of 1.5, \( K_1 \) of 0.01, and \( K_2 \) of 0.03.
Finally, standard deviation value can be used to represent the amount of noise present in an image. To get a value that is not affected by the variations of pixel values in the cellular structure, we restrict our standard deviation calculation to just the background. We do this by applying Otsu’s method of thresholding the image. We chose this metric to confirm that our model does not add noise to the images.

2.2.2: cycleGANs

The pix2pix model is restricted to being trained with images that share the same general spatial orientation. Creating an experimental confocal dataset with LR and HR image pairs with the same spatial orientation poses many challenges. Among these difficulties, the most significant one is finding the same sample among the hundreds on a microscope slide with both microscope objective lenses. Due to this, we explore the application of the cycleGAN model for our experimental validation. The cycleGAN was developed by Zhu et al. for image-to-image translation with unpaired image data (Zhu et al., 2017). In our case, this means that rather than our training dataset containing LR inputs and their exact HR counterparts, the dataset will contain random LR images and random HR images. The cycleGAN is based on the concept of cycle consistency, a regularization technique that enforces the forward function with the backward function. This means that an image translated from domain A to B should be able to be accurately translated back from B to A. The cycleGAN model does this by training two cGANs with each other. One generator, which we call G, is tasked with learning the forward process, while the other generator, which we call F, learns the backward process. The cycle consistency loss can be described by

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} \left[\|F(G(x)) - x\|_1\right] + \mathbb{E}_{y \sim p_{data}(y)} \left[\|G(F(y)) - y\|_1\right]. \quad 2.11$$

The total loss function for the generators becomes,
\[ G^*, F^* = \arg \min_{G,F} \max_{D_x,D_y} \mathcal{L}_{GAN}(G, D_Y) + \mathcal{L}_{GAN}(F, D_X) + \lambda \mathcal{L}_{Cyc}(G, F) \]

where \( \mathcal{L}_{GAN} \) refers to the loss function for the cGANs (Zhu et al., 2017).

**Figure 2.8. Architecture and dataflow diagram for the cycleGAN model used.**

To circumvent the challenge of recording paired LR and HR confocal microscope images, we propose to train a cycleGAN for lateral resolution enhancement. Figure 2.8 shows the architecture and dataflow diagram for the cycleGAN model. The generators are U-Nets with 3 encoding layers, 9 residual layers, and 3 decoding layers. The first encoding layer uses 64 filters of size of 7×7, the second uses 128 filters of size of 3×3, the third uses 256 filters of size 3×3. Each encoding layer applies the ReLU activation at the end. Each residual layer applies a convolution with 256 filters of size 3×3 with the ReLU activation function, another convolution operation with the same number of filters and filter size, and then concatenates this output with the input of the block. The decoding layers mirror the encoding layer, with the last layer applying the tanh activation function rather than the ReLU function. The discriminators used are 70×70 patchGANs, which are the same as the pix2pix patchGAN. To test the accuracy of the cycleGAN
on completely unpaired data, we trained the model on a random half of the LR images and the other half of the HR images from the simulated dataset’s training split. This resulted in a training dataset of unpaired 362 LR images and 362 HR images. We also trained the cycleGAN with the experimental dataset.
Chapter III: Results and Numerical Validation

In this chapter, we provide the results of both learning-based model studies. For the cGAN model, we report the plots of our convergence metrics for different epochs as well as the results of the cGAN model using simulated datasets with different SNR levels. We also provide the results of the cycleGAN model using the high SNR simulated dataset and experimental images.

3.1 cGAN Results

During training, the image intensities were scaled from (0, 255) to (-1, 1). We have analyzed the cGAN model by comparing the generated HR images outputted by the model to their corresponding ground-truth HR images. Quantitatively, this comparison was performed by calculating the PSNR, SSIM, and background standard deviation values between the images. To calculate PSNR and SSIM, the model output and ground truth images were scaled to (0,1). For the background standard deviation, the images were scaled to (0, 255). For each epoch, we calculated these values after running our model with 50 images from the training and validation dataset. Figure 3.1 shows the average of the 50 values of the PSNR, SSIM, and background standard deviation values for the training and validation sets of varying SNR levels. From Figure 3.1, we see that the model convergence metrics fluctuate rapidly. This effect is more pronounced for the models training and processing lower SNR images. From Figure 3.1, we also notice that the model performs similarly for both the training and validation sets. For the 45 dB SNR dataset, the PSNR values fluctuate around 35 dB, the SSIM values fluctuate around 0.96, and the background standard deviation fluctuates around 7.5 AU. For the 30 dB SNR dataset, the PSNR values fluctuate around 34.5 dB, the SSIM values fluctuate around 0.95, and the background standard deviation fluctuates around 30 AU. For the 15 dB SNR dataset, the PSNR values fluctuate around 27 dB, the SSIM values fluctuate around 0.5, and the background standard deviation fluctuates around 75 AU.
deviation fluctuates around 20 AU. Clearly, the lower the SNR value on the images, the lower the performance of the cGAN model. We do, however, notice that the background standard deviation values unexpectedly fluctuate around a lower value for the 15dB SNR model than the 30 dB SNR model. This may be because of the structural inaccuracies of the 15 dB SNR model shown in Figure 3.3.

Figure 3.1. Performance of the cGAN model based on the convergence metrics (e.g., PSNR, SSIM and standard deviation of the background) between the generated high-resolution image and the ground truth. The blue and orange curves represent the behavior of the training and validation set, respectively.
To further show the capability of the model, Figure 3.2 shows the performance of the trained 45 dB cGAN for three different FOVs. Based on the convergence metrics shown in Figure 3.1, we chose epoch 30 to produce our predicted HR images. We also show the line profiles along the cellular structure of the LR, HR, and predicted HR 45dB SNR images to view the structural accuracy of the model outputs. Overall, the cGAN model generates HR images with significant structural accuracy in high SNR images.

<table>
<thead>
<tr>
<th>FOV 1</th>
<th>Low Resolution</th>
<th>High Resolution</th>
<th>Predicted</th>
<th>Line Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR: 32, SSIM: 0.96</td>
<td>PSNR: Inf, SSIM: 1</td>
<td>PSNR: 43, SSIM: 0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOV 2</td>
<td>PSNR: 29, SSIM: 0.88</td>
<td>PSNR: Inf, SSIM: 1</td>
<td>PSNR: 39, SSIM: 0.96</td>
<td></td>
</tr>
<tr>
<td>FOV 3</td>
<td>PSNR: 27, SSIM: 0.88</td>
<td>PSNR: Inf, SSIM: 1</td>
<td>PSNR: 39, SSIM: 0.97</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.2.** Three examples of the performance of the cGAN model from the 45dB SNR validation dataset showing the low-resolution, high-resolution ground truth, and cGAN predicted high-resolution images. The last column shows the line profiles along cellular structures.
Figure 3.3 compares the performances of the model for the same FOV but different SNR by showing its ground-truth HR image, predicted HR image, and line profiles along the cellular structure. Based on the convergence metrics shown in Figure 3.1, we chose epoch 60, and epoch 90 for the 30- and 15-dB models, respectively. By comparing the ground-truth line profiles between the 45-, 30-, and 15-dB SNR images, we can see that the noise presents itself in the line profiles as a jaggedness in the curves. Based on the 30- and 15-dB ground truth and predicted line profiles, we see that the cGAN model is unable to reconstruct the noise present in the ground truth high-resolution images. This gives our model a denoising effect, at the expense of structural accuracy.
Figure 3.3. Performance of the cGAN model at different SNR levels using the validation dataset. In the line profiles plots, the orange line corresponds to the ground-truth HR images, and the green line corresponds to the model-predicted HR images.

We have compared the Fourier transforms of the LR and the predicted HR images to quantify the resolution improvement. Since finer details correspond to high-frequency content, we should expect the predicted HR images to have a larger radius of signal content in the Fourier domain. Figure 3.4 shows the Fourier transforms of the high-resolution (e.g., ground truth), predicted high-resolution, and low-resolution images for one FOV in the 45 dB SNR validation set. The radii of the circles were calculated by taking a diagonal profile from the center of the Fourier transform image to the top right corner. The profile was then curve fitted to an
exponential function to help ease analysis, as shown in Figure 3.5. The radius was found at the point at which the curve was 1% of the max value of the line profile.

**Figure 3.4.** Evaluation of the resolution improvement achieved by the cGAN model using a single FOV from the 45dB SNR dataset. The Fourier transforms contain a circle which surround the area in which most of the signal’s energy is contained.
Figure 3.5. Line profile from the center of the Fourier transform image to the upper right corner along with its exponential fits. The red line corresponds to the line profile with noise and the blue line corresponds to the exponential fit. The blue shaded area ends at the cutoff frequency defined as the location where the curve is 1% of the max value of the line profile.

The resolution improvement is estimated by dividing the radius of the generated HR Fourier transform image with the radius of the LR Fourier transform image. For an estimate of the resolution enhancement performance of the models, we took ten images from the validation set and averaged the resolution enhancement for each of the images, as shown in Figure 3.6-3.8. For comparison, we also calculate the resolution improvement between the LR and the ground-truth HR images with the same method. In summary, the 45-dB model produced images with a resolution improvement (mean ± standard deviation) of $2.45 \pm 0.543\times$ for the LR/predicted HR comparison vs. $2.065 \pm 0.204\times$ the LR/ground-truth HR comparison. The resolution improvement of the 30dB model is $2.750 \pm 0.857\times$ for the LR/predicted HR comparison vs.
2.181 ± 0.810× for the LR/ground-truth HR comparison. The 15dB model produced the lowest resolution improvement at 1.819 ± 0.153× for the LR/predicted HR comparison and 2.065 ± 0.1621× for the LR/ground-truth HR comparison. Surprisingly, the average resolution enhancements for both the model output and ground-truth high resolution are higher for the 30 dB SNR model than the 45 dB SNR model. However, the standard deviation of the 30 dB SNR model is much higher. This emphasizes that the resolution enhancement achieved by the cGAN model is highly sample and noise dependent. We also conclude that our resolution enhancement estimation does not correlate to the accuracy of the model’s outputs.

**Figure 3.6.** Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 45 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair.
Figure 3.7. Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 30 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair.

Figure 3.8 Low-resolution and cGAN predicted high-resolution Fourier transform image pairs from the 15 dB SNR dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair.

3.2 cycleGAN Results

This section is devoted to analyzing the performance of the cycleGAN model, which does not require paired datasets. We split cycleGAN results into two sections: the simulated dataset results and the experimental dataset results.

3.2.1 cycleGAN Simulated Dataset Results

Like with the cGAN model, we have analyzed the cycleGAN model by comparing the generated HR images outputted by the model to their corresponding ground truth HR images. Quantitatively, this was done by calculating the PSNR, SSIM, and background standard deviation values between the images. The image intensities were scaled to (-1, 1) for training, (0,
1) for calculating PSNR and SSIM, and (0, 255) for calculating background standard deviation. For each epoch, we calculated these values after running our model with 50 images from the training and validation dataset. Figure 3.9 shows the average PSNR, SSIM, and background standard deviation for different epochs. For the cycleGAN model, we only trained on the 45 dB SNR dataset. The cycleGAN model converges at lower PSNR and SSIM values and higher background standard deviation values than its cGAN counterpart for the same SNR level. The validation PSNR values converge around 27.5 dB, the SSIM values around 0.65, and the background standard deviation around 30.

![Figure 3.9](image)

**Figure 3.9.** Performance of the cycleGAN model based on the convergence metrics (e.g., PSNR, SSIM, and standard deviation of the background) between the generated high-resolution image and the ground truth. The blue and orange curves represent the behavior of the training and validation set, respectively.

Figure 3.10 shows the performance of the cycleGAN model of three examples of the training and validation dataset. Based on the convergence metrics, we chose to use the model at epoch 50 to produce the predicted high-resolution images. We have shown the low-resolution (first column), the ground truth (second column), and the predicted high-resolution (third column) images. For comparison purposes, we have also plotted lines profiles along cellular
structures. Overall, the similarity between the ground-truth and predicted HR images is high. However, we noticed that the model seems to add spikiness to the curves of the generated high-resolution images. In FOV 3, we also see that the model adds a slight shift as well.

![Figure 3.10](image)

**Figure 3.10.** Three examples of the performance of the cycleGAN model from the 45dB SNR validation dataset showing the low-resolution, high-resolution ground truth, and cycleGAN predicted high-resolution images. The last column shows the line profiles along cellular structures.

As before, we have analyzed the Fourier spectrum of the images to quantify the resolution improvement. Figure 3.11 shows ten Fourier transform pairs of low-resolution and predicted high-resolution images used to quantify the resolution enhancement of the cycle
model. We found that the cycleGAN achieved an average resolution enhancement of 2.454 with a standard deviation of 0.299. In terms of resolution enhancement for the 45 dB SNR dataset, the cycleGAN produces similar results to the cGAN model whose resolution enhancement was $2.45 \times \pm 0.543 \times$. However, we see that the structural accuracy and PSNR values are higher for the cGAN model. We assume that this could be because the generator used in the cGAN model has many more parameters than the generator in the cycleGAN model. Also, this result can be related to the fact that the learning task is generally easier when the dataset is paired.

3.2.2 cycleGAN Experimental Dataset Results.

Finally, we have trained and validated the cycleGAN model using our experimental dataset. Since we do not have the corresponding ground-truth images for each LR image, we are not able to calculate and plot the convergence metrics used previously. Nevertheless, we will compare line profiles between LR and predicted HR images and quantify the resolution enhancement based on the cutoff frequency estimated from their Fourier transform content. Since the experimental dataset contains color images, we converted the images to grayscale before taking

**Figure 3.11.** Low-resolution and predicted high-resolution Fourier transform image pairs from the 45 dB SNR dataset for the cycleGAN model. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the LR and predicted HR images are shown below each pair.
the line profiles and calculating Fourier transforms. We produced the predicted HR images for the HR generator trained at epoch 100.

<table>
<thead>
<tr>
<th>FOV 1</th>
<th>FOV 2</th>
<th>FOV 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Low-Resolution" /></td>
<td><img src="image2.png" alt="Predicted" /></td>
<td><img src="image3.png" alt="Line Profiles" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Low-Resolution" /></td>
<td><img src="image5.png" alt="Predicted" /></td>
<td><img src="image6.png" alt="Line Profiles" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Low-Resolution" /></td>
<td><img src="image8.png" alt="Predicted" /></td>
<td><img src="image9.png" alt="Line Profiles" /></td>
</tr>
</tbody>
</table>

**Figure 3.12.** Experimental performance of the cycleGAN model to generate high-resolution images from low-resolution confocal images.

Figure 3.12 shows that the cycleGAN model generally generates images with lower intensity values than the LR images, but with more defined edges, enabling us to visualize more structural details previously hidden. Though, experimentally, we cannot evaluate the structural
accuracy of the cycleGAN model due to the lack of the ground-truth dataset, it is clear that the cycleGAN model generates outputs with more high-frequency content. To confirm this, we have compared the Fourier transforms of the low-resolution and predicted high-resolution images, see Figure 3.13. Qualitatively, we can see that the cycleGAN model predicts images with more high frequency content. However, because the Fourier transforms of these images are not circular like those found in the simulated dataset, we cannot use the method used on the simulated dataset to quantify resolution enhancement. This is because the diagonal line profiles of the Fourier transforms do not flatten out as they did for the simulated dataset. To reduce this effect on our resolution enhancement calculation, we set the cutoff frequency as 5% of the max value of the line profile rather than 1%. Figure 3.14 shows the calculated cutoff frequency from the low-resolution and predicted high-resolution images. The cycleGAN model generates experimental high-resolution images with a resolution enhancement on average of $1.829 \times$ and a standard deviation of $0.8179 \times$. For comparison purposes, we have also calculated resolution enhancement using the 5% cutoff frequency for the cycleGAN trained on the 45dB SNR simulated dataset for comparison. We found that with this calculation, the resolution enhancement of the 45 dB SNR cycleGAN model was $2.929 \pm 0.219 \times$. As we saw previously, the cGAN performs significantly worse on lower SNR images. Because of this, we believe that the difference in noise level between the 45 dB simulated dataset and the experimental dataset is what caused this stark difference in resolution enhancement capability in our two cycleGAN tests.
**Figure 3.13.** Comparison of the Fourier Transforms of the low-resolution and predicted high-resolution images to show the increase of the spatial frequency content.

**Figure 3.14.** Low-resolution and cycleGAN predicted high-resolution Fourier transform image pairs from the experimental dataset. Circles whose radii correspond to the cutoff frequency of the Fourier transform are superimposed on each image. Resolution enhancement achieved between the low- and predicted high-resolution images are shown below each pair.
Chapter IV: Conclusion

Confocal Microscopy is a widely available imaging modality that produces multi-colored, optically-sectioned images. One of the major drawbacks of confocal microscopy is its diffraction-limited spatial resolution. Super-resolution confocal systems have been investigated by requiring additional hardware or computational approaches. Deep learning models are computational methods that, once trained, do not require any prior information about the confocal microscope system, being more user-friendly. In this Master's thesis, we have trained and validated two deep learning algorithms to overcome the diffraction resolution limit in confocal microscopy: the cGAN and the cycleGAN. The cGAN model has been trained and tested on three simulated datasets with different noise levels: a 15 dB SNR dataset, a 30 dB SNR dataset, and a 45 dB SNR dataset. On the other hand, the cycleGAN has been tested on the 45 dB SNR simulated dataset and an experimental dataset with images recorded at the University of Memphis Integrated Microscopy Center.

Using line profiles through the cellular structures in our images, we find that the cGAN model produces accurate HR images when trained with the 45 dB SNR dataset. But, the performance of the cGAN models fails in the presence of noise. Although the cGAN model seems useful as a denoising tool, it loses its accuracy in producing high-frequency structural details. By analyzing the Fourier content of the low-resolution and predicted high-resolution images, we estimate a resolution improvement of $2.451 \pm 0.543$ times for the 45dB SNR dataset, $2.750 \pm 0.857$ times for the 30dB SNR dataset, and $1.819 \pm 0.153$ times for the 15dB SNR dataset. These results show that the resolution improvement is sample dependent, as shown by the high values on the standard deviation, and heavily noise dependent, as shown by the low performance of the 15-dB-SNR cGAN model.
The cycleGAN model was trained and validated on the 45dB SNR simulated dataset used for the cGAN model but without pairing the images. Overall, the cycleGAN model generates high-resolution images with high similarity to the ground truth high-resolution image. However, the cycleGAN model seems to add spikiness to the generated high-resolution images, shown in the line profiles. By analyzing the frequency spectrum, we find that the cycleGAN model achieves a resolution enhancement of 2.454 ±0.299 times to the low-resolution input using the simulated dataset. The generated high-resolution images provided by the cycleGAN model have, in general, lower intensity values than the low-resolution images, but they present more defined edges and less blur, enabling us to visualize more structural details previously hidden. Experimentally, the cycleGAN model provides a resolution enhancement of 1.829 ±0.8179 times to the low-resolution experimental images.

Based on these results, we conclude that the cGAN model can produce accurate high-resolution confocal images from low-resolution images as long as the training dataset contains high SNR paired images. The cycleGAN model also has the potential to be used for resolution improvement, though in our simulated results, it performed worse than the cGAN model. This result may be related to the fact that the generators used in the cycleGAN model did not have as many parameters as the cGAN generator did. Future work for this project includes the creation of a larger experimental dataset of low- and high-resolution confocal images with more sample diversity, experimentation with different generator loss functions, and different sizes of the generators used in the cycleGAN models. We will also explore another general hyper-parameter tuning.
References


https://doi.org/10.1016/j.ymeth.2015.06.004


https://doi.org/10.1364/oe.23.015003


48

https://doi.org/10.1016/S0006-3495(01)76214-5


https://doi.org/10.1046/j.1365-2818.2000.00710.x


https://doi.org/10.1073/pnas.0406877102


https://doi.org/10.1364/oe.16.004154


https://doi.org/10.1038/nmeth929

https://doi.org/10.1038/s41556-018-0251-8


https://doi.org/10.1364/oe.25.002162.

https://doi.org/10.1038/s41592-018-0239-0

https://doi.org/10.1364/ol.440561


https://doi.org/10.1101/pdb.top91

Appendix

PSF Convolution (MATLAB):

```matlab
a = dir('*.*tif');

psf = load('psf_lowNA.dat');
psf = imresize(psf, [128,128]);
psf = gpuArray(psf);

for i = 1:length(a)
    img = imread(a(i).name);
    img = rgb2gray(img);
    img = gpuArray(img);
    conimg = conv2(img,psf,'same');
    conimg = uint16(conimg);
    imwrite(conimg, ['../kinomeatlasLowNA/LowNA' a(i).name]);
end
```

AddPoissnNoise (MATLAB):

```matlab
function [g_N SNR] = AddPoissnNoise(g, scale)

% Author: Ana Doblas; Date: 12/06/2018

% This function adds noisy to the 3D forward image of SIM
```
Note that the input contains the 3D forward image for the three different phases X, Y.

Scale is an user defined parameter. Changing this parameter we will tune the SNR value.

% b = % of background signal. [1.5% to 2%]
% Syntax: [noisy_image SNR] = AddPoisnNoise(image, scale)

[X Y Z Ph_no]=size(g);

b = 0.0168; % fixed value of the background level
no = 10;

for i = 1:Ph_no
    i
    g_Poisson=zeros(X,Y,Z);
    temp = g(:,:,i);
    g_noisy = temp./max(temp(:))*scale;
    g_noisy = g_noisy + g_noisy./max(g_noisy(:))*b;

    % Repeat the process 100 times
    for k = 1:no
        k
        g_Poisson = g_Poisson+poissrnd(g_noisy);
    end
g_Poisson = g_Poisson/no;

mu = mean(sqrt(g_Poisson(:)));
SNR(i) = 10*log10(mu);

g_N (:,:,i)=g_Poisson(:,:);

clear temp

clear g_noisy

clear g_Poisson

clear mu

end

end

Adding Poisson Noise To All Dataset (MATLAB):

a = dir('kinomeatlasHighNA/*.tif');

for i = 1:length(a)

    highNApath = ['kinomeatlasHighNA/' a(i).name];
lowNApath = ['kinomeatlasLowNA/' a(i).name];

highNAimg = imread(highNApath);
highNAimg = double(highNAimg);

lowNAimg = imread(lowNApath);
lowNAimg = double(lowNAimg);

noisyHighNA = AddPoissonNoise(highNAimg,0.015,10);
noisyLowNA = AddPoissonNoise(lowNAimg,0.015,10);

noisyHighNA = rescale(noisyHighNA,0,255);
noisyHighNA = uint8(noisyHighNA);

noisyLowNA = rescale(noisyLowNA,0,255);
noisyLowNA = uint8(noisyLowNA);

imwrite(noisyHighNA, ['./KinomeAtlasHighNA5dB/' a(i).name]);
imwrite(noisyLowNA, ['./KinomeAtlasLowNA5dB/' a(i).name]);

end
Dataset Split (MATLAB):

clear;
clc;

validSplit = .2;

rng(692);

a = dir('*tif');

datasetLen = length(a);

numValid = uint32(validSplit*datasetLen);

b = randperm(datasetLen,numValid);

for i = b

    filename = a(i).name;

    movefile(filename,'ValidSet')

end