2022-09

DEBLURRING OF OPTICALLY ABERRATED SATELLITE IMAGERY WITH DEEP LEARNING (UNET)

Siew, Jun Jie

Monterey, CA; Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

DEBLURRING OF OPTICALLY ABERRATED SATELLITE IMAGERY WITH DEEP LEARNING (UNET)

by

Jun Jie Siew

September 2022

Thesis Advisor:	Brij N. Agrawal
Co-Advisor:	Jae Jun Kim
Second Reader:	Leonardo Herrera, National Research Council (NRC)

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2022	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
DEBLURRING OF OPTICALLY ABERRATED SATELLITE IMAGERY WITH DEEP LEARNING (UNET)

5. FUNDING NUMBERS

6. AUTHOR(S) Jun Jie Siew

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

Satellite imaging performance can degrade due to optical aberrations. To maximize a satellite's imaging output over its useful lifespan, deep learning presents a cost-effective alternative to traditional adaptive optics for deblurring satellite images. This is because deep learning is essentially a post-processing technique that relies on algorithms and a large dataset. This research focuses on applying deep learning algorithms based on the UNET Convolutional Neural Network, which is widely used in the bio-medical imaging field, to deblur optically aberrated satellite imagery. The XVIEW dataset, which is composed of images taken by the Worldview-3 satellite, is used. The XVIEW images are then simulated with optical aberrations (defocus and spherical) using Zernike polynomials. The blurred images are subsequently deblurred with UNET and UNET variants (UNET++ and UNET3+) before final performance evaluation with various image quality metrics. The results showed that (1) UNET algorithms can effectively deblur optically aberrated satellite images, and (2) UNET3+ modified with additional convolutional layers (deep-UNET3+) provided the best deblurring performance. Based on the positive results, this thesis recommends that the UNET algorithm be applied on actual field cases of optically aberrated satellite imagery and be further developed to perform better even on super-resolution applications.

14. SUBJECT TERMS deep learning, computer vision, satellite imagery, deburring	**15. NUMBER OF PAGES** 55
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

DEBLURRING OF OPTICALLY ABERRATED SATELLITE IMAGERY WITH DEEP LEARNING (UNET)

Jun Jie Siew
Military Expert 5, Singapore Army
BME, Nanyang Technological University Singapore, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SPACE SYSTEMS OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
September 2022

Approved by: Brij N. Agrawal
Advisor

Jae Jun Kim
Co-Advisor

Leonardo Herrera
Second Reader

James H. Newman
Chair, Space Systems Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Satellite imaging performance can degrade due to optical aberrations. To maximize a satellite's imaging output over its useful lifespan, deep learning presents a cost-effective alternative to traditional adaptive optics for deblurring satellite images. This is because deep learning is essentially a post-processing technique that relies on algorithms and a large dataset. This research focuses on applying deep learning algorithms based on the UNET Convolutional Neural Network, which is widely used in the bio-medical imaging field, to deblur optically aberrated satellite imagery. The XVIEW dataset, which is composed of images taken by the Worldview-3 satellite, is used. The XVIEW images are then simulated with optical aberrations (defocus and spherical) using Zernike polynomials. The blurred images are subsequently deblurred with UNET and UNET variants (UNET++ and UNET3+) before final performance evaluation with various image quality metrics. The results showed that (1) UNET algorithms can effectively deblur optically aberrated satellite images, and (2) UNET3+ modified with additional convolutional layers (deep-UNET3+) provided the best deblurring performance. Based on the positive results, this thesis recommends that the UNET algorithm be applied on actual field cases of optically aberrated satellite imagery and be further developed to perform better even on super-resolution applications.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
CAE	convolutional autoencoder
CNN	convolutional neural network
EO	electro-optical
ESF	edge spread function
FFT	fast Fourier transform
GAN	generative adversarial network
GSD	ground sample distance
IQM	image quality metric
LSF	line spread function
ML	machine learning
MTF	modulation transfer function
MSE	mean squared error
PSNR	peak signal-to-noise ratio
RGB	red-green-blue
SSIM	structural similarity index
PSF	point spread function

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I thank my advisors, Distinguished Professor Dr. Brij N. Agrawal, and Dr. Jae Jun Kim, for their guidance, expertise, and support in this thesis research. I also thank my second reader, Dr Leonardo Herrera, for providing support in MATLAB coding.

I thank the Naval Postgraduate School (NPS) for its excellent academic faculty and infrastructure support that greatly aided my learning experience.

I thank the Space Systems Academic Group (SSAG) faculty led by the esteemed Chair James Hansen Newman.

Finally, I thank the Singapore Armed Forces (SAF) Command, Control, Communications, Computer, and Intelligence (C4I) community for providing me the opportunity to further my studies at NPS as well as National University of Singapore (NUS) Temasek Defence Systems Institute (TDSI) and NPS International Graduate Programs Office (IGPO) for their administrative support.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

High resolution electro-optical (EO) satellite imaging in the visible light spectrum, is important for commercial and military applications of earth observation. However, maximum ground sample distance (GSD) resolution is limited by the physics of diffraction and consequently is a function of the imager's aperture size (the larger the size, the higher the resolution). In addition, the diffraction-limited imaging performance of a satellite can be further degraded by optical aberrations such as defocus and spherical aberrations, which distort the image quality. Optical aberrations can be caused by pre-launch factors such as imperfect optics and design flaws, and/or post-launch factors such as degradation of optics and misalignment of optical system components during operations in the space environment. The result of optical aberrations on satellite imaging is image blurring.

Once a satellite is in operations, there are two main computer vision post-processing approaches to restoring raw blurred images caused by optical aberrations: (1) conventional physics-driven approach with image optimization routines, and (2) data-driven approach using Artificial Intelligence (AI) to predict image using a technique called deep learning [1]. The premise of deep learning is that given a dataset that is adequately representative of the problem (blurred image) and solution (clean image), an optimized deep neural network that has been trained with relevant data can quickly restore a new blurred image in real time, which can be more efficient than traditional image-by-image optimization methods [2]. High potential use cases of deep learning in restoring optically aberrated blurred imagery can be described in the following scenarios: (1) enhancement - using clean imageries from other satellites with similar GSD resolution but better optical performance and (2) sustainment - using previously clean imageries from the same degraded satellite.

As higher resolution imaging satellites require high optical performance (maximum of 30 nm root mean square [RMS]) to counter the effects of optical aberrations, this can significantly increase the project cost, schedule, and risk. This issue is most pertinent with wide aperture telescope exceeding 3m, thus requiring segmented and deployable mirrors, which necessitates complex hardware such as adaptive optics to correct optical aberrations.

As articulated, the potential of deep learning to restore satellite imagery with optical aberrations presents an opportunity to reduce satellite cost by relaxing optical performance requirements.

A. RESEARCH PURPOSE

The main purpose of this thesis is to investigate the deblurring performance of deep learning as an end-to-end solution in restoring the optically aberrated high-resolution Worldview-3 Satellite imagery. Specifically, a popular deep learning convolutional neural network (CNN) architecture known as UNET will be evaluated. UNET, first applied in biomedical imaging for semantic segmentation [3], has also found recent success in various deblurring applications employing variants of UNET [4–6].

B. RESEARCH OBJECTIVES

There are three main research objectives. The first objective is to evaluate the deblurring performances for the common defocus and spherical aberrations, both respectively and in combination. The second objective is to evaluate the performances of UNET and state-of-the-art UNET variants, namely UNET++ [7] and UNET3+ [8] for deblurring of high-resolution satellite imagery. The third objective is to evaluate the performances of UNET and UNET variants with various quantitative image quality metrices (IQM) as well as qualitative visual inspection.

C. THESIS ORGANIZATION

Chapter II provides the technical background of (1) fundamentals of optical imaging and computer vision; (2) effects of defocus and spherical aberrations; (3) modelling of optical aberrations with Zernike polynomials; (4) fundamentals of machine learning (ML), deep learning, and CNN; and (5) a review of UNET. Chapter III details the simulation of blurred imagery as well as the modeling and implementation of UNET and UNET variants for image deblurring. Chapter IV provides the detailed results and analysis of the deblurring performances of UNET and UNET variants. Chapter V summarizes the

thesis research objectives and corresponding findings and insights, as well as propose future works to improve UNET deblurring performance.

THIS PAGE INTENTIONALLY LEFT BLANK

II. TECHNICAL BACKGROUND

This chapter provides the technical background for the following: (1) limitation of EO systems and computer vision sensing, (2) effects of optical aberrations on satellite imaging, (3) modelling of optical aberrations, (4) image deblurring with deep learning, and (5) broad review of UNET and UNET variants.

A. ELECTRO OPTICAL SATELLITE IMAGING SYSTEM

1. Diffraction Limitation of Optical Systems

For simplicity of explanation, consider a satellite imaging system which uses a refracting telescope with curved lens to direct light to a focal plane where the light photons are captured. When light passes through the lens, a circular aperture, the phenomenon known as diffraction occurs, which is the bending of light's wavefront. The result is the distribution of light intensity, forming an airy disk pattern of light on the focal plane, with a series of maxima and minima as shown in the red image intensity Sinc function in Figure 1, representing the wave-like nature of light.

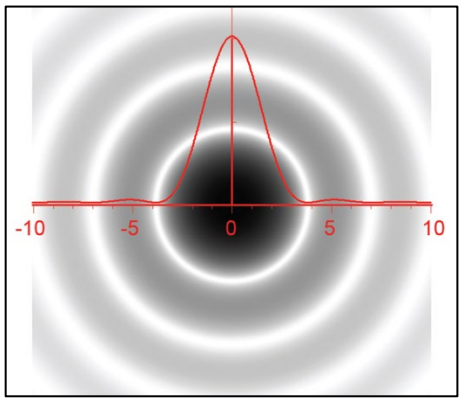

Figure 1. Diffraction Airy Pattern for Circular Aperture.
Source: [9].

The first maxima, denoted in black, represents the peak light intensity, while the first minima, denoted in white, indicates the boundary of the first wave as well as the

angular resolution ($\Delta\theta$) of the camera lens and can be described by the Rayleigh criteria formula for circular aperture. The Rayleigh criteria implies that the resolution of an imaging system is limited by wavelength of light (λ) and the size (diameter) of the lens aperture

$$\Delta\,\theta = 1.22\times \frac{\lambda}{diameter}.$$

To obtain the diffraction limited GSD resolution (X) of a satellite imaging system, multiply the angular resolution (θ) by the range (R) from the lens to the target (see Figure 2). The GSD resolution indicates how well the imaging sensor can discriminate two objects that are X distance apart, the smaller the GSD resolution, the better it can detect smaller targets.

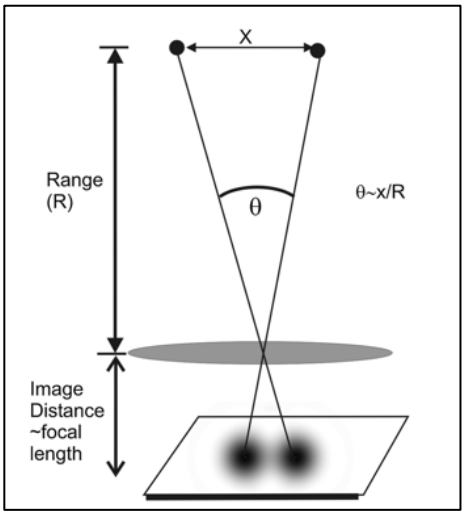

Figure 2. GSD Resolution X as Described by Rayleigh Criteria. Source: [9].

2. Computer Vision Image Sensing

Computer vision is essentially the field of making machines see. While the human eye is a versatile vision capability, it is typically not good at quantitative tasks, such as taking measurements, something the machine can be made to excel in. This section describes briefly how computers interpret images from an imaging sensor. Light photons,

reflecting off a target, pass through the telescope aperture and are collected on the camera focal plane, which is an array of pixels. When light photons interact with the camera focal plane made of semi-conductors, electrons are released. The movement of electrons create currents that can be measured, allowing the light intensity of each pixel on the focal plane to be determined. The Analog measurement for each pixel is converted digitally to a number value and stored by the computer in typical data sizes of 8 bits, thus storing values between 0 to 255. Essentially, a panchromatic image as interpreted by a computer is simply a matrix of number values as in Figure 3, representing the intensity of light between 0 (black) and 255 (white). A color image has three channels (red, green, and blue [RGB]) for each pixel.

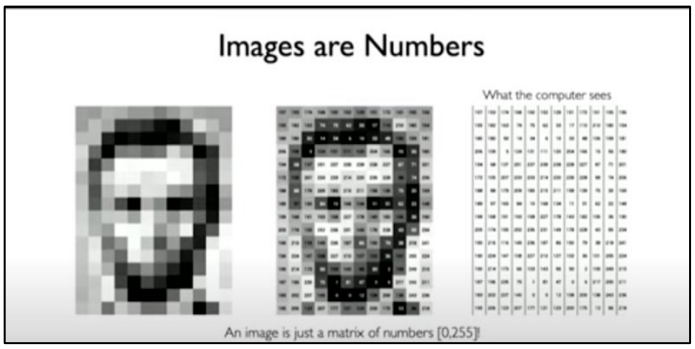

Figure 3. Image as a Matrix of Numbers.
Source: [10].

B. EFFECTS OF OPTICAL ABERRATIONS ON SATELLITE IMAGING

Further reducing the diffraction limited GSD resolution of EO systems are optical aberrations. Technically, optical aberrations occur when light reflected from a point source of the target do not converge perfectly at the same point on the image focal plane, where light photons are collected. Two specific optical aberrations for the purpose of this thesis are discussed, namely, (1) defocus aberration and (2) spherical aberration.

1. Defocus Aberration

The defocus aberration occurs when there is a wavefront error and the image formed deviates longitudinally away from the ideal gaussian image point (best focal point). The result is a diminished intensity at the first central maxima of the image intensity Sinc function, with the energy spreading to the adjacent wave maximas or intensity rings. The larger the wavefront error, the lower the central maxima intensity and the bigger the defocus ring, resulting in more severe blurring, as seen in Figure 4.

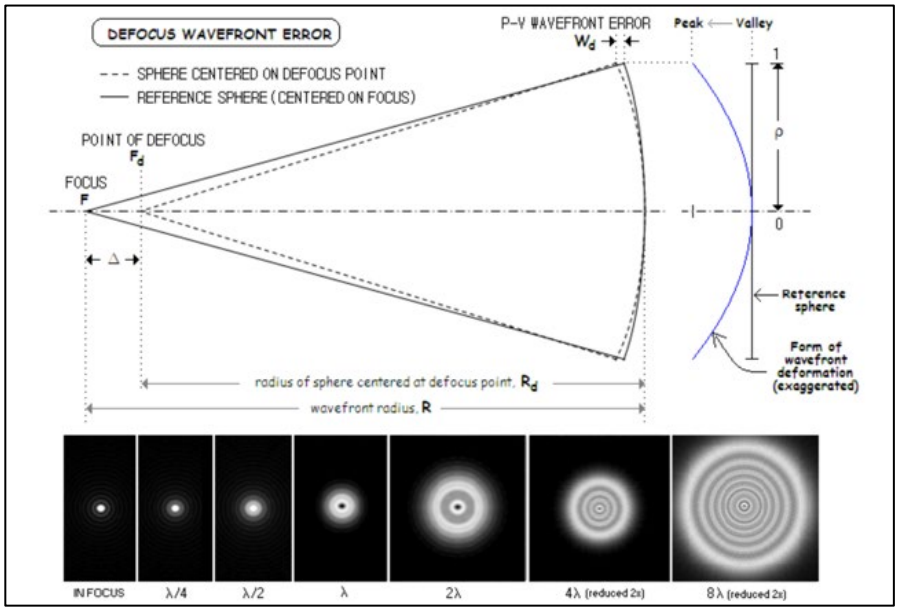

Figure 4. Defocus Error vis-à-vis Wavefront Error.
Source: [11].

2. Spherical Aberration

The spherical aberration occurs when multiple wave front errors result in multiple image points being formed around the ideal gaussian image point. This can occur when light refract at varying degrees along the lens in a way that leads to multiple converging points around the gaussian point, as seen in Figure 5. The result is multiple image intensity Sinc functions being formed with diminished central maxima intensity due to the energy spread across focal points, causing a blurring more severe than defocus.

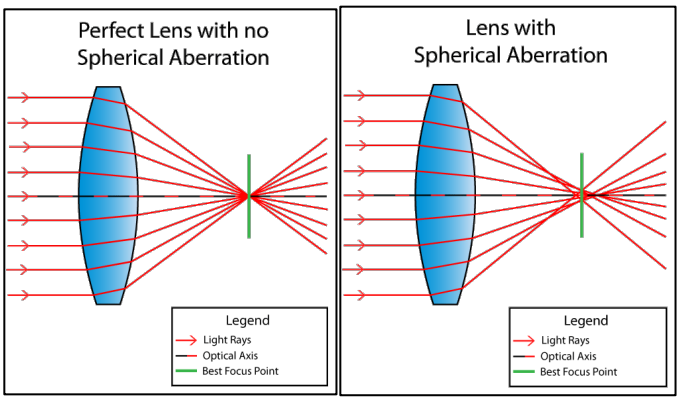

Figure 5. Spherical Aberration.
Source: [12].

C. MODELING OF OPTICAL ABERRATIONS WITH ZERNIKE POLYNOMIAL

Mathematically, the light wavefront errors (W) of the above-mentioned defocus and spherical aberrations can be modelled with the ZERNIKE polynomial, which has orthogonal functions that can effectively describe the light wavefront in a circular aperture. The ZERNIKE polynomial for defocus and spherical aberrations can also be expressed in the cartesian coordinate for efficient computation, as visualized in Figure 6.

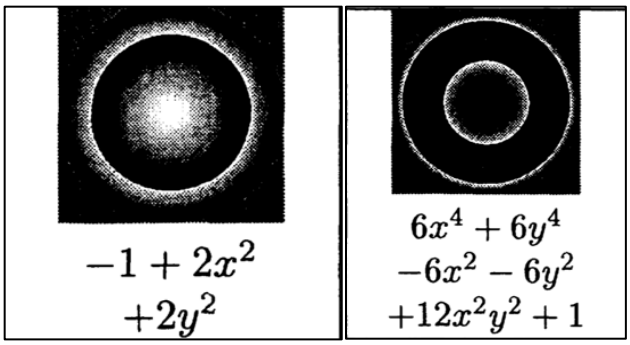

Figure 6. Zernike Polynomial for Defocus and Spherical Aberrations.
Source: [13].

However, to accurately describe the gaussian (normal) distribution of light waves in a blur circle, also known as the point spread function (PSF), h, a Fourier transform F is applied to the wavefront error W.

$$h = |F[W]|^2.$$

To generate a blur image, g, from a clean image, f, convolve the PSF, h, with f.

$$g = f \times h.$$

D. COMPUTER VISION DEBLURRING WITH DEEP LEARNING

Traditionally, computer vision tasks have been undertaken by so-called expert systems, which models images based on physical laws, a highly specialized and application-specific enterprise that has difficulty in scaling for complex applications such as self-driving cars. For deblurring, one key problem is that it is an ill-constructed problem as the blurred kernel PSF is often unknown. As such, one typical method for deblurring is with the blind deconvolution algorithm, which requires an initial input of estimated noise or blurred kernel PSF and then further iterated to obtain the ideal result, which can be time consuming and presents several limitations [14].

In recent years, ML, which models images based on a data-driven approach, has been gaining traction due to advancements in computing power, storage capacity, and algorithms. ML is a subset of AI which allows a computer software to derive a model from training data (see Figure 7). The model, having undergone *model training* with data, can perform various computer tasks such as object classification, semantic segmentation, deblurring and super-resolution. For example, a blurred image can be fed into the *trained model* as an input data and be rapidly output as a deblurred image, also known as *model inference* (see Figure 7), thus having the advantage of speed over traditional methods such as the blind deconvolution algorithm. This advantage of ML makes it ideal for near real-time applications that makes it suitable for military and civil operations which require time-critical responses such as in military strike and humanitarian rescue missions.

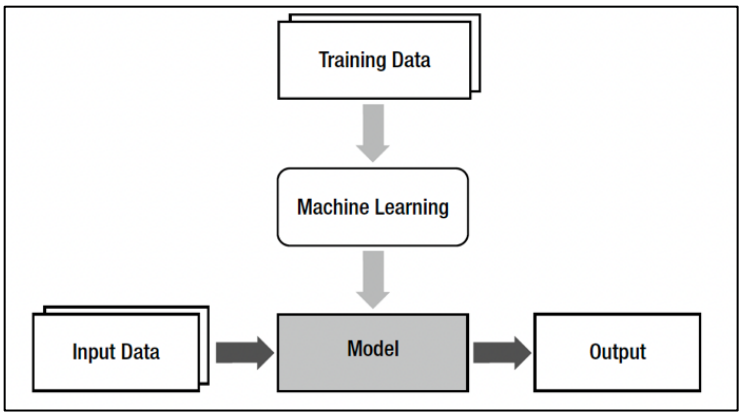

Figure 7. ML Process and Model Inference.
Source: [1].

However, the implicit challenge in ML is the training data, where an issue known as *over-fitting* can arise if the model can only perform well for a narrow set of new data that closely resemble the training data. In contrast, a good model has a quality known as *generalization*, where the model can perform well for a broad set of new data with similar characteristics as the training data.

Deep learning is the state-of-the-art technology of ML which utilizes deep neural networks that have more than one hidden layer (see Figure 8).

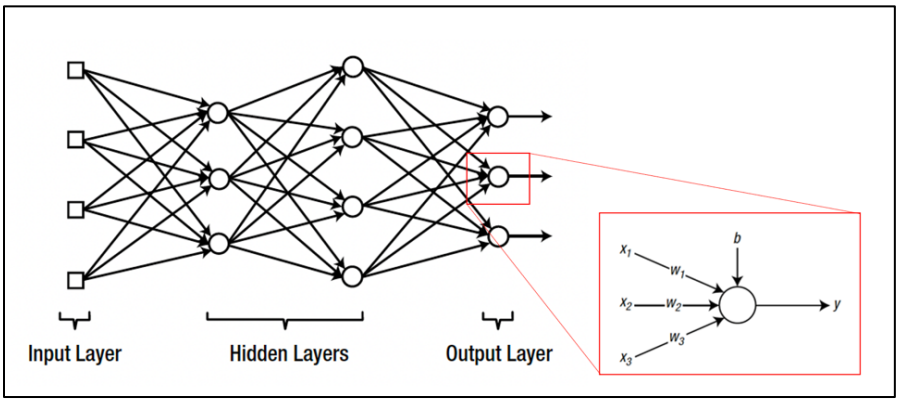

Figure 8. Deep Neural Network.
Source: [1].

A deep neural network comprises neural nodes that are interconnected, with each connection represented by a different weight (W). Preceding nodes are taken in as inputs (X) for each node and multiplied by the respective weight values, and with the addition of a bias (b), produces the weighted sum of the inputs v [1].

$$v = (X1 \times W1) + (X2 \times W2) + (X3 \times W3) + b.$$

To obtain the final output (Y) of a node, an activation function (φ) is applied to the weighted sum of inputs (v) [1].

$$Y = \varphi(v) = \varphi(W \times X + b).$$

To maximize the performance of ML, this thesis adopts a ML type known as *supervised ML* (see Figure 9), where the training data comprises blurred images (input) as well as the pristine/original sharp images (correct output).

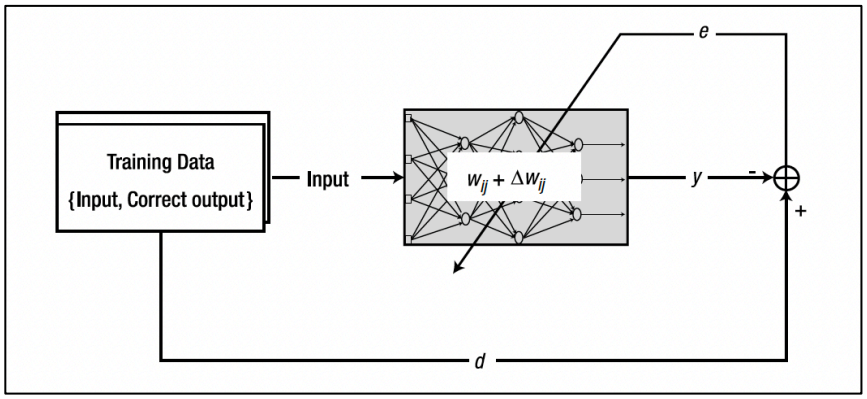

Figure 9. Supervised ML.
Source: [1].

The output (Y) obtained from the input is compared with the correct output (d) to obtain an error (e), which is then back propagated through the neural network to adjust the weights of the neural network via a *learning rule*. This process is typically repeated until the error reaches an accepted tolerance. Therefore, it can be understood that a *trained model* is essentially a deep neural network with adjusted weight values.

For computer vision, the deep learning architecture used is known as CNN. The main architectural design of a CNN is the use of an image feature extraction network which produces various feature maps via convolution (image filter function) and pooling (data reduction function) operations, as opposed to the traditional method of having expert designed feature extractors (see Figure 10) [1].

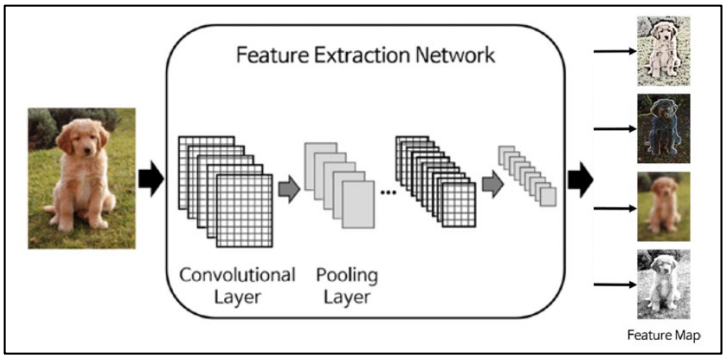

Figure 10. Architecture of Convolutional Neural Network. Source: [1].

E. REVIEW OF UNET

UNET is essentially a type of CNN with an encoder-decoder architecture (see Figure 11), first developed in the field of bio-medical imaging for semantic segmentation. UNET is used to better segment out the cell of interest from the surrounding cells, which can be challenging due to low contrast and clustering of cells [3].

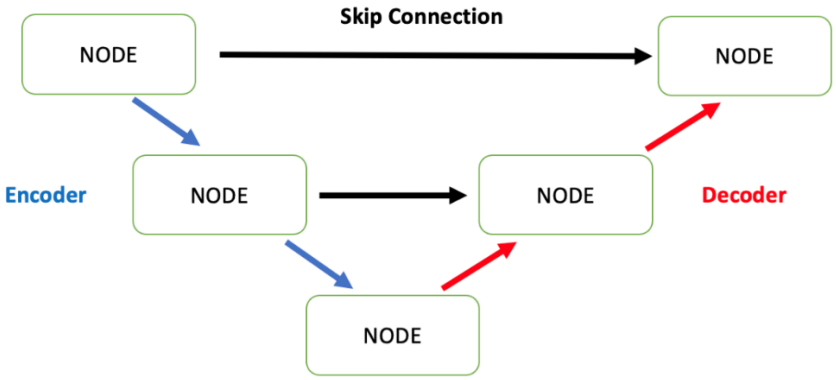

Figure 11. General Architecture of UNET

The key innovation of the original UNET lies in the skip connections between the encoder-decoder nodes at the same stages, where the higher resolution features from the encoder nodes are combined with the better pixel-localized features from the decoder nodes, resulting in more precise image segmentation [3]. Further innovations in the scale of skip connections, in particular the supervision of the first stage encoder node on other nodes, further advanced the UNET architecture, namely with UNET++ [7] and UNET3+ [8] (see Figure 12 for summary), which are also applied in the bio-medical imaging fields.

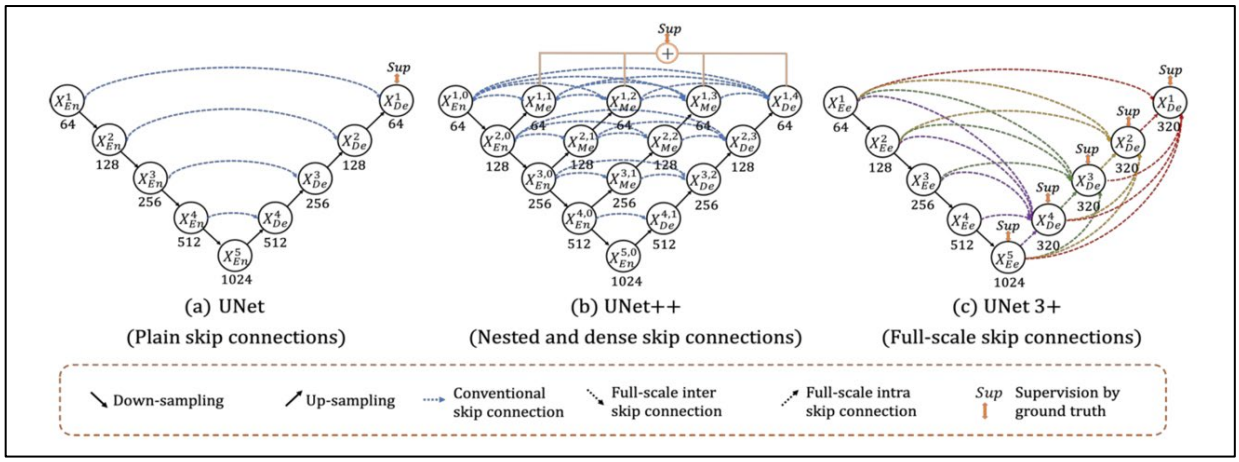

Figure 12. Summary of UNET and Two Main UNET Variants.
Source: [8].

In the UNET3+ research, UNET3+ not only outperformed UNET++ in the selected datasets (liver and spleen) based on an IQM (i.e., dice similarity coefficient), but also had the advantage of being faster due to it having lesser parameters [8].

Due to its effectiveness and simplicity, UNET variants have also been applied successfully in applications such as deblurring of satellite images affected by atmospheric turbulence, notably in [4], where adding additional convolutional layers (also known as convolutional autoencoder [CAE]) to UNET at both the encoder beginning and decoder end yielded better performance (see Figure 13).

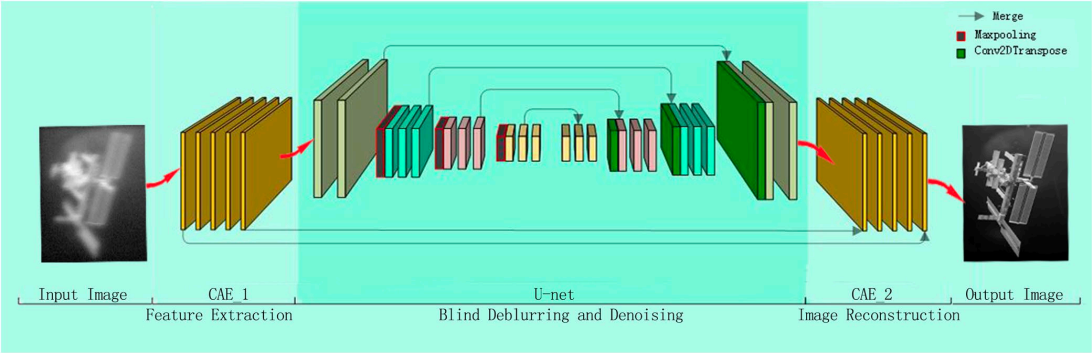

Figure 13. UNET with Deep Convolutional Autoencoder Architecture. Source: [4].

THIS PAGE INTENTIONALLY LEFT BLANK

III. STUDY APPROACH AND SETUP

This chapter describes the study approach and setup, which includes (1) the generation of clear images, (2) modeling and simulation of blurred images, (3) implementation of UNET and UNET variants, and (4) performance evaluation methods.

A. EQUIPMENT SETUP

This study utilizes the MATLAB software for data generation and evaluation, the setup of the UNET CNNs, and deep learning implementation (*model training* and *model inference*). To enable the *model training* of the large dataset as well as the UNET CNNs, the hardware utilized is the NVIDIA DGX station which has four Tesla V100 GPUs.

B. CLEAR DATA GENERATION

The dataset selected for this research is the open-source XVIEW dataset, which comprises satellite images taken by the Worldview-3 Satellite, at a GSD resolution of 1.24m. There is a total of 1,129 RGB colored images, at varying dimensions of ~3000 by ~2000 pixels.

1. Clear Data Generation

It is necessary to have a large number of data to generate better results for deep learning. In addition, the pixel dimensions must be uniform and small to allow for efficient training with the equipment setup. As such, the XVIEW images are first cropped to typical dimensions of 512 by 512 pixels to retain sufficient scene information, generating a total of 35,701 images from the original 1,129 images. These images are then resized to dimensions of 256 by 256 pixels for optimal computation, at the expense of losing some pixel information.

2. Dataset Ratio

For optimal deep learning implementation, the dataset, which comprises clear and blurred image pairs, is split into three separate segments, each with distinct images:

(1) training, (2) testing, and (3) validation. This ensures that the trained model is tested and validated on images it has not seen before. The training and validation datasets are used during the training process and the testing dataset is used with the trained model for model performance evaluation. For this study, the percentage ratio used for training, testing, and validation are ~75%, ~12.5%, and ~12.5%, respectively (see Table 1).

Table 1. Dataset Ratio

Training	Testing	Validation
26,542 images (~75%)	4,580 images (~12.5%)	4,579 images (~12.5%)

C. BLURRED DATA GENERATION

For this study, the blurred images are modelled with Zernike polynomials as detailed in Chapter II and simulated with MATLAB. Table 2 summarizes the defocus and spherical aberrations that are generated with various factor types and values. A fixed factor would mean that the entire dataset is uniformly blurred with the same factor value whereas a mixed factor type would mean that the dataset is randomly blurred with varying values. The higher the factor value, the greater the optical aberration.

Table 2. Blurred Dataset Generation Summary

Aberration Type	Factor Type and Values
Defocus	Fixed (value of 1)
Defocus	Random (values between 0 to 1.1)
Spherical	Fixed (value of 1)
Mix Defocus and Spherical	Fixed (value of 1)

D. UNET SETUP AND TRAINING PARAMETERS

Table 3 summarizes the UNET architectures that are tested in this study. For this study, the general architectures of the UNET and various UNET variants are adopted with the key training parameters as detailed in Table 4. Notably, Deep-UNET3+ is created for this thesis study by combining UNET3+ with additional convolutional layers as in the deep UNET CAE. These UNET CNN variants are designed and implemented in MATLAB's deep learning designer application. Note that as this is a deblurring problem, the output layer would be a regression layer.

Table 3. Summary of UNET Architectures

UNET Architectures
UNET
UNET++
UNET3+
Deep-UNET3+

Table 4. Summary of Key UNET Training Parameters

Parameter Type	Parameter Detail
UNET Stages	Two Stages
Optimizer	ADAM @ 0.001
Max Epoch	10
Mini-Batch Size	16
Padding	Yes
Dropout Layer	Yes
Loss Function	Mean Squared Error

E. IMAGE QUALITY METRICS

The deblurred images are obtained by inferencing the testing dataset with the trained UNET models. The deblurring performance are mainly evaluated quantitatively with three main image quality metrics (IQMs) (see Table 5) commonly used in this area of research [15], as well as qualitatively with human vision for additional verification.

Table 5. Summary of IQMs

Image Quality Metric (IQM)
Mean Squared Error (MSE) Measurement
Peak Signal to Noise Ratio (PSNR)
Structural Similarity (SSIM) Index

1. Mean Square Error Measurement

The MSE measures the difference of the pixel values between the clear and blurred images:

$$MSE = \left(\frac{1}{N} \sum_{i=1}^{N} \left| x_i - y_i \right|^2 \right),$$

where x and y are i -th samples of the clear and blurred images respectively and N is the total number of image samples [15]. The formula penalizes greater differences between the clear and blurred images. A lower value indicates higher similarity between the clear and blurred images. This measurement is chosen as it is mathematically simple and because the regression layer's loss function is also based on MSE. However, the limitation is that the MSE measurement may not be accurate in representing human-perceived quality measurements [16].

2. Peak Signal-to-Noise Ratio

The PSNR measurement is a derivative of the MSE measurement and traditionally a highly popular image quality metric:

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right),$$

where 255 represents the maximum value of an eight-bit pixel [15]. Therefore, the formula further penalizes higher MSE. A higher value indicates better image quality. However, like the MSE measurement, it suffers from inaccuracies related to representing human perceived quality measurements.

3. Structural Similarity Index

The SSIM index is a human vision system based framework that seeks to provide a good approximation to human perceived image distortion [15]:

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right) \times \left(2\mu_{xy} + C_2\right)}{\left(\mu_x{}^2 + \mu_y{}^2 + C_1\right) \times \left(\sigma_x{}^2 + \sigma_y{}^2 + C_2\right)},$$

where $\{\mu_x, \sigma_x\}$ and $\{\mu_y, \sigma_y\}$ refer to the mean intensity and standard deviation of the clear image x and blurred image y respectively while μ_{xy} is the mean intensity of image x and blurred image y i.e., their cross correlation, and C_1 and C_2 are constants to ensure the denominator will not be too close to zero [15]. Conceptually, the SSIM is closer to human's perception of image quality by considering the entire image structure. A value closer to 1 indicates higher image similarity.

4. Slanted Edge Modulation Transfer Function

Where appropriate, depending on the availability of suitable image samples, the main IQMs in Table 5 are supplemented with the slanted edge modulation transfer function (MTF) to describe image quality with contrast and resolution. The MTF is a method that measures the ability of an imager to transfer the object modulus (contrast) to the detector at a given spatial frequency (resolution) [17]. It defines how well an imager can resolve the target contrast at a given resolution [18]. In general, an image's modulation decreases (from

modulation value of 1) with increasing spatial frequency until the image's limiting resolution is reached, the spatial frequency where the modulation value is zero. The slanted edge MTF based on the ISO standard 12233 is a commonly used technique to approximate a digital imager's MTF using the edge spread function (ESF)/edge profile (b) of a knife-edge/line-pair image (a) (see Figure 14).

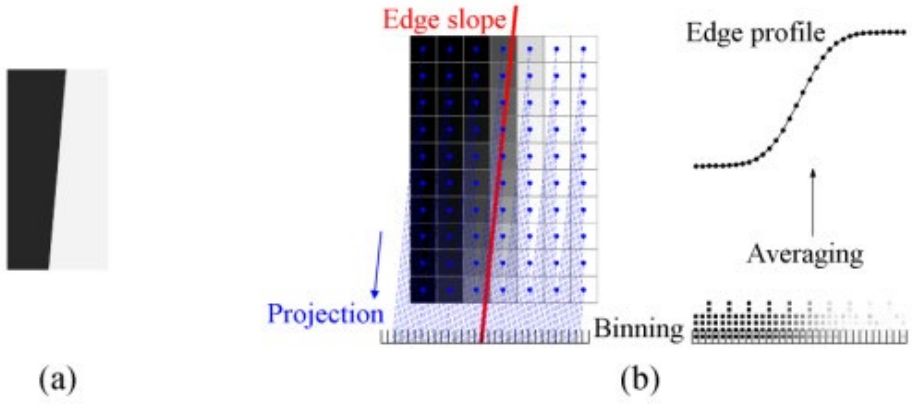

Figure 14. Edge Spread Function/Edge Profile of Knife-Edge Image.
Source: [18].

Subsequently, the line spread function (LSF) is obtained from the derivative of the ESF. Finally, the MTF is obtained by applying the fast Fourier transform (FFT) on the LSF. See Figure 15 for process to obtain MTF from ESF.

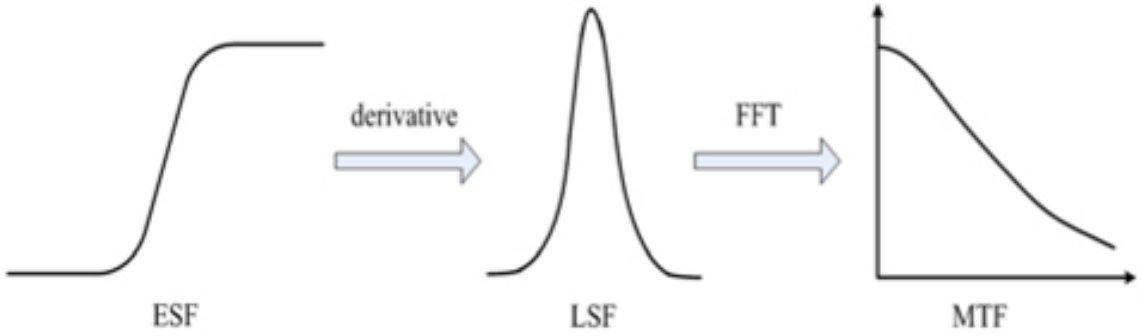

Figure 15. Process to Obtain MTF.
Source: [19].

IV. STUDY RESULTS AND ANALYSIS

This chapter describes the study results and analysis for the various blurred test cases of the dataset as described in Table 2 using the UNET architectures in Table 3. Overall, the IQMs and visual results show that the UNET architecture is effective in deblurring the optically aberrated images. In addition, deep-UNET3+ provided the best performance amongst the UNET architectures.

A. OPTICAL ABERRATION: FIXED DEFOCUS

For the training dataset using a fixed defocus factor of 1, Table 6 summarizes the deblurring performance as quantified by the various IQMs. UNET3+ and UNET provided much better performance than UNET++ across all 3 key metrics. However, while UNET3+ performed the best under the MSE measurement, UNET performed the best under the PSNR measurement. In this case, it is reasonable to give more weight to the MSE measurement given that the loss function used during training is the MSE.

Table 6. Summary of IQM for Defocus Factor of 1 using Various UNETs

UNET Variant	Training Time	MSE	PSNR	SSIM
UNET	~5 hours	1.047e+03	20.63	0.68
UNET++	~7 hours	1.489e+03	19.25	0.63
UNET3+	~10 hours	1.0393e+03	20.57	0.68

Figure 16 shows the visual results of deblurring with the various UNET architectures for the three selected images. These visual results are consistent with the IQM results. It can be observed from the first image that there is a clear improvement in resolution (runway line pairs and numerical markings are resolved) after deblurring, with UNET3+ appearing to provide the best results. Notably, UNET++ appeared to perform worse than UNET for this test case. In the second and third images, deblurring managed to

resolve object details that were lost during blurring. Specifically, vehicles absent in the blurred images were now visible in the images deblurred by UNET3+.

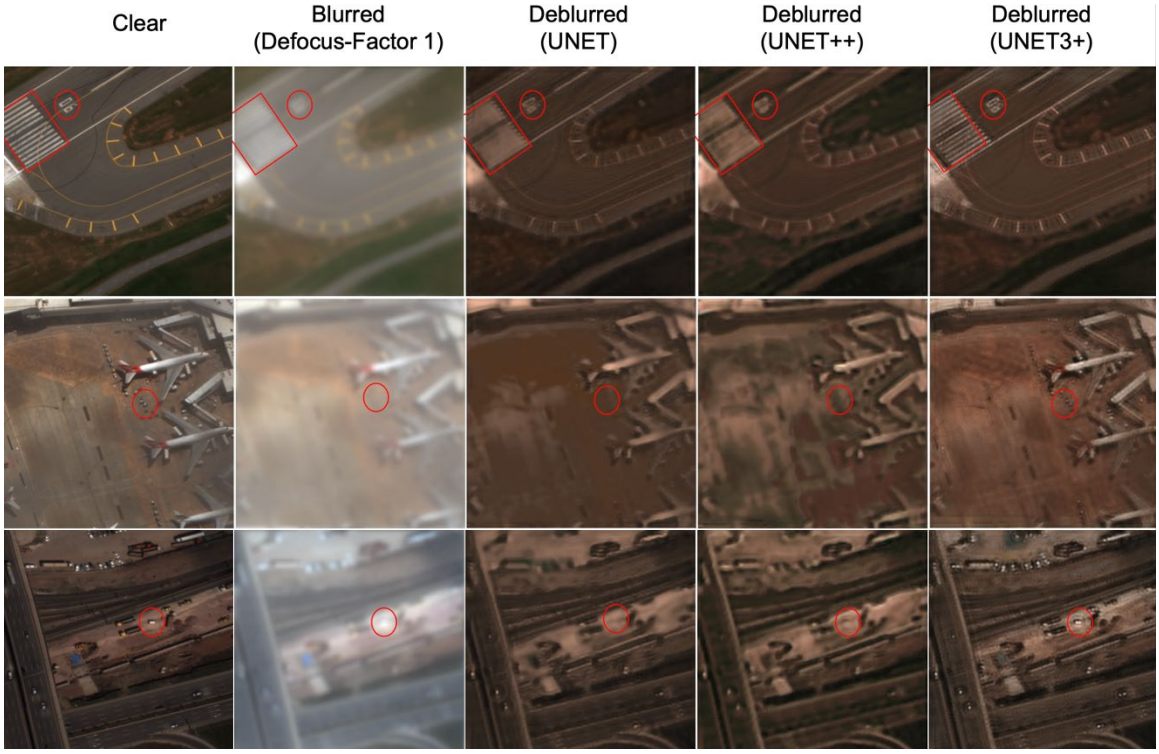

Figure 16. Visual Results for Fixed Defocus using Various UNETs

As an example to quantify the improvement of resolution as observed in Figure 16, the slanted edge MTF technique is applied on a sample knife-edge image from the XVIEW dataset (see Figure 17).

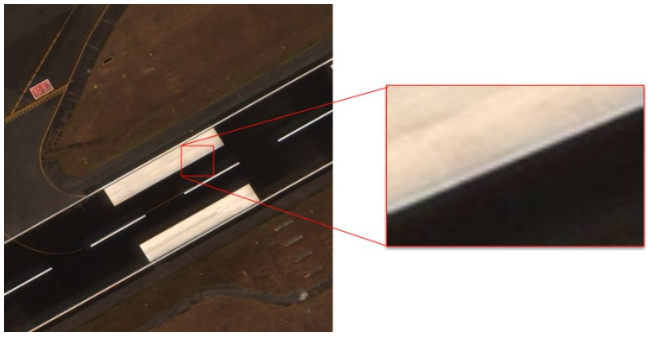

Figure 17. Image Sample for Slanted Edge MTF

Figure 18 shows the slanted edge MTF results based on Nyquist frequency (spatial frequency corresponding to 0.5 cycles per pixel) for the pristine/original image, blurred image (defocus aberration), and deblurred image (using trained UNET3+). As observed, the limiting resolution of the blurred image decreases significantly (from 0.5 cycles per pixel to 0.3 cycles per pixel) as compared to the pristine image. However, the limiting resolution of the deblurred image is restored back to 0.5 cycles per pixel. In addition, the deblurred image also has slightly higher contrast values at the various spatial frequencies when compared to the blurred image, indicating higher image quality. As such, the slanted edge MTF results are consistent with the visual results in Figure 17.

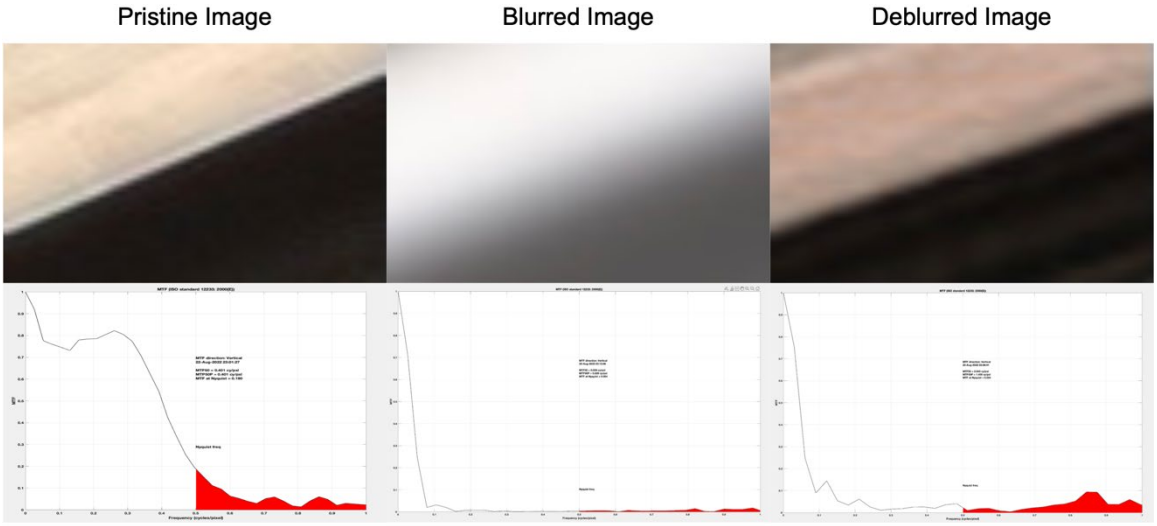

Figure 18. Slanted Edge MTF Results for Defocus Aberration

B. OPTICAL ABERRATION: RANDOM DEFOCUS FACTOR

For the training dataset using random defocus factors ranging from 0 to 1.1, only the best performing UNET3+ CNN based on results from Section A. is used for training. The testing dataset comprises two blurred datasets using fixed and random defocus factors. Table 7 summarizes the results for the training dataset of random defocus factors on different testing datasets, as well as provides a comparison versus the training dataset of fixed defocus factor. Notably, the UNET3+ trained on the mixed defocus factor also

performed better on the testing dataset that had the mixed defocus factor, suggesting that a CNN trained on a generalized dataset would also perform better on a generalized testing dataset.

Table 7. Comparison of IQM for Fixed vs. Random Defocus using UNET3+

Training Data	Testing Data	MSE	PSNR	SSIM
Fixed Factor	Fixed Factor	1.04e+03	20.57	0.684
Random Factor	Fixed Factor	1.57e+03	19.40	0.643
	Random Factor	711	21.98	0.688

As observed in Figure 19, when tested on a defocus with a fixed factor, while there is an improvement from the deblurring, the CNN trained on the mixed defocus factor did not perform as well when compared to the CNN trained specifically on the test data, as expected due to the differences in the weight adjustments during *model training*.

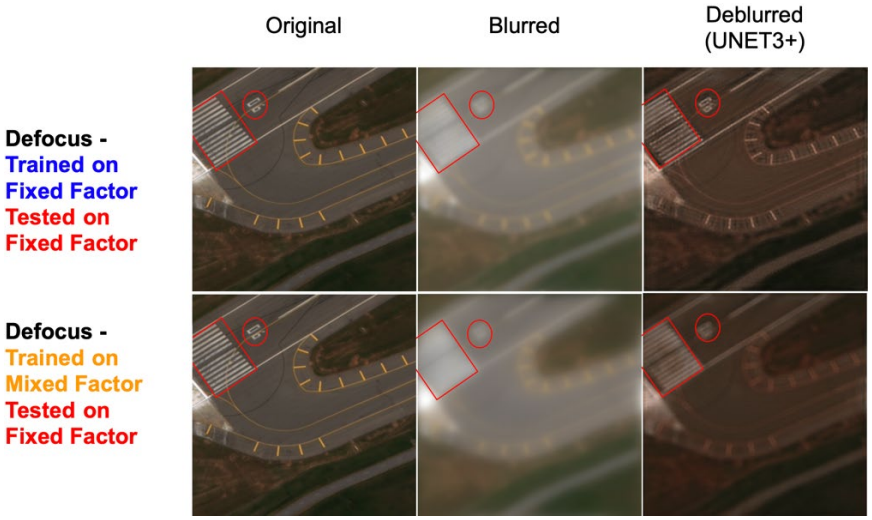

Figure 19. Visual Comparison for Random Defocus using UNET3+

C. OPTICAL ABERRATION: FIXED SPHERICAL

For the training dataset using the fixed spherical factor of 1, other than the best-performing UNET3+, Deep-UNET3+ is also trained for better deconstruction and construction of the image. Table 8 summarizes the deblurring performance as quantified by the various IQMs. Notably, Deep-UNET3+ slightly out-performed UNET3+.

Table 8. Comparison of IQM for Spherical Factor of 1 using UNET3+ and Deep-UNET3+ with Defocus Factor of 1 using UNET3+

Aberration	UNET Variant	Training Time	MSE	PSNR	SSIM
Defocus	UNET3+	~10 hours	1.0393e+03	20.57	0.68
Spherical	UNET3+	~10 hours	919	20.94	0.677
	Deep-UNET3+	~12 hours	711	21.98	0.688

Figure 20 shows the visual results of deblurring with UNET3+ and Deep-UNET3+. Consistent with the IQM results, Deep-UNET3+ provided better visual performance than UNET3+. However, even though the IQM results of the UNET3+ CNN trained using the spherical aberration are better than for the UNET3+ CNN trained using defocus aberration, the visual performance of the UNET3+ CNN trained using the spherical defocus (which is a more complex optical aberration than defocus), as observed from the runway line pair and numerical marking, is worse. This suggest that the IQMs alone may be inaccurate in determining image quality.

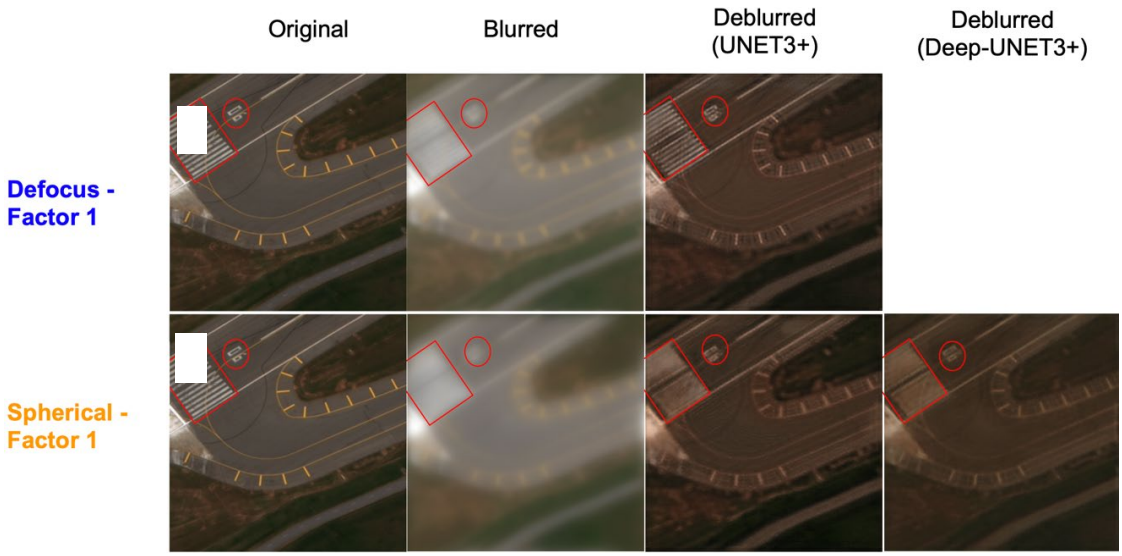

Figure 20. Visual Comparison of Fixed Defocus vs. Spherical with UNET3+ and Deep-UNET3+

D. OPTICAL ABERRATION: MIX OF DEFOCUS AND SPHERICAL

For the training dataset using a mix of defocus and spherical optical aberrations of factor 1, Deep-UNET3+ is used because it provided the best performance based on the results in Section C. Table 6 summarizes the deblurring performance as quantified by the various IQMs using various testing datasets. The IQM results are relatively good across the testing dataset using (1) fixed defocus factor of 1 only, (2) fixed spherical factor of 1 only, and (3) a mixture of defocus and spherical factor of 1.

Table 9. Summary of IQM for Mixture of Defocus and Spherical with Deep-UNET3+

Training Data	Testing Data	MSE	PSNR	SSIM
Defocus and Spherical	Defocus	898	21.20	0.656
	Spherical	944	21.10	0.655
	Mixed	919	21.15	0.656

Figures 21 and 22 show the visual performance of the test cases as summarized in Table 9. In addition, even though the Deep-UNET3+ CNN is trained on the dataset using mixed defocus and spherical aberrations, it appears that its visual performance is comparable to the Deep-UNET3+ trained on the dedicated training dataset using spherical aberrations. This suggest that there is potential for a UNET trained on a broad-based optical aberration to be effective.

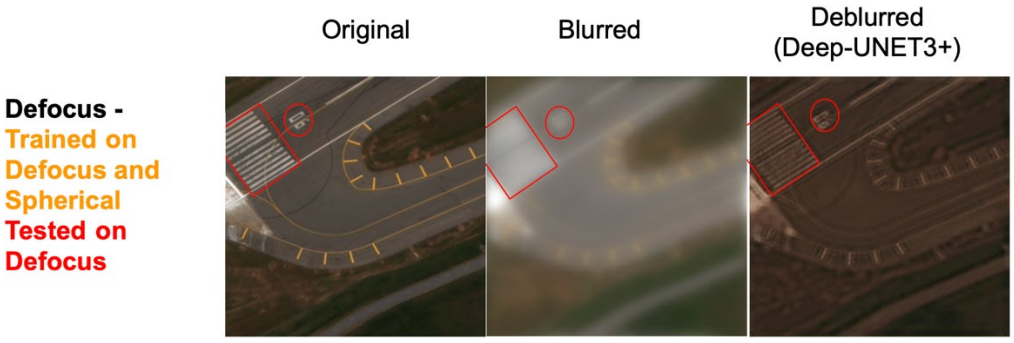

Figure 21. Visual of Mixed Training Dataset using Deep-UNET3+

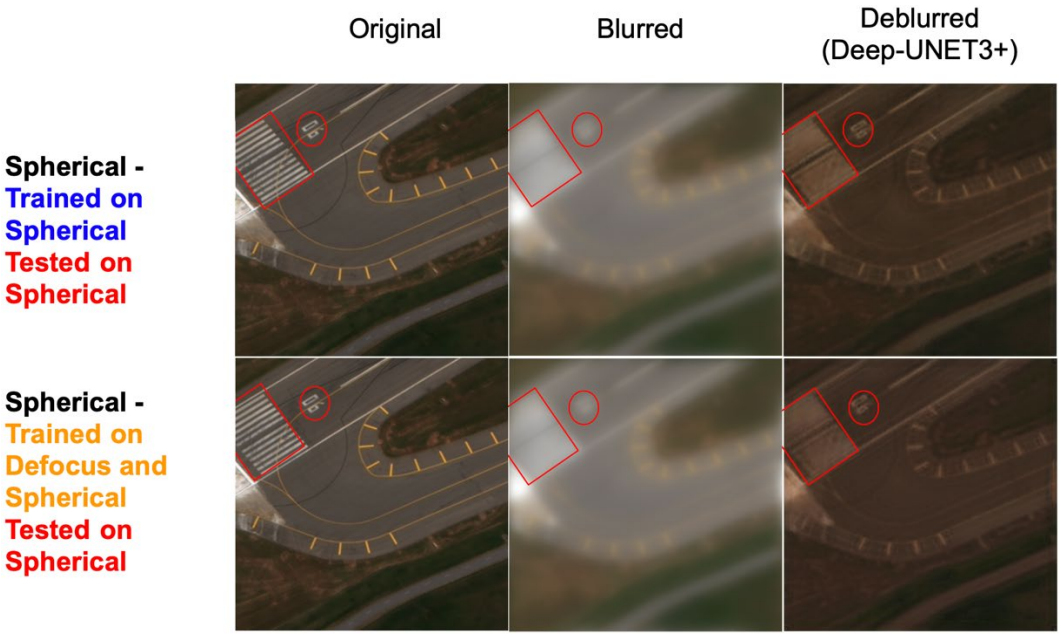

Figure 22. Visual Comparison of Spherical vs. Mixed Training Dataset using Deep-UNET3+

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSION

This chapter provides key insights from this thesis, highlights its limitations, and recommends future work. In summary, this thesis has two main contributions: (1) demonstrating that UNET can be used for deblurring optically aberrated satellite imagery, and (2) developing Deep-UNET3+, which combines UNET3+ and deep UNET CAE, and demonstrating that it achieved the best deblurring performance.

A. UNET IS USEFUL FOR DEBLURRING OPTICALLY ABERRATED SATELLITE IMAGERY

As shown by the results in Chapter IV, UNET can deblur satellite images with the defocus and spherical aberrations, demonstrating improved spatial resolution and resolving small objects that are not visible when blurred. In addition, UNET can provide relatively good performance even when the training dataset uses mixed optical aberrations of defocus and spherical, as well as varying factors of aberrations such as for the defocus optical aberration. This outcome means that when the cause of blur is unknown, a UNET trained on a sufficiently large and representative dataset comprising ~~of~~ multiple optical aberrations and varying factors could be useful as a preliminary method.

B. DEEP-UNET3+ PROVIDES THE BEST DEBLURRING PERFORMANCE

UNET and two variants (UNET++ and UNET3+) are tested in this thesis, and amongst these, UNET3+ provides the best performance which is consistent with the results when they are applied on bio-medical images for semantic segmentation [8]. In addition, UNET3+ is modified with additional convolutional layers as per deep UNET CAE to obtain Deep-UNET3+, which outperformed UNET3+ when tested. This result validates the concept that additional convolutional layers enhance the image deconstruction prior to the UNET encoder and image construction after the UNET decoder [4].

C. POTENTIAL UNET APPLICATIONS

Deblurring of satellite images with UNET can be used to sustain existing satellite capabilities that have degraded over time, since we will have images pre and post degradation, which can be used for the training of UNET. In addition, the improvement of spatial resolution by UNET indicates that UNET can be potentially used for super-resolution applications as well, where images from a higher-resolution satellite can be used to enhance the image from a lower-resolution satellite.

D. KEY LIMITATIONS AND SOURCES OF ERROR

One key limitation of this thesis arises from the equipment setup and therefore the computational resources available, which limit the image dimensions and depth of the UNET CNN employed. As such, errors can result from information lost during image pre-processing with the resizing of the image and less information being extracted from the image with a shallower UNET CNN.

Another limitation is in the use of IQMs. As seen in Chapter IV, different IQMs produce different results and may not all be consistent in indicating superior performance of one UNET variant over another. As such, MTF and visual results are needed for additional comparison and validation. The effectiveness of IQM is also tied to the loss function, which explains why MSE seemed to be the best indicator of performance.

E. RECOMMENDATIONS FOR THE WAY FORWARD

There are several key areas that can be explored to further the research and performance of deblurring satellite imagery with UNET. One, other optical aberrations with higher factors of aberrations can be tested to expand the envelope of UNET's application; two, improving deblurring performance with deeper UNET layers; three, testing different hyper-parameters such as loss functions, e.g., focal loss; four, combining UNET with other deep learning algorithms such as Generative Adversarial Networks (GAN) to improve the robustness of the training process; and five, use of other evaluation

metric such as the slanted edge MTF and use of object detection neural networks to validate deblurring performance. In addition, research can be conducted with UNET on actual operational datasets and on super-resolution problems as well as with infrared and radar images.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] Kim, P. *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence.* Apress Berkeley, California, 2017, pp 1–155.

[2] Wu, X., Li, J., Zhou, G., Lü, B., Li, Q., and Yang, H. "RRG-GAN Restoring Network for Simple Lens Imaging System." *Sensors (Basel, Switzerland)*, Vol. 21, No. 10, 2021, p. 3317. https://doi.org/10.3390/s21103317.

[3] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, No. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., Munich, Germany, 2015, pp. 234–241.

[4] Chen, G., Gao, Z., Wang, Q., and Luo, Q. "U-Net like Deep Autoencoders for Deblurring Atmospheric Turbulence." *Journal of Electronic Imaging*, Vol. 28, No. 5, 2019, p. 053024. https://doi.org/10.1117/1.JEI.28.5.053024.

[5] S. -J. Cho, S. -W. Ji, J. -P. Hong, S. -W. Jung, and S. -J. Ko. Rethinking Coarse-to-Fine Approach in Single Image Deblurring. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, No. 1, 2021, pp. 4621–4630.

[6] Mamidibathula, B., and Biswas, P. K. SVDocNet: Spatially Variant U-Net for Blind Document Deblurring. 2019.

[7] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation." *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S*, Vol. 11045, 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.

[8] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., and Wu, J. "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation." *Arxiv*, 2020. https://doi.org/10.48550/arXiv.2004.08790.

[9] Olsen, R. C. *Remote Sensing from Air and Space, 2nd ed.* SPIE, Bellingham, Washington, USA, 2016, pp. 65–69.

[10] Massachusetts Institute of Technology. "MIT Deep Learning 6.S191." Accessed May 14, 2022. [online] http://introtodeeplearning.com.

[11] “Defocus Aberration.” Accessed May 14, 2022. [online] https://www.telescope-optics.net/defocus1.htm.

[12] Photography Life. “What Is Spherical Aberration?” Accessed May 14, 2022. [online] https://photographylife.com/what-is-spherical-aberration.

[13] Brug, H. H. van. Efficient Cartesian Representation of Zernike Polynomials in Computer Memory. In *Fifth International Topical Meeting on Education and Training in Optics*, No. 3190, 1997, pp. 382–392.

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and Evaluating Blind Deconvolution Algorithms. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[15] K. -H. Thung and P. Raveendran. A Survey of Image Quality Measures. Presented at the 2009 International Conference for Technical Postgraduates (TECHPOS), 2009.

[16] Mathworks. “Image Quality Metrics—MATLAB & Simulink.” Accessed Aug. 1, 2022. [online] https://www.mathworks.com/help/images/image-quality-metrics.html.

[17] Ahmed, S. N. 7 - Position-Sensitive Detection and Imaging. In *Physics and Engineering of Radiation Detection (Second Edition)* (S. N. Ahmed, ed.), Elsevier, 2015, pp. 435–475.

[18] Estribeau, M., and Magnan, P. Fast MTF Measurement of CMOS Imagers Using ISO 12333 Slanted-Edge Methodology. In *Detectors and Associated Signal Processing*, No. 5251, 2004, pp. 243–252.

[19] Hang, L., Changxiang, Y., and Jianbing, S. “Measurement of the Modulation Transfer Function of Infrared Imaging System by Modified Slant Edge Method.” *Current Optics and Photonics*, Vol. 20, No. 3, 2016, pp. 381–388. https://doi.org/10.3807/JOSK.2016.20.3.381.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California