# A Comprehensive Study of Conditional Generative Adversarial Networks for Noise Reduction in Optical Coherence Tomography

by

Youwei Chen

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences - Radiology and Diagnostic Imaging

University of Alberta

# Abstract

Optical coherence tomography (OCT) has been widely adopted as an imaging modality for various clinical applications, such as breast cancer screening, retinal imaging, and vascular assessment, due to its non-invasive nature. However, OCT is affected by coherent speckle noise, which impairs OCT images' contrast and detailed structural information. This presents a challenge for accurate clinical analysis. To improve image quality, one can adopt frame-wise averaging of OCT images, where multiple images of the same field of view are consecutively acquired and averaged. However, this approach is time-consuming and impractical for point-of-care clinical applications. To overcome the time-consuming shortcomings and to make OCT more suitable for surgical settings, we propose the application of conditional generative adversarial networks (cGAN). The cGAN was trained to learn how the signal and noise characteristics change via the averaging process, using non-clinical data for training and then testing on unseen clinical breast tissues. This method has demonstrated strong robustness and generalizability, significantly enhancing signal and contrast without compromising sharpness and reducing speckle noise in OCT B-scans of human breast tissue. The proposed method offers a potential replacement for frame-wise averaging approaches.

# Preface

This thesis was submitted as a partial fulfillment of the Master of Science (MSc) degree in Radiology and Diagnostic Imaging at the University of Alberta. The thesis is an original work by Youwei Chen and was completed between September 2022 and May 2024. The material presented in this thesis is based on the following paper: Youwei Chen, Mark Nguyen, Yanir Levy, Michelle Noga, Ersin Bayram, Kumaradevan Punithakumar. A Comprehensive Study of Conditional Generative Adversarial Networks for Noise Reduction in Optical Coherence Tomography, Computer Methods and Programs in Biomedicine, submitted.

# Acknowledgements

I am thankful to Dr. Ersin Bayram for giving me the opportunity to engage in this project and to Yanir Levy and Dr. Mark Nguyen for their insightful contributions and unwavering support. I learned a lot from the AI and Data Science team and am very thankful to be a part of it.

I also want to express my deepest gratitude to my supervisors, Dr. Michelle Noga and Dr. Kumaradevan Punithakumar. They guided this project in the right direction, and I could not have succeeded without their consistent help and support.

Finally, I want to thank my family and friends for their continued encouragement throughout my pursuit of my goals and dreams.

Thank you to everyone who has been part of this journey.

# Contents

# List of Tables

# List of Figures

# Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| cGAN | Conditional Generative Adversarial Networks |
| CNN | Convolutional Neural Network |
| CNR | Contrast-to-Noise Ratio |
| CT | Computed Tomography |
| DCIS | Ductal Carcinoma In Situ |
| DCNN | Deep Convolutional Neural Networks |
| DCGAN | Deep convolutional generative adversarial networks |
| FDA | Food and Drug Administration |
| FID | Fréchet inception distance |
| GAN | Generative Adversarial Networks |
| IDC | Invasive Ductal Carcinoma |
| LSGAN | Least Squares Generative Adversarial Networks |
| OCT | Optical Coherence Tomography |
| PSNR | Peak Signal-to-Noise Ratio |
| ResNet | Residual Network |
| ROI | Region of interest |
| SAR | Synthetic aperture radar |
| SNR | signal-to-noise Ratio |
| SRGAN | Super-Resolution Generative Adversarial Network |
| SSIM | Structural Similarity Index |
| U-Net 128 | U-Net for 128×128 input images |
| U-Net 256 | U-Net for 256×256 input images |
| VGG | Visual Geometry Group |
| WF-OCT | Wide-Field Optical Coherence Tomography |

# Chapter 1

# Introduction

## 1.1 Overview

As the second leading cause of cancer death among women in the United States, breast cancer represents a significant portion of incidents[SMWJ23]. Projections by the American Cancer Society estimate 2,001,140 new cancer cases and 611,720 cancer-related deaths across all sexes in the United States for 2024, with breast cancer responsible for 313,510 new cases and 42,780 deaths, predominantly affecting the female population. The incidence rates of breast cancer increased by 0.6–1 percent annually from 2015 to 2019 [SGJ24]. Numerous risk factors for breast cancer have been identified, including family history, genetic mutations, personal habits, environmental factors, age, and particularly female sex, as the disease is most prevalent among women. Figure 1.1 shows the increasing trend for breast cancer diagnoses in women. Despite advancements in technology and increased awareness, women remain particularly vulnerable to breast cancer. This vulnerability largely stems from the fact that women's breast cells are highly sensitive to hormones, especially estrogen, and progesterone, in contrast to men, who have negligible levels of estrogen [ŁCF+21]. In addition, the breasts serve as accessory organs of the female reproductive system, containing mammary glands to produce milk for feeding babies [AM22]. Beyond breasts' biological functions, breasts also possess social significance and cosmetic value. Given their importance, breast conservation therapy becomes crucial when a tumor can be removed with clear margins, ensuring an acceptable cosmetic outcome [JO19], [HMLMM14]. Consequently, there is currently a strong demand for early diagnosis and precise treatment of breast cancer, particularly related to surgical care.

In this thesis, the primary focus is on producing clear images of breast tissues, leveraging significant advancements in breast cancer diagnosis and treatment. In particular, we explore recent advances in optical coherence tomography (OCT), an intraoperative imaging technique that offers high-resolution, real-time microscopic images up to 2 mm beneath the tissue surface. In breast cancer surgery, margins are defined in the edge or border of the tissue removed [HMLMM14]. The OCT technique plays a crucial role in the evaluation of surgical margins of breast tumors. However, it suffers from interference noise generated during the scanning process, which degrades image quality and increases assessment difficulties for surgeons. Therefore, supporting surgeons by providing clear OCT image scans is vital. Doing so improves the outcomes of breast conservation therapy by enabling the detection of clear margins free of cancer cells at the edge of the tissue while preserving as much healthy breast tissue as possible.

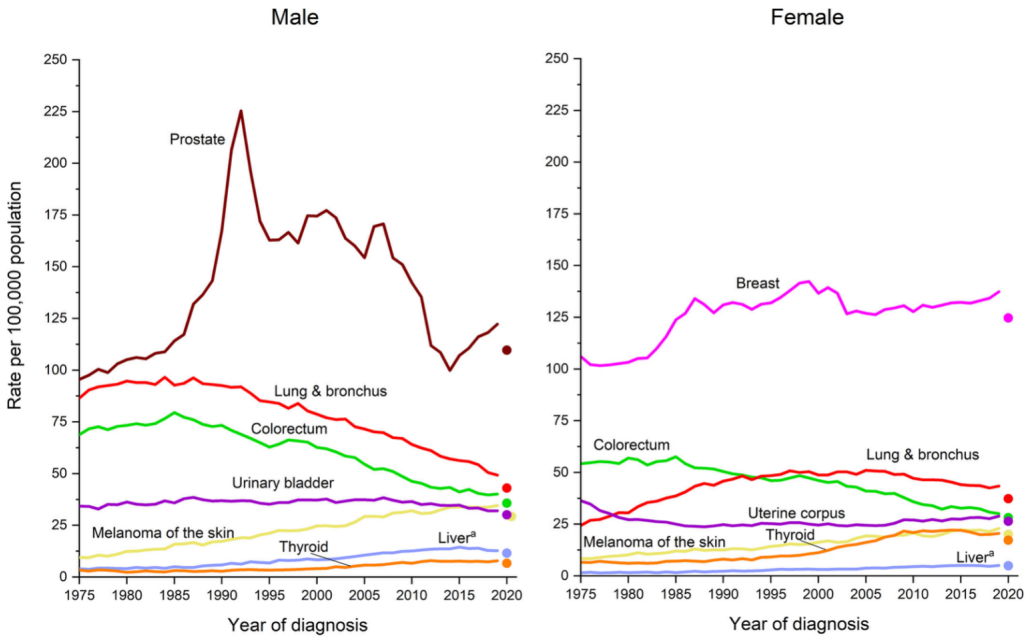


Figure 1.1: The trends in incidence rates cancers by sex, United States, 1975–2020. Rates are age-adjusted to the 2000 US standard population and adjusted for delays in reporting. Incidence data for 2020 are shown separately from the trend lines.
Source [SGJ24]

## 1.2 Current situations, challenges and objectives

Excluding non-melanoma skin cancer, breast cancer is the most commonly diagnosed cancer in 109 countries, including Canada [AM22]. Decades of oncology research, alongside advancements in breast cancer screening and technology, have enabled the detection of smaller, more numerous lesions at an early stage. This progress has significantly increased the number of breast-conserving surgeries [BW12], [HJKSN17], [Mas12]. The aim of breast-conserving surgery, or lumpectomy, is the complete removal of malignant tissue while preserving cosmetic appearance in a single procedure [KFH21]. Landmark trials have established lumpectomy followed by radiation as the standard of care for many patients [KFH21]. However, re-excision due to positive margins from local recurrence remains a significant concern after breast-conserving surgeries [KFH21]. In response, OCT has been introduced as an intraoperative imaging technique. It provides high-resolution, real-time microscopic images beneath the tissue surface, aiding in the evaluation of surgical breast tumor margins. Given that OCT imaging is an additional procedure during surgery, it is crucial to perform this task quickly to minimize its impact on surgery duration and provide near real-time guidance to surgeons. The duration of a single margin's OCT scan can range from one to two minutes, depending on the size of the excised tissue. Thus, in a worst-case scenario involving the assessment of all six margins, the scan time could extend to 6-12 minutes. Like other light-based imaging methods, OCT is also susceptible to speckle noise, and a faster scanning approach can significantly degrade the quality of OCT images. This necessitates the demand for a rapid denoising method that ensures that the quality of the denoised results is not compromised.

Recent advances in computer vision techniques, with increased availability of computational resources, have facilitated the development of image translation tasks using deep learning. Essentially, this process involves transforming an image into a target image. Inspired by these advancements and the growing accessibility of datasets from FDA-approved OCT scanning devices [RBDS+22], we aim to develop an AI-based solution to reduce or eliminate noise in OCT images through reconstruction, thereby enhancing the signal-to-noise ratio. As the device scans specimens during surgeries or when training and onboarding surgeons, AI solutions would provide a level of transparency and trust for enhanced images.

## 1.3 Thesis contributions

This thesis proposes a fully automatic deep learning-based method that denoises OCT images acquired with 1x averaging tumor images in real-time. It serves as image synthesis to significantly improve image quality without compromising important clinical features. The main contributions of this work are categorized as follows.

### 1.3.1 Application/Method Contribution

This research introduces a customized conditional generative adversarial network (cGAN) based denoising system for OCT. To tackle the complex training challenges associated with cGAN, we employ an evaluation performance method that uses the pre-trained InceptionV3 model to generate Fréchet inception distance (FID) scores. This approach not only validates our system but also captures the nuances of denoising from a unique perspective [SVI+16]. Moreover, this thesis presents a comparative analysis of various loss functions and conducts an architecture-wise evaluation, thereby illustrating the versatility of the proposed cGAN-based system for diverse applications.

Through ablation studies, we have investigated the significance of different loss functions, GAN architectures, and predefined weights, evaluating their impact on the system's performance. Our findings reveal that the proposed cGAN-based OCT denoising system significantly reduces speckle noise, demonstrating robustness and excellent generalization capabilities. Additionally, we highlight the system's ability to be trained effectively on small datasets, which is particularly advantageous in scenarios where data availability is limited. Given these strengths, our system could be used for data generation processes, augment clinical training, and lay the groundwork for further developments in artificial intelligence (AI) systems, including those focused on classification and segmentation.

### 1.3.2 Dataset perspective

The deep learning system is trained and validated using an averaging technique for raw data collected from identical specimens, and it is tested on unseen breast cancer images. The dataset, acquired from organic and inorganic specimens and phantoms, utilizes Perimeter Medical Imaging AI S-Series and equivalent systems. This dataset not only assists the development and validation of our proposed cGAN-based OCT denoising system but also

serves as a valuable resource for further research into OCT image enhancement and analysis. Using real clinical images for validation ensures that our system is evaluated under practical, real-world conditions, underscoring its potential for clinical application and its contribution to the field of medical imaging. This represents a significant step towards accelerating the application of practical OCT solutions.

The methodologies and experimental design principles highlighted in our study could drive advancements in imaging modalities afflicted with noise issues, such as ultrasound, synthetic aperture radar (SAR) imaging, and low-dose computed tomography (CT). This is particularly relevant in situations where high-signal clinical data is difficult to collect, such as using ultrasound in children or collecting data from other types of tissues to train neural networks.

## 1.4 Outline

The thesis is divided into five chapters:

- Chapter 2 reviews previous works, including OCT, clinical background, speckle noise literature, and methodological studies.
- Chapter 3 covers materials and methods.
- Chapter 4 presents experiments and results.
- Chapter 5 discusses extension studies and provides a discussion.
- Chapter 6 concludes the thesis.

# Chapter 2

# Literature Review

This section reviews denoising methods across multiple domains and modalities, highlighting the shortage of sufficient deep learning-based OCT denoising methods. It then examines the inspiration behind the ideas of image translation tasks and their extensions. Furthermore, it discusses the pre-trained model literature and the loss function literature.

## 2.1 Oncology and Optical coherence tomography

For many cancer patients, the first line of treatment is surgical removal of the tumor. When the tumor is removed, it is important to ensure that the specimen does not have cancerous cells at the margin for the patient's prognosis. Positive margins increase locoregional recurrence rates in patients with breast, colorectal, oral cavity, bladder, and potentially uterine cancer; positive surgical margins also decrease disease-specific survival rates in patients with breast and bladder cancer and decrease the overall survival rate in patients with colorectal, oral cavity, and lung cancer [OTC+18]. Surgeons currently do not have adequate intraoperative assessment tools during initial breast tumor removal surgery to ensure that the cancer has been completely removed. In addition to adjuvant chemotherapy and/or radiotherapy, positive margins may lead to re-resection if there is enough tissue that can be removed [OTC+18]. The lack of an adequate intraoperative assessment tool results in 23.6% of women having to return for at least one additional operation because their tumor was not completely removed during their primary surgery [HMLMM14], [BBR+09]. In the current clinical setting, histological assessment is still the most extensively used method to

reduce tumor regional recurrence [KFH21]. A pathologist typically sends a pathology report to the doctor after a biopsy or surgery. However, only a small number of hospitals perform breast-conserving surgery utilizing intraoperative pathological assessment of tumor margins because it is a time-consuming process requiring significant expertise and has difficulty detecting ductal carcinoma in situ [BBR+09], [LRN+23]. These additional treatments for cancer recurrence for women who undergo repeated breast cancer surgeries not only result in pain, suffering, and disfigurement but also negatively impact the patient's prognosis, leading to a higher risk of complications and increased costs for the patient and the healthcare system [BBR+09].

One potential solution is the use of OCT, which offers detailed views of tissue structure on the micron scale in situ and in real time. Different from conventional histopathology, OCT does not require the removal and processing of a tissue specimen for microscopic examination, enabling it to detect malignant breast cancer types, e.g., invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS) in real-time [FPBB00], [LRN+23]. Additionally, OCT image resolution is notably superior to common imaging techniques such as ultrasound, MRI, and X-ray imaging [ADFM19]. Research has shown that OCT is capable of generating images with the clarity and contrast needed to differentiate benign from malignant breast tissues [YZR+19]. The introduction of the wide-field optical coherence tomography (WF-OCT) system, crafted explicitly for intraoperative application in breast-conserving surgery (BCS), addresses the challenge of comprehensive lumpectomy margin examination by providing real-time visualization. This WF-OCT technology achieves a 10-micron resolution to a depth of 2 mm, surpassing the capabilities of specimen radiography or ultrasound for BCS margin assessment. Such high-resolution imaging facilitates comparison with histopathological findings, thereby establishing a reliable benchmark for training AI models [SCJ+20], [LRN+23]. Research has verified that the tissue structures observed in the WF-OCT images closely match those seen in corresponding histological samples for both normal and abnormal tissues [BSL+23]. Trained clinicians have clearly distinguished between the specific layers, characteristics, and microstructures of healthy and diseased tissues in these images [RC23], [BSL+23]. Figure 2.1 illustrates the comparative analysis of WF-OCT images and histological slides across various tissue types. Figure 2.2 gives a preview of DCIS and its corresponding histology image.

While OCT has demonstrated significant promise in tumor imaging and evaluation, offering a potential solution to delays in histology report processing and reducing the necessity for follow-up surgeries [VFJB12], its adoption in contemporary cancer treatment and sur-

gical practices remains limited. One reason is the novelty of Wide-Field Optical Coherence Tomography (WF-OCT) technology. Although OCT has been pinpointed as a promising tool, it has yet to be extensively tested across the most prevalent tissue types encountered in surgical oncology on a unified, standardized platform [RC23]. Furthermore, OCT's susceptibility to coherent noise, particularly speckle noise, significantly impacts the clarity, contrast, and visibility of detailed structural information within OCT images [ADFM19]. This susceptibility restricts the diagnostic utility of OCT and presents substantial obstacles for both manufacturers of existing OCT systems and clinicians who rely on OCT imaging for intraoperative decisions. Additionally, OCT's application in cancer diagnosis and treatment is still in its early stages, with clinicians and radiologists facing a scarcity of interpretative data and requiring extensive training periods to achieve proficiency in OCT image analysis and diagnosis.

## 2.2 Speckle reduction algorithms

As a multiplicative noise, speckle noise is a granular texture that degrades quality due to interference among coherent waves caused by the microscopic structure and geometry of the surface within the limited bandwidth of the measured systems [Goo76]. In OCT, the measurement technique relies on the spatial and temporal coherence of optical waves backscattered from tissue. The downside is that this same coherence gives rise to speckle noise [SXY99].

Denoising algorithms have two main streams: conventional filter-based methods and deep learning-based methods. These filter-based method formulations have been well studied across modalities, most targeting specific problems and domain-specific issues that require hardware-level optimization or software-level processing [ESA12]. In hardware optimizations, noise can be sufficiently reduced, but they usually have specific equipment requirements that make it harder for closed-source commercial devices or require the development of specific corresponding systems [LLS+17]. Software-level filter-based optimizations remain a hot topic as different kinds of processing algorithms have their pros and cons. The balance between image quality and computation/acquisition time cost remains challenging across multiple domains such as OCT, ultrasound imaging, and SAR images [PN21], [CJ19].

Moreover, in OCT imaging settings, selecting the appropriate level of image enhancement remains a challenge for users, as speckle information may be clinically important

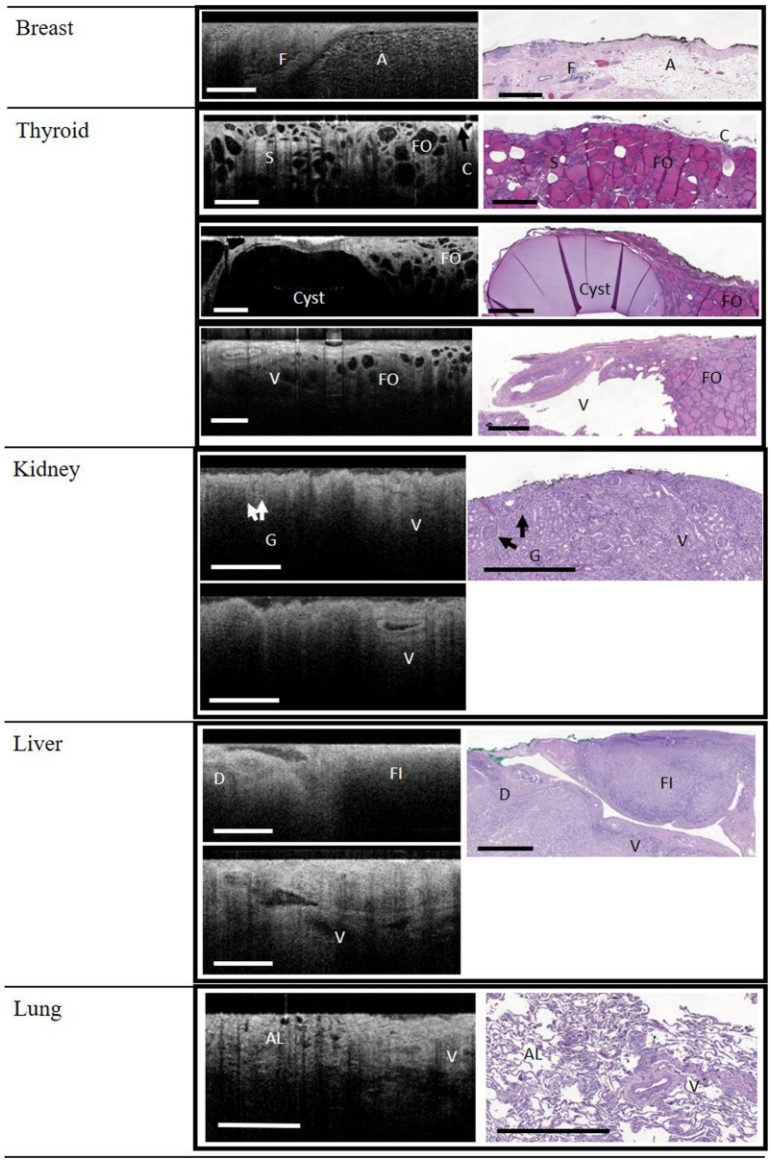


Figure 2.1: WF-OCT images and correlated histology from Breast, Thyroid, Kidney, Liver, and Lung. Reference images for the kidney and liver (lower panels) demonstrate that vessels (V) could be followed across multiple WF-OCT image slices. Abbreviations: A, adipose tissue; C, Capsule; F, fibrous tissue; FO, follicle; S, Stroma; AL, alveoli; D, Duct; FI, fibrosis; G, glomerulus; V, vessel; WF-OCT, wide-field optical coherence tomography. Scale bar: 1 mm. Source: [BSL+23]

[SADJK+22]. Traditionally, customized and device- and tissue-specific frame-averaging is a standard method used in image enhancement methods. However, frame-averaging, in the presence of object motion, can degrade lateral resolution and acquisition time as it requires multiple scans of the same object [RJHT+22]. Due to more available computational training

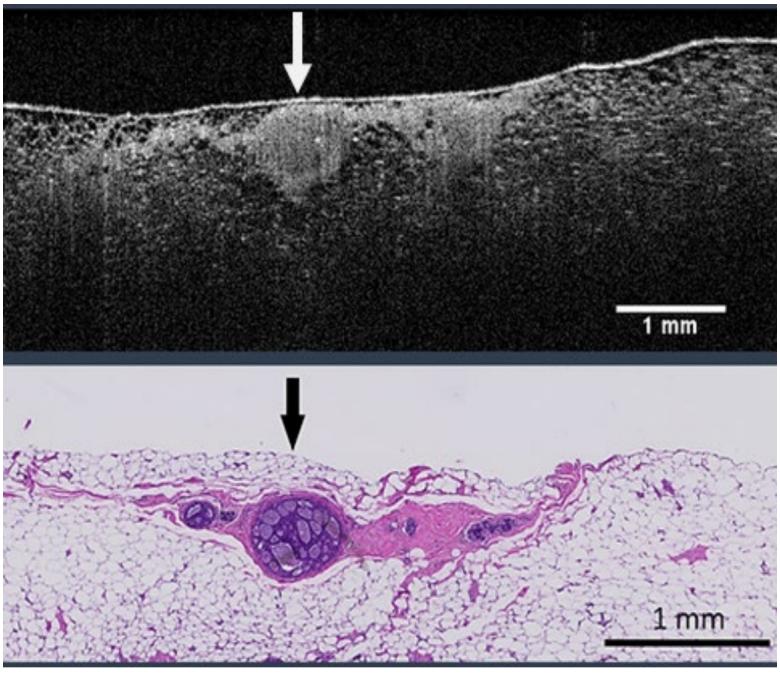


Figure 2.2: WF-OCT image of breast tissue (top) and the corresponding digital pathology image (bottom). The arrow in the pathology image points to ductal carcinoma in situ. Source: [LRN+23]

resources, machine learning-based, or more specifically, deep learning-based denoising algorithms have been extensively explored while balancing the effect of denoising and preserving OCT imaging with moving flow information.

Qiu et al. proposed using feed-forward convolutional neural networks (CNN) and loss functions targeting structural similarity metrics to reduce noise in OCT-B scans of eye images [QHL+20]. Moreover, Mehdizadeh et al. suggested combining fixed pre-trained perceptual loss with deep convolutional neural networks (DCNN) to denoise OCT retina images and increase perceptual sharpness [MMX+21]. Given the success of U-shaped Convolutional Neural Networks in multiple medical imaging tasks [RFB15], U-net has also been adapted for image translation tasks that transform one image into another. In OCT, most methods use U-Net-like deep learning architectures adapted to tissue-type-specific problems such as those with flow features across multiple scans. This makes frame-wise averaging-based denoising methods challenging to perform as they introduce issues with moving features. In single-frame denoising, Schottenhamml et al. proposed an unsupervised method combined with U-Net to preserve temporal information, aiming to retain information of moving features such as flow [SWP+23]. In edge-sensitive loss-based conditional generative networks, a U-Net was

employed as a generator where edge-sensitive loss was proposed to capture edge information in retinal OCT images [MCZ+18]. In the self-fusion neural network, a U-Net model was trained to learn the functionalities of fusing three frames to minimize computational overhead and offer benefits similar to direct three frames of discrete Fourier transform based rigid image-registration, thereby reducing the impact of moving flows in image fusion [RJHT+22].

In OCT tumor imaging, since minimal environmental movement can be avoided during scan times and the breast tissue remains relatively static, the problem of frame-wise averaging blurriness is minimized, leading to a different demand.

## 2.3 Generative adversarial networks

As one of the prominent topics in generative networks, Generative Adversarial Networks (GANs) have been extensively studied [GPAM+14], [MLX+17], [RMC15]. With the continual advancement and development in GAN-based methods in the computer vision domain, Conditional GANs were proposed to control the behavior of GANs and have shown significant generalization abilities [MO14].

Isola et al. proposed Pix2PixGAN, achieving impressive results in image translation tasks by utilizing paired information from both the conditional GANs perspective and the generator/discriminator visibility perspective to fully use the entire dataset for two adversarial networks[IZZE17]. Subsequently, CycleGAN combined cGAN with neural style transfer and super-resolution network backbones to solve the problem of some applications without paired images[ZPIE17]. Compared to Pix2PixGAN, CycleGAN targets image translation/synthesis tasks without a definitive outcome while demonstrating relatively good performance.

The concept of GANs has also been applied in the medical imaging enhancement domain, such as low-dose CT images denoising[WLVI17], CT images super-resolution [AASA22], and medical image synthesis [SJL23].

## 2.4 Pre-trained neural network models

In image restoration tasks, the combination of perceptual loss functions with L1 and L2 loss yields the best results with the same neural network [ZGFK16]. Furthermore, Johnson et al. proposed the use of perceptual loss from pre-trained models with CNNs in image trans-

formation tasks [JAFF16]. In the natural image super-resolution task, the Super-Resolution Generative Adversarial Network (SRGAN), which utilizes ResNet and incorporates deep feature loss to leverage perceptual features from the 19 layers of the Visual Geometry Group (VGG) model, achieved significant visual improvements and enhancements in mathematical metrics such as PSNR and SSIM [LTH+17]. Afterward, Mehdizadeh et al. suggested combining deep features/VGG Loss with feed-forward CNNs targeting virtual features in the OCT denoising task as well[MMX+21]. The study of deep feature loss is facilitated by publicly accessible pre-trained networks trained on large datasets.

Moreover, to overcome the instability of cGAN's performance evaluation, where the minimum loss values in the generator or discriminator may not necessarily indicate the best performance, the Fréchet Inception Distance (FID) scores were used. This approach captures the visual performance of the cGANs by utilizing pre-trained Inception networks to find the optimal Nash equilibrium point [HRU+17].

# Chapter 3

# Materials and Methods

This section will discuss the complete process of end-to-end training of the deep learning pipeline in our work. We introduce the application of cGAN to enhance WF-OCT scans.

## 3.1 From CNN to skip connections

In recent years, deep learning methods have played pivotal roles across various domains. Among these approaches, CNNs have emerged as dominant performers in diverse computer vision tasks. Initially proposed for image recognition tasks, CNNs are constructed with various consequence layers and activation functions. They have demonstrated favorable results in image-to-image and feature-to-feature translation tasks, excelling in pattern recognition. With the evolution of CNNs, researchers have dedicated significant efforts towards developing deeper and more extensive networks, as larger models theoretically yield better performance. Yet, one notable issue with deeper CNNs is the diminishing gradient problem, where the gradient signal decreases in the network layer by layer.

To address this issue, the concept of residual blocks was developed, creating pathways that connect early layers directly to the later layers to ensure critical information is effectively transmitted throughout the network [HZRS16]. In their groundbreaking paper, He et al. introduced residual blocks into Deep CNNs and proposed the Residual Network (ResNet), thus facilitating the conservation of residual values across layers through these short connections. This approach reimagines convolutional layers as learners of residual functions, which are shaped by a blend of inputs from desired preceding layers and outputs from the last layer,

rather than simply using the output of the previous layer as input for the next [HZRS16]. Figure 3.1 showcases a residual block that exemplifies a skip connection. Figure 3.2 shows how network architectures evolved from DeepCNN to ResNet.

In 2015, the U-net architecture was introduced. It has a long-range skip connection, and as a convolutional neural network characterized by its U-shaped structure, this structure ensures vital values are retained in the deeper layer [RFB15]. The U-shaped structure, inspired by encoder and decoder architectures, is enhanced with long skip connections, with connection points selected based on experimental results in the cell segmentation tasks [RFB15]. U-net has notably surpassed classification-based convolutional networks in segmentation tasks due to its effectiveness. Its high utility has made it a preferred tool for various applications within the medical image analysis community.

A typical U-net model consists of two main components: a contracting path that forwards context information to higher resolution layers and an expansive path that upsamples the information, as illustrated in Figure 3.3. This configuration enables U-net to effectively handle segmentation tasks by maintaining essential information throughout the network. In the machine learning community, U-net's adaptability and flexibility have led to its widespread adoption as a primary tool for numerous tasks across different domains. Its versatile architecture allows for modifications to suit various sizes and can be integrated with other architectures.

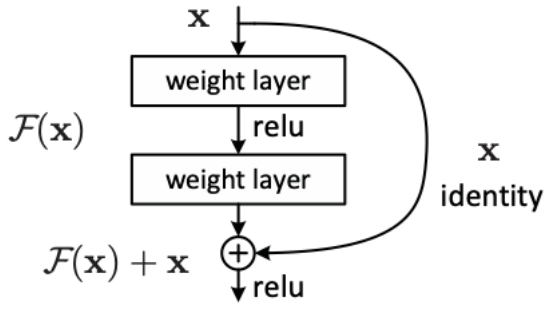


Figure 3.1: Residual Block Source: [HZRS16]

## 3.2 Generative models

Building on the successes of CNNs across various domains, Goodfellow introduced the concept of Generative Adversarial Networks (GANs). GANs, as a machine learning method,

are more deeply rooted in game theory than in traditional optimization-based approaches. They involve a generative CNN network and a discriminative CNN network, formulating a zero-sum game framework [GPAM+14]. Despite their promise, basic GANs often struggle with generating comprehensible images in generative tasks. Nonetheless, GANs have been extensively studied, particularly for generative applications, with a surge in research focused on producing more consistent outputs and adapting them to diverse tasks.

One influential trend in deep learning involves building deeper and wider architectures. Following studies on GANs and deep CNNs, deep convolutional generative adversarial networks (DCGAN) were proposed to explore combining deep convolutional layers with GANs [RMC15]. This approach potentially allows training on higher resolutions and achieving better generative results. As merely scaling up GANs did not yield satisfactory results compared to other types of CNNs, DCGAN introduced architectural constraints to enhance GANs and evaluate discriminator performance over classification tasks. These design rules, tested across multiple datasets and through visualization of CNNs' intermediate outputs, include replacing pooling layers with strided or fractional-strided convolution, utilizing batch normalization, removing fully connected layers in deep structures, using ReLU activation functions for all generator's layers except the output, which uses Tanh, and employing LeakyReLU activation function in the discriminator for all layers. Following the development of DCGAN, Least Squares Generative Adversarial Networks (LSGAN) was introduced, adopting a least squares loss function (L2) for the discriminator rather than the traditional sigmoid cross-entropy loss used in regular GANs [MLX+17]. LSGAN has been shown to generate higher-quality images than standard GANs and offers more stability during the training process.

With the developments in GANs for generative tasks and advancements in other types of deep CNNs, Johnson et al. proposed a style transfer network that utilizes a feed-forward convolutional neural network. Instead of a general per-pixel loss, a perceptual loss utilizing pre-trained models was introduced, enhancing the network's ability to mimic human visual perception [JAFF16]. Johnson et al.'s style transfer network is tailored for image translation tasks and refines several details compared to previous deep CNN architectures. In the style transfer network, all non-residual convolutional layers are followed by spatial batch normalization and ReLU nonlinearities, except for the output layer, which employs a scaled tanh to ensure that the output image pixels fall within the appropriate data range. Notably, such networks modify the residual blocks to exclude zero padding in convolution, instead opting for $3 \times 3$ convolutional layers to mitigate border artifacts. This innovation builds upon the residual block concept introduced by He et al. [HZRS16] and is further enhanced

by adjustments that omit a ReLU activation or batch normalization following the addition operation in the original residual block from Sam et al. [GW16]. Figure 3.4 shows the modifications between Sam et al.'s residual block, and He's original residual block. Figure 3.5 presents the residual block utilized in Johnson et al.'s style transfer network.

Another key development in this domain is the introduction of conditional GANs. In this variation, both the generator and discriminator are conditioned on additional information, $y$, a concept proposed by Mirza et al. [MO14]. This innovative approach has spurred further research and practical applications. Pix2PixGAN, a type of conditional GAN where the discriminator assesses both generated and real images, was introduced by Isola et al. [IZZE17]. This model has demonstrated exceptional performance in image translation tasks, marking another milestone in the evolution of GANs. Pix2PixGAN employs fully convolutional networks, with operations ranging from pixel level to patch level for the discriminator. This flexibility allows the discriminator to be tailored according to the specific requirements of the task. Meanwhile, it employs a U-net architecture for the generator. This setup combines the principles of cGAN with long skip connections to form a U-shaped network, further enhancing the model's effectiveness. Following the success of Pix2PixGAN, Zhu et al. [ZPIE17] introduced CycleGAN, which establishes a cyclical relationship between the generator and discriminator, catering specifically to "no-target" applications. This means that CycleGAN is designed for image translation/synthesis tasks where a definitive outcome is not preferred, thereby bypassing the need for a paired relationship between images and labels. Instead, CycleGAN explores image relationships through a concept known as cycle consistency. Despite this approach, CycleGAN still delivers satisfactory results in image translation tasks. To stabilize training, CycleGAN incorporates the least square (L2) loss from LSGAN for the GAN loss [MLX+17]. Another distinguishing feature of CycleGAN is its generator architecture, which includes three convolution layers, multiple residual blocks depending on the task (six blocks for 128x128 images, nine for 256x256 images), fractionally strided convolution layers, and lastly one convolutional layer. This setup is specifically designed to map features from the deep learning-based style transfer network used in super-resolution tasks [JAFF16]. Figure 3.6 illustrates the architecture of the style transfer network, which is subsequently employed as the generator in CycleGAN.

Except for GANs, there are still many other variants of generative models. The main streams are diffusion models and auto-encoder based models [KW13], [HJA20]. With recent advances in denoising methods, diffusion models were proposed. Diffusion models, with long inference times, are relatively large, require large datasets, and do not necessarily outperform

GANs [HJA20]. They still need time to develop, making them more challenging for our application scope.

On the other hand, although auto-encoders have been well-studied, GANs have had better results in many cases. For instance, in Wasserstein GANs, Variational auto-encoders (VAEs) has been compared, and GANs have shown strong generative results by employing different loss functions and discriminator settings [ACB17]. Because VAEs focus on the approximate likelihood of the examples, they share the limitations of standard models and need to handle additional noise terms. GANs offer much more flexibility in defining the objective functions [ACB17]. In short, training GANs remains the most difficult, but they usually produce better results compared to diffusion models and VAEs [HJA20], [ACB17].

Our method combines the benefits of Pix2PixGAN and CycleGAN as they belong to Conditional GANs. More specifically, we adopted the super-resolution and style transfer network from Johnson et al., used in CycleGAN, into the original Pix2Pix framework as the generator [JAFF16], [IZZE17], [ZPIE17], to meet the resolution and image quality demands of the oct specimen image denoising task.

## 3.3 Conditional Generative Adversarial Networks for Noise Reduction in Optical Coherence Tomography

In this study, we introduce the application of cGANs to enhance WF-OCT scans. We finalize our cGAN model not only from a research perspective but also from an engineering standpoint. Based on the intuition of scaling laws, we start with a small dataset and examine different models. Considering previous experiments' conclusions, we iteratively involve the dataset and test it with improved models. The dataset consists of WF-OCT scans obtained by conducting multiple simultaneous scans with a probe within a designated time frame. This approach assumes the retention of speckle noise information during this period. Our data processing technique, which uses an eight times (8x) averaging method as the ground truth, proves superior. By training the cGAN model with various specimens, our objective is to enable the model to learn the characteristics of speckle noise, enhancing the analysis of Ductal Carcinoma In Situ, Invasive Ductal Carcinoma diseases, and human breast tissues. This approach ensures that the network has truly learned speckle suppression ability rather

than simply over-fitting to specific tissue types. More specifically, we apply the cGAN method to learn a mapping behavior from an observed image x and a noise vector z to a target image y. In other words, this denoising network system learns the representation of a statistical vector that converts the observed image into the target image, thereby removing noise. In our case, the objective of the OCT denoising network is to produce a consistently enhanced reconstructed image under the WF-OCT system. Such noise vectors are ignored to ensure deterministic mapping behavior, especially considering the input being conditioned on is already sufficiently complex and contains necessary speckle noise information in a high-dimensional space. This is based on the assumption that CNNs can learn such speckle noise characteristics while preserving the necessary clinical information, increasing contrast without degrading image quality. The goal of the generator is to produce enhanced 1× averaging images that are as close as possible to the 8× real images, effectively 'fooling' the discriminator. The discriminator aims to differentiate between the generated results and the actual 8x images. A key distinction between conditional GANs and traditional GANs is that the discriminator evaluates two inputs: the generated image G(x), x', and the ground truth 8× OCT image. Figure 3.7 shows the overview of the proposed OCT denoising framework.

Additionally, we suggest using the Fréchet Inception Distance (FID) to determine the optimal checkpoints during each training phase. The FID calculation formula is given in equation 3.1 from [HRU+17]. The Fréchet distance $d(.,.)$ between the Gaussian with mean $(m, C)$ obtained from $p(.)$ and the Gaussian with mean $(m_w, C_w)$ obtained from $p_w(.)$ is known as the "Fréchet Inception Distance" (FID), where $C$ is the covariance, $p_w(.)$ is the generated image distribution, and $p(.)$ the target images distributions [HRU+17]. FID employs a pre-trained Inception network to score the similarity between the network's output and the ground truth. A stabilization in FID scores indicates minimal changes in the perceived differences between image groups. Given that FID is a biased estimate, epochs exhibiting local minimum FID scores should be prioritized for further evaluation over those with global minimum scores. This strategy helps in conserving both human visual inspection efforts and computational resources.

$$d^2((m, C), (m_w, C_w)) = ||m - m_w||_2^2 + \mathrm{Tr}(C + C_w - 2(CC_w)^{1/2}). \tag{3.1}$$

### 3.3.1 Architecture

The generator (G) of the cGAN translates a single-frame noisy B-scan (1x averaging) into a noiseless image, simulating the process of multi-frame scan averaging. It operates under the assumption that $Img_{noisy} = Img_{clean} + N$, where both the clean and noisy images share the same underlying signal structure. This assumption allows the model to learn de-speckling translation procedures.

During training, the OCT denoising network's generator integrates features from CycleGAN and Johnson et al.'s ResNet-based generator network, as our experiments have shown that ResNet architectures yield superior image quality compared to U-Net type generators [JAFF16] [ZPIE17]. The network begins with an initial convolution layer, followed by two downsampling convolutional layers that double the number of filters while halving the spatial dimensions. This setup is succeeded by nine residual blocks, each designed to perform a series of transformations while preserving the feature maps' spatial dimensions. Subsequently, two upsampling stride convolutions transpose, leading to a final convolutional layer that transforms the feature maps to a single output channel in our denoising translation task. Finally, a Tanh activation function normalizes the output.

The discriminator is selected using a one-by-one PixelGAN classifier, as described in the original pix2pix paper [IZZE17]. The model is specifically designed to discriminate based on assessing whether individual pixels are real or fake, providing better image quality in the OCT denoising task than the pix2pix favored 70 by 70 PatchGAN. This outcome may be because our use of black and white two-channel images, instead of the RGB color images as in the original Pix2Pix image translation task, results in less image information. In denoising tasks, the focus shift to speckle reductions applies to individual pixels rather than spatial information translation. The network begins with a convolutional layer with a kernel size of 1x1, a stride of 1, and no padding. Another one-by-one convolutional layer then increases the depth to 128. After this, a normalization layer is applied, where the presence of bias is determined by the batch normalization layer, followed by a LeakyReLU activation. The final convolutional layer further processes the feature maps into a single output channel with a 1x1 kernel that provides the discriminator's verdict on each pixel's authenticity.

### 3.3.2 Objective functions

The objective of the cGAN is learning a mapping from an observed image $x$ and a noise vector $z$ to a target image $y$. The expectation of the conditional GAN loss function is

$$\begin{aligned} L_{cGAN}(G, D) =& \mathbb{E}_{x,y}[\log D(x, y)] \\ &+ \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \end{aligned} \tag{3.2}$$

To understand the adversarial relationship from a loss perspective, the discriminator tries to maximize the difference between the real image and the generated fake image and, in turn, maximize this objective. Note that $D(x, y)$ represents the probability that $(x, y)$ or $(1x \text{ image}, 8x \text{ image})$ is a real pair according to the discriminator's judgment. $D(x, y) = 1$ indicates the discriminator's classification of real, and $D(x, y) = 0$ indicates the discriminator's classification of fake. The first term $D(x, y)$ encourages the discriminator to output high probabilities (close to 1) for real images. Taking the logarithm of a high probability (close to 1) results in a value close to 0. The second part $\mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))$ is the expected value of the logarithm of one minus the discriminator's output for fake images generated by the generator. This term encourages the discriminator to output low probabilities (close to 0) for fake images. When taking the logarithm of the value $(1 - D(x, G(x, z)))$ that is close to 1, the output would be close to 0; maximizing the function $\log(1 - D(x, G(x, z)))$ pushes the discriminator to recognize fake images as fake. At the same time, the generator tries to make $D(x, G(x, z))$ as close to 1 as possible; minimizing this function pushes the generator to fool the discriminator into treating generated fake denoised images as real $8x$ images. During the iteration, the discriminator is always optimized first. The loss function of the discriminator can be considered as:

$$L_D = (l_{cGAN}(D(\text{real}), 1.0) + l_{cGAN}(D(\text{fake}), 0.0)) \times 0.5 \tag{3.3}$$

Where $l_{cGAN}(D(\text{real}), 1.0)$ refers to the selected adversarial cGAN loss value between the real image pair $(1x, 8x)$ and its target label 1.0, and $l_{cGAN}(D(\text{fake}), 0.0)$ refers to the selected cGAN loss value between the fake image pair $(1x, G(1x))$ and its target label 0.0. We choose the mean squared error loss (L2 loss) as the adversarial cGAN loss. As the L2 loss is more stable during the training [ZPIE17], [MLX+17].

After the discriminator is updated, the generator loss function is given by combining the

cGAN objective with another generator loss function term:

$$L_G = l_{cGAN}(G(x,z),1.0) + \lambda \times l_{\mathrm{L1}}(G(x,z),y) \tag{3.4}$$

Where the parameter $\lambda$ represents a weight for the second part of the generator loss, $l_{\mathrm{L1}}$, which can be chosen from any loss function. In our case, we use $L1$.

Then, cGAN's final objective becomes:

$$\begin{aligned} cGAN =& _{GD} L_{cGAN}(G,D) \\ &+ \lambda \times l_{\mathrm{L1}}(G(x,z),y) \end{aligned} \tag{3.5}$$

### 3.3.3 Implementation

To ensure training consistency, the cGAN's generator and discriminator are trained and optimized simultaneously. Initially, one gradient step is taken on the discriminator, followed by one step on the generator. Based on recommendations from previous GAN literature, we halve the discriminator's loss to slow down its training and prevent it from overpowering the generator early in the training process. Both the generator and discriminator use the Adam optimizer [KB14], with an initial learning rate set to $10^{-4}$ and momentum parameters set as $\beta_1 = 0.5$ and $\beta_2 = 0.999$. We initiate our networks with weights from a Kaiming normalization distribution [HZRS15]. This choice of weight initialization is motivated by the presence of outlier pixels in the image that have pixel intensities close to the minimum values. Kaiming normalization provides the generator with the initial capability to converge quickly. We maintain the initial learning rate for the first $x_1$ number of epochs, then linearly decay the rate to zero over the following $x_2$ epochs. These epoch durations vary depending on the model's convergence rate. The training batch size is set to 1, as it allows the network a greater potential to avoid local minima, compared to using a batch size greater than 1, which results in updating gradients based on the average batch result. This is particularly important in image denoising tasks.

As described in the objectives function section, we use the LSGAN approach, employing an L2, least square loss, which has proven stable during training and effective in generating high-quality images [MLX+17], [IZZE17]. In the context of the original art style transfer purpose, the cGAN noise vector can be set randomly, and a dropout layer might be added.

However, to ensure deterministic results in the output of OCT deep learning reconstruction for denoising purposes without manipulating image features, the random noise vector $z$ is set to zero, and the dropout layer is removed.

The code was adapted from the CycleGAN GitHub repository using PyTorch and the wandb package [PGM+19], [Bie20], [ZPIE17]. Experiments were primarily conducted on a V100 GPU with 16 GB of memory and a T4 GPU with 16 GB of memory.

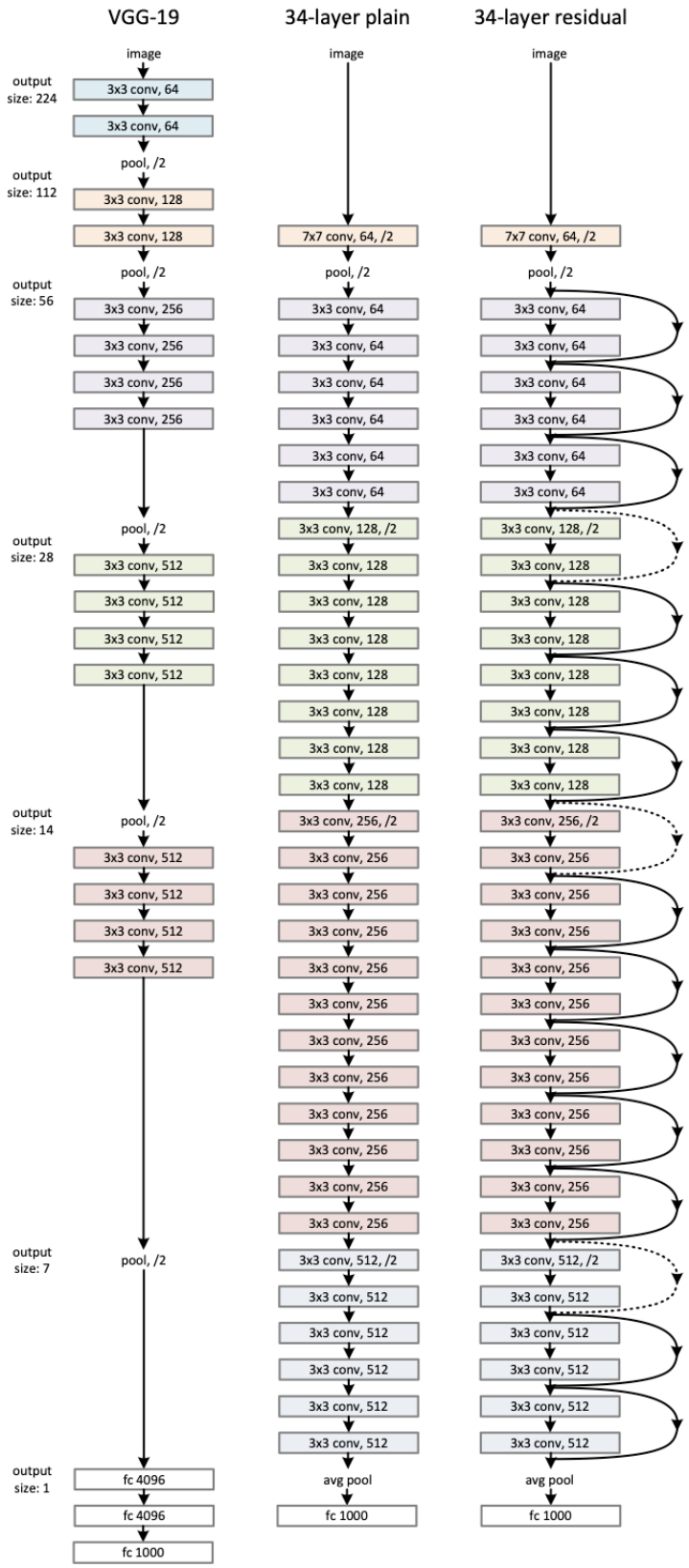




Figure 3.2: Network architectures evolve from DeepCNN to ResNet. From left to right: a Deep CNN model (VGG-19), a plain Deep CNN with 34 layers, and a Residual Network with 34 layers. Source: [HZRS16]

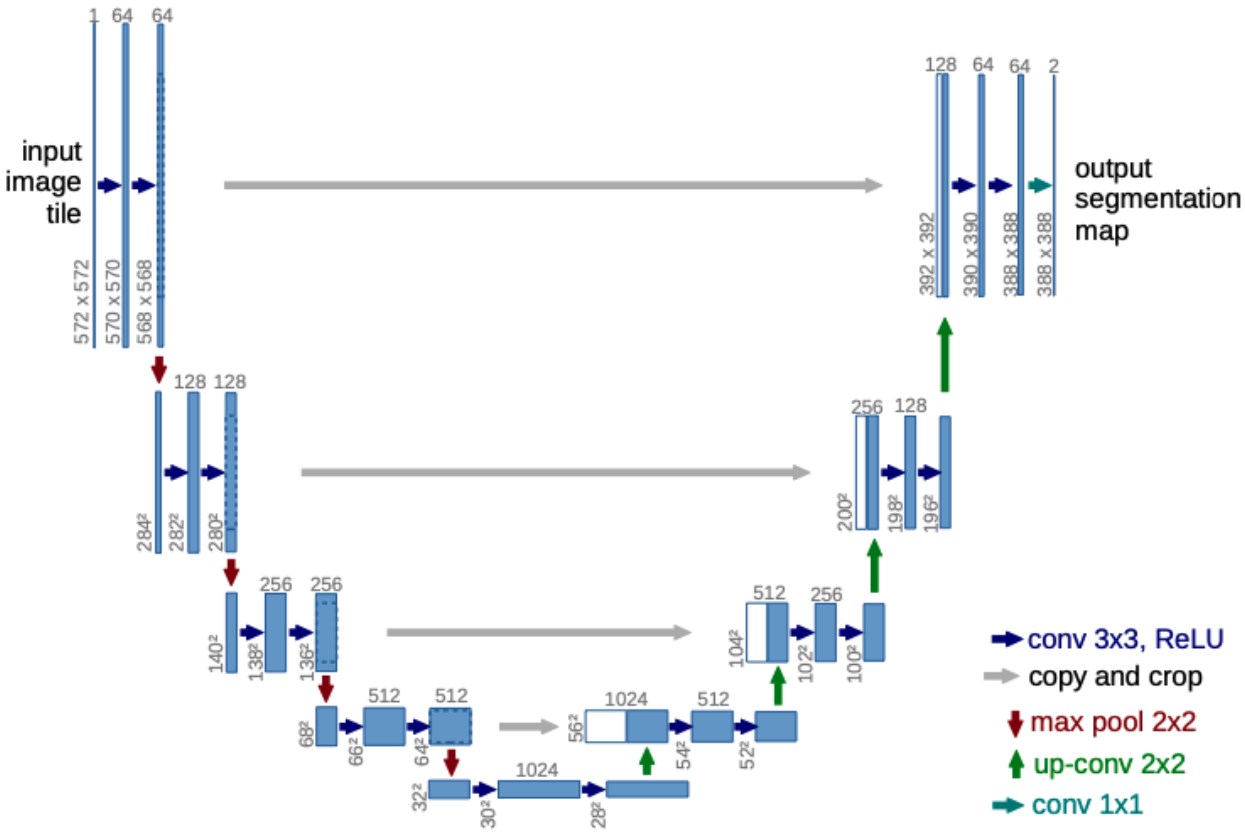


Figure 3.3: The U-net architecture, originally designed for segmentation tasks, incorporates skip connections between early and later layers as the network deepens. This design enables the model to learn and transmit image features across layers without losing essential information. Source: [RFB15]

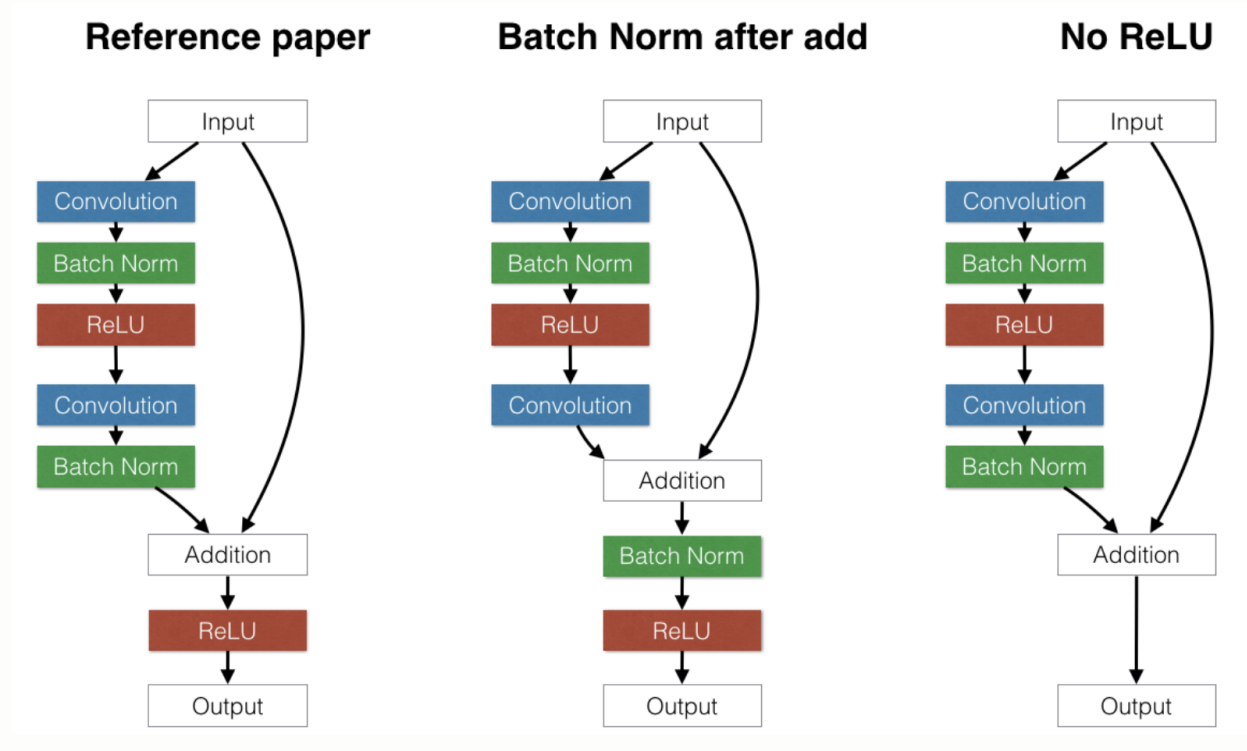


Figure 3.4: Different residual block architectures tested on the Image-Net dataset; from left to right: He's original recommended residual block, a residual block with a batch normalization operation after addition, and the best result residual block. Source: [GW16]

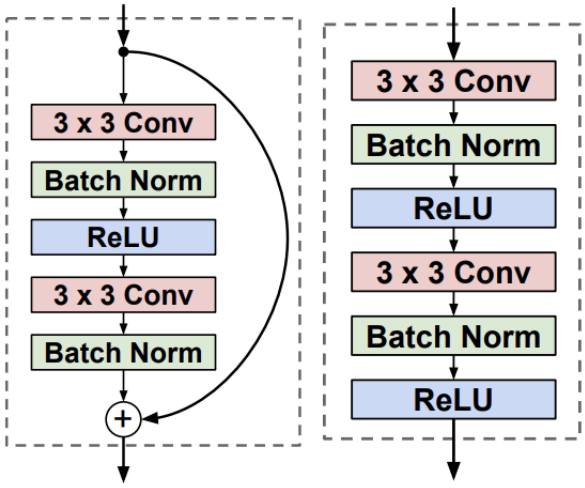


Figure 3.5: Style transfer task improved residual block Source: [JAFF16]

| ×4 | | ×8 | |
|---|---|---|---|
| **Layer** | **Activation size** | **Layer** | **Activation size** |
| Input | $3 \times 72 \times 72$ | Input | $3 \times 36 \times 36$ |
| $64 \times 9 \times 9$ conv, stride 1 | $64 \times 72 \times 72$ | $64 \times 9 \times 9$ conv, stride 1 | $64 \times 36 \times 36$ |
| Residual block, 64 filters | $64 \times 72 \times 72$ | Residual block, 64 filters | $64 \times 36 \times 36$ |
| Residual block, 64 filters | $64 \times 72 \times 72$ | Residual block, 64 filters | $64 \times 36 \times 36$ |
| Residual block, 64 filters | $64 \times 72 \times 72$ | Residual block, 64 filters | $64 \times 36 \times 36$ |
| Residual block, 64 filters | $64 \times 72 \times 72$ | Residual block, 64 filters | $64 \times 36 \times 36$ |
| $64 \times 3 \times 3$ conv, stride 1/2 | $64 \times 144 \times 144$ | $64 \times 3 \times 3$ conv, stride 1/2 | $64 \times 72 \times 72$ |
| $64 \times 3 \times 3$ conv, stride 1/2 | $64 \times 288 \times 288$ | $64 \times 3 \times 3$ conv, stride 1/2 | $64 \times 144 \times 144$ |
| $3 \times 9 \times 9$ conv, stride 1 | $3 \times 288 \times 288$ | $64 \times 3 \times 3$ conv, stride 1/2 | $64 \times 288 \times 288$ |
| - | - | $3 \times 9 \times 9$ conv, stride 1 | $3 \times 288 \times 288$ |

Figure 3.6: Style transfer Network architectures used for ×4 and ×8 super-resolution. Source: [JAFF16]

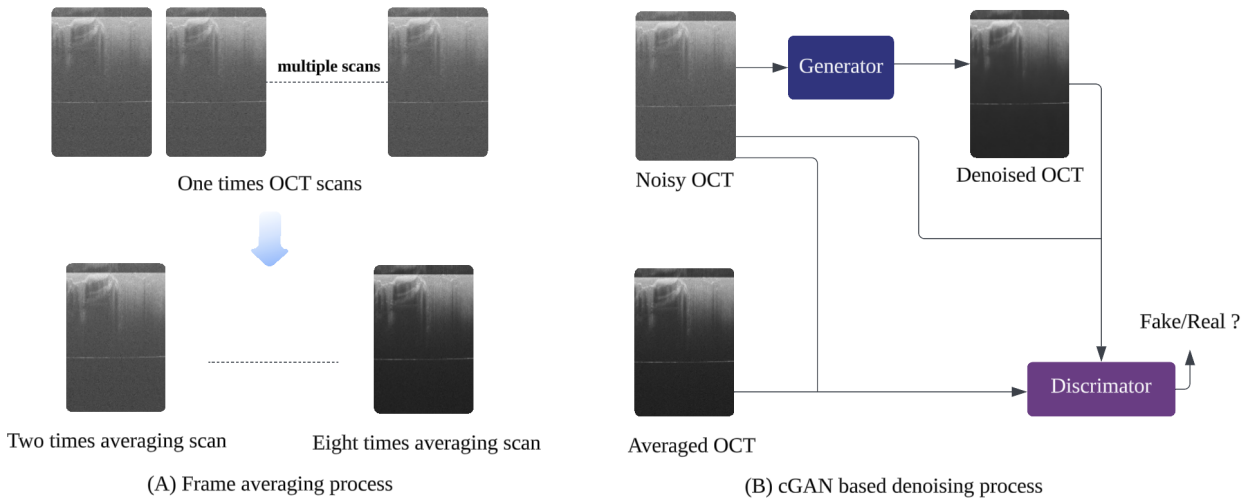


Figure 3.7: The overview of the proposed OCT denoising framework consists of two main components. Figure A illustrates the averaging process, transforming aligned raw OCT data (1x) into averaged OCT data. Figure B depicts the two components of the framework: the generator and the discriminator. The generator takes a high-noise image as input and produces a high-quality image as output. Meanwhile, the discriminator evaluates the artificially generated images by the generator and the real, low-noise images resulting from higher frame averaging. The discriminator's objective is to effectively differentiate between the two types of images.

# Chapter 4

# Experiment

## 4.1 Datasets overview

This work utilizes a combination of privately acquired organic and inorganic specimens or phantoms from Perimeter Medical Imaging AI S-Series and equivalent systems, along with clinical breast tissue samples previously obtained, such as cancerous DCIS tissue from the company. These datasets result from a partnership with Perimeter Medical Imaging AI Inc. Special attention has been paid to ensuring that clinical breast tissue samples are not used in the training or validation phases but are exclusively reserved for testing purposes.

The data collection process includes several crucial steps: preparing the samples, configuring the service tool and device, capturing images of the specimens, and storing the data securely. Each phase is accompanied by specific guidelines and must be executed consistently to maintain data integrity. Data collection was conducted multiple times with a Perimeter scanner for every specimen. To thoroughly assess differences between devices, specimens were imaged using two different devices. The acquired raw OCT datasets were then utilized to produce frame-wise averaged PNG images. For instance, to generate data averaged 8×, we first acquired 1x images of each specimen and then averaged eight of these images. The acquired data produces high-quality images, highlighting regions separated by the glass line. For example, Figure 4.1 for a human specimen illustrates the glass region and the user's region of interest, which encompasses both the signal area corresponding to tissue and the noise area within the glass line.

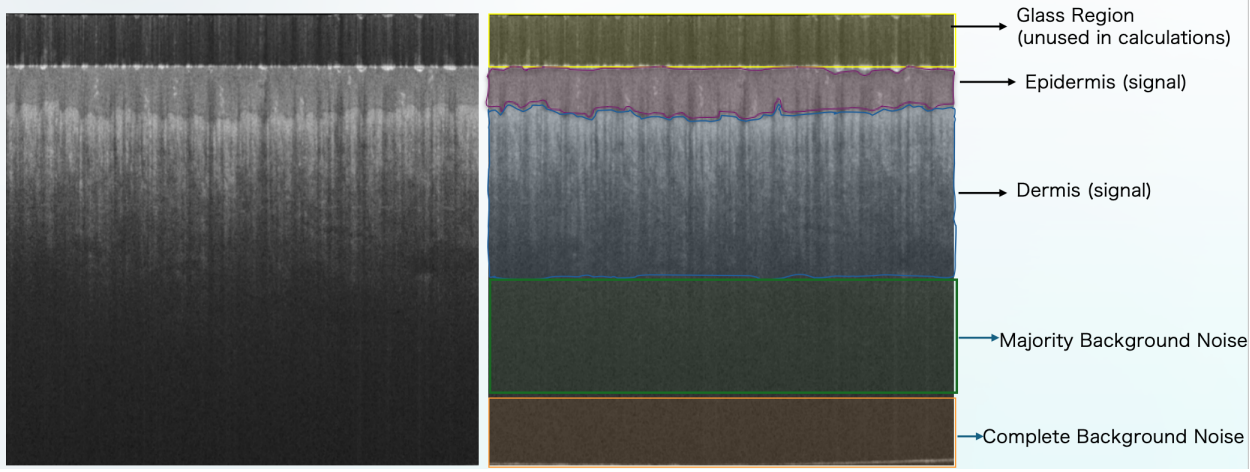


Figure 4.1: Human finger OCT image Region breakdown

## 4.2 Designing Data: Data Science for OCT Dataset

Our approach to designing data involves an iterative process rooted in our data collection and applied ML case study. We continually update our dataset based on experimental results. At the onset of the concept phase, our dataset included a limited variety of data types—totaling five, with a significant skew towards one data type. Each type had 1-2 samples, except for one category, which included 50 distinct samples. We observed that models trained on more balanced datasets tend to generalize better. Interestingly, a model achieved relatively good outcomes with only 810 images, outperforming models trained with a larger number of images but more focused on the same sample region of a single tissue type. To further explore the data imbalance issue and determine whether models perform better with diverse types and regions, we collected additional data to create a more balanced dataset. Consequently, we expanded the dataset to include nine types of cases, which led to improved model performance.

In the final production phase, we directly addressed the challenge of data imbalance by amassing 12 data types in a balanced fashion, except for outliers. More specifically, we identified outlier data types, such as 'air,' demonstrating little to no signal-to-noise Ratio improvement through the averaging process. Data from two devices were labeled for analysis to discern how variations between devices influence the data and, subsequently, model performance.

## 4.3 Evaluation metrics

To comprehensively review the performance of the OCT denoise system, this work utilizes evaluation metrics such as the signal-to-noise (SNR) and the contrast-to-noise (CNR), with the selection of CNR and SNR being guided by expert observers' input, which are also integral to the evaluation process, as highlighted in devalla2019deep.

SNR measures the signal level relative to background noise, indicating the clarity of the signal. Equation (4.1) outlines the calculation of SNR in decibels.

$$\mathrm{SNR}_{dB} = 10 \log_{10} \left( \frac{\mu^2_{\text{signal}}}{\sigma^2_{\text{noise}}} \right) \tag{4.1}$$

CNR assesses an image's contrast relative to its noise. However, metrics like CNR require a professional level of background tissue labeling in breast cancer use cases. Thus, to provide a fair comparison, only selected clinical images demonstrate CNR changes as proof. The metric is shown in Equation (4.2).

$$\mathrm{CNR} = \frac{|\mu_{\text{signal}} - \mu_{\text{background}}|}{\sigma_{\text{noise}}} \tag{4.2}$$

Figure 4.2 captures the SNR and CNR equations and shows how they can be calculated in clinical OCT images.

## 4.4 Ablation study

The experiment is structured into the concept phase and the production phase. The concept phase features a less restrictive training set and a smaller volume of data, while the production phase employs a more refined experimental setup. The primary goals across these phases are identifying the optimal configuration for the cGAN model and curating the most effective dataset for model training.

Throughout both phases, the evaluation of cGAN models was thorough, involving over 60 experiments from model and data science perspectives. These experiments explored a variety of configurations, including differences in architecture, such as distinct generator/discriminator networks, and variations in hyperparameters, such as weight initialization

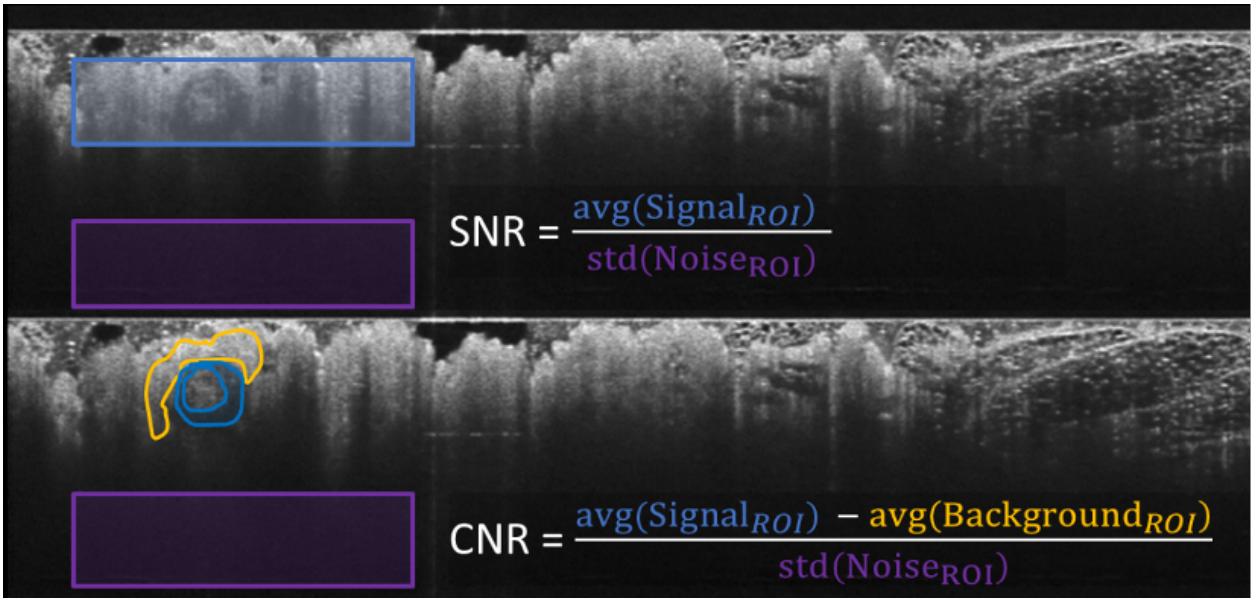


Figure 4.2: Measurement methodology of SNR and CNR in WF-OCT images, where the bottom portion of the image allows extraction of noise parameters while features like the DCIS in this image allow signal measurement.

methods, learning rates, loss functions, optimizers, and epoch decay. From the dataset perspective, experiments included increasing the number of paired images or experimenting with different data configurations.

### Concept phase

During the training and hyperparameter search in the initial concept phase dataset, we discovered that the cGAN with a generator employing ResNet and a discriminator using 1-1 pixel convolutional neural networks yielded the best results. As the U-Net 256 ( U-Net for 256×256 input images) and U-Net 128 ( U-Net for 128×128 input images) generator settings did not produce satisfactory results on the validation set, we observed that significant speckle noise information was still preserved. Example inference images for the validation set are provided in Figure 4.3 and Figure 4.4.

As an open question regarding the optimal training dataset configurations for the OCT denoising task, the finalized cGAN was experimented with various dataset configurations, including OCT images sourced from tomatoes, air (no target), NPL, chicken, and tomatoes, yielding numerous results. In the concept phase, the training dataset that produced the

best-performing model contained 3220 images. This included 2500 images of tomatoes (50 images/regions × 50 regions). The remainder were non-tomatoes, comprising 60 each for Ginger, Wedges, Air, Daiko, Chicken A 50, Chicken B 50, NPL 50, No Target 50, and Lamb Brain 180 (60 for each brain OCT sequence). We observed that models trained with various specimen types achieved better outcomes than those trained with multiple views or regions of a single specimen type. Collecting both organic and non-organic tissues could help the deep-learning model learn the noise characteristics. With a preference for the OCT dataset, we conducted hyperparameter searches to improve model convergence and performance. These included weight initialization methods such as Kaiming normalization, Xavier normalization, and Gaussian distribution N(0, 0.02) cited from the CycleGAN paper [GB10], [HZRS15], [ZPIE17]. This work's experiments show that Kaiming normalization yielded the best results, as it preserved the majority of signals and minimized the impact of the ReLU activation function's tendency to eliminate signals, a problem often referred to as the 'dead ReLU' issue. Kaiming normalization setting is particularly important since the cGAN's generator maintains the original ReLU settings in the residual blocks. An illustrative example of these results can be seen in Figure 4.5. The reason for this phenomenon is that a batch size of 1 provides the cGAN model with a more effective way to denoise images. A larger batch size tends to average the differences between various inputs, which can lead to the model overfitting and negatively impacting the noise statistics. Figure 4.5, using hold-out validation, demonstrates that models with larger batch sizes are more prone to overfitting in OCT denoising tasks. Altering other hyperparameters, such as the learning rate, does not necessarily improve the model's generalization ability. Figure 4.6 illustrates the impact of different learning rates on the training outcomes of the denoise cGAN with batch size = 2. While a lower learning rate resulted in fewer spikes during training, the learning rate of $10^{-4}$ still demonstrated the best performance. It offers a greater potential to avoid local minima without causing substantial image differences.

The experiments have demonstrated the challenge of balancing the generator loss function, the adversarial loss, and discriminator losses for real and fake images. Oscillations may occur in the discriminators when a powerful generator is trained, yet such a strong de-speckling generator is desirable. Instead of solely relying on changes in training loss functions to determine the optimal stopping point, we use the change in Fréchet Inception Distance (FID) as a secondary indicator to decide when to stop training the denoise cGAN model. The FID score, which captures the visual differences between images, stabilizes as the model approaches convergence, as previously proposed for finding the Nash equilibrium

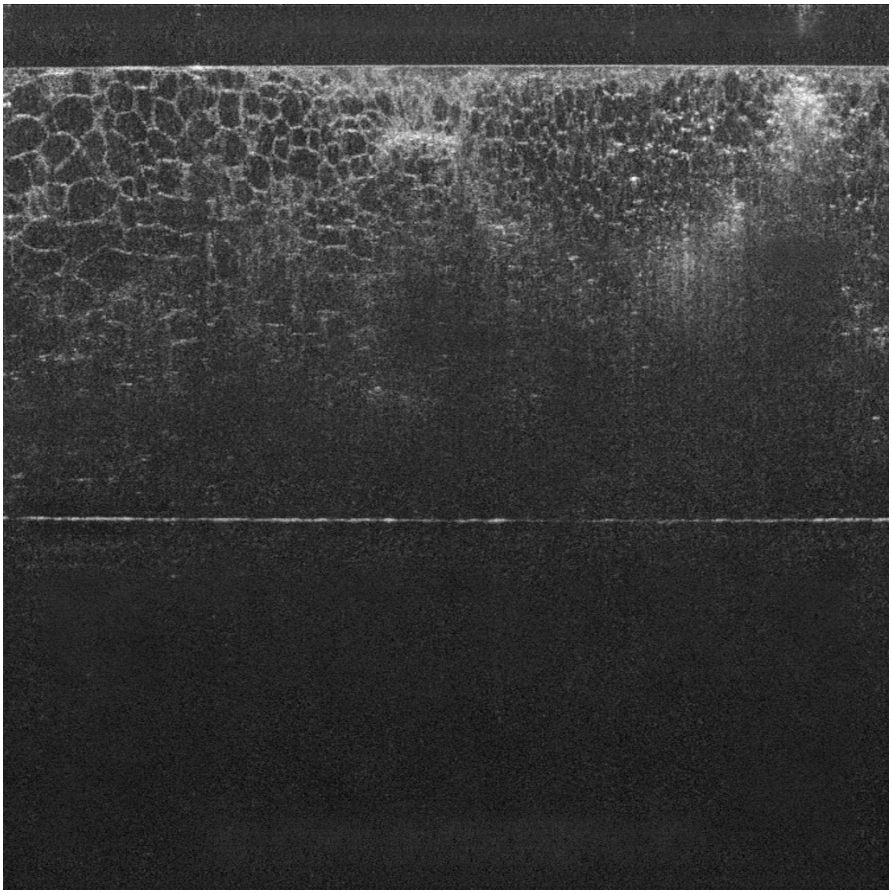

Figure 4.3: result of U-Net 128 denoised 1x grape validation image. The training dataset has 810 images.

in traditional generative adversarial networks heusel2017gans. As illustrated in Figure 4.7, sharp fluctuations in FID values indicate that the model has not yet converged, as evidenced by significant alterations in the cGAN denoised images that are perceptible to humans. As training progresses, FID values tend to stabilize, indicating convergence toward more consistent and reliable model performance.

### Production phase

The objective of the production phase experiment is to address two key questions: Firstly, can the performance of the cGAN be enhanced through the provision of higher-quality data for the experiment? Secondly, how does the model perform when applied to a larger dataset?

A larger and more balanced dataset is provided in production. It includes nine different

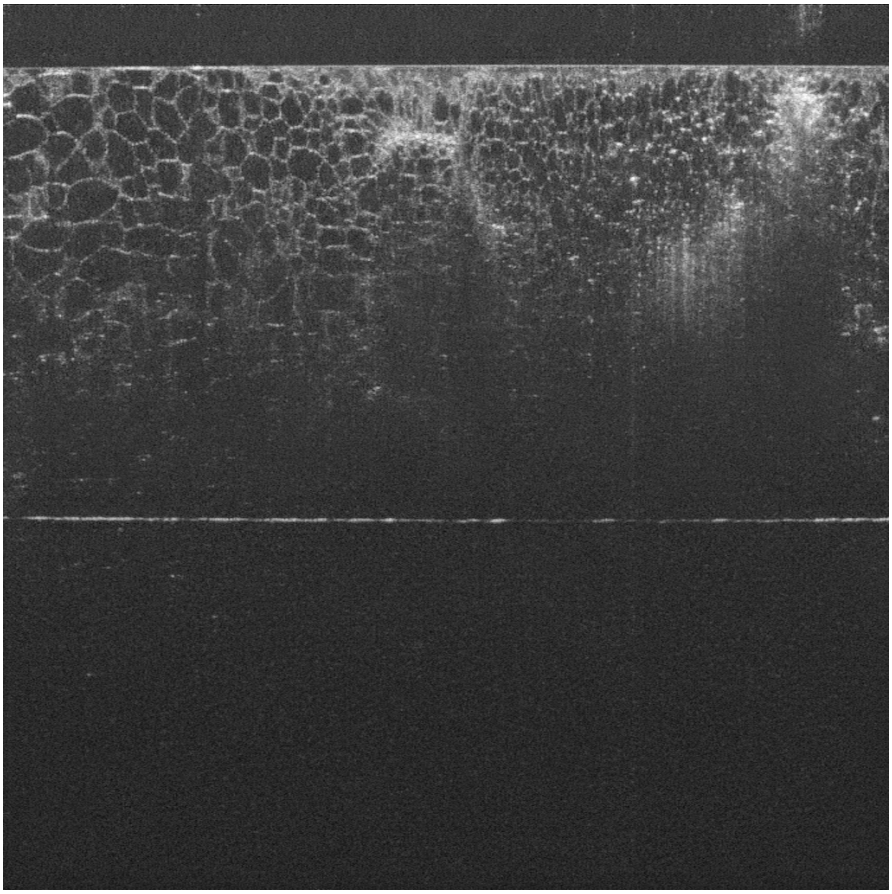

Figure 4.4: result of U-Net 256 denoised 1x grape validation image. The training dataset has 810 images.

organic specimens, each collected from five distinct regions, and five non-organic specimens, each with its unique regions, are also collected. All specimen regions have 16 1x scans, and 16 1x averaging images would result in $12,870$ (calculated from $16C8$) 8x averaging combinations. To achieve sufficiently reliable training results, we only used the first 100 8x averaging images as the ground truth. Similarly, creating a 1 to 1 relationship between 1x averaging and 8x averaging would result in $100 \times 16$ pairs for each region of specimens, giving a total of $100 \times 16 \times 49$ regions $= 78,400$ pairs for each device. This still represents a large dataset; thus, we only use a randomly sampled small percent of images from those pairs for training. The thumb finger specimen validation set is completely excluded from the training, as we aim to assess the model's generalization ability.

In the production phase experiment, we use the FID scores as indicators to further evaluate the optimal training stop point and provide evidence of convergence after the generator's

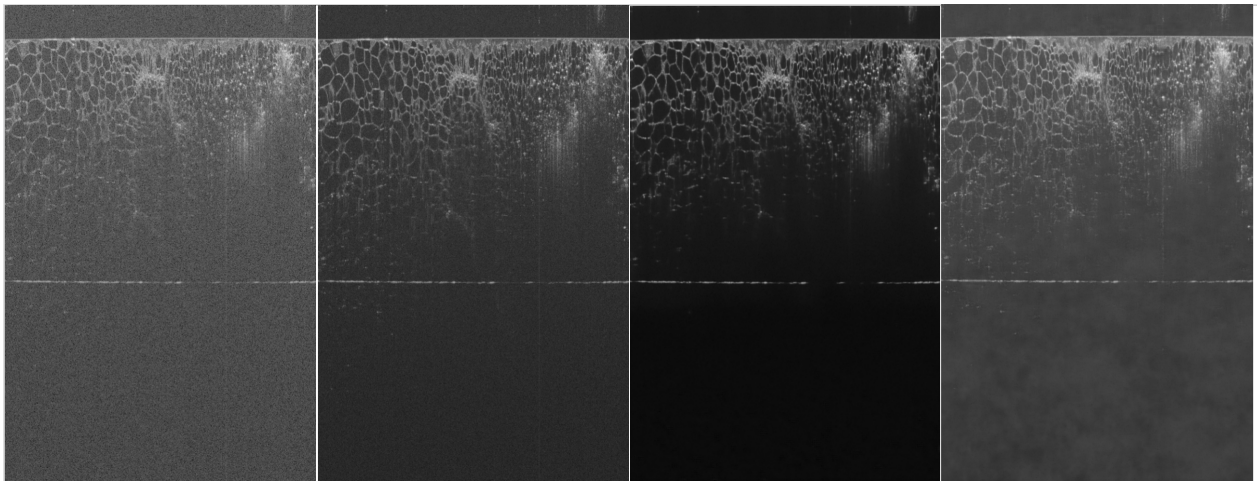

Figure 4.5: Grape Validation Performance: Comparison of models trained with 3320 images, utilizing different batch sizes. From left to right: 1x image input, 8x averaging ground truth, cGAN model denoised 1x image with batch size = 1, and cGAN model denoised 1x image with batch size = 2.

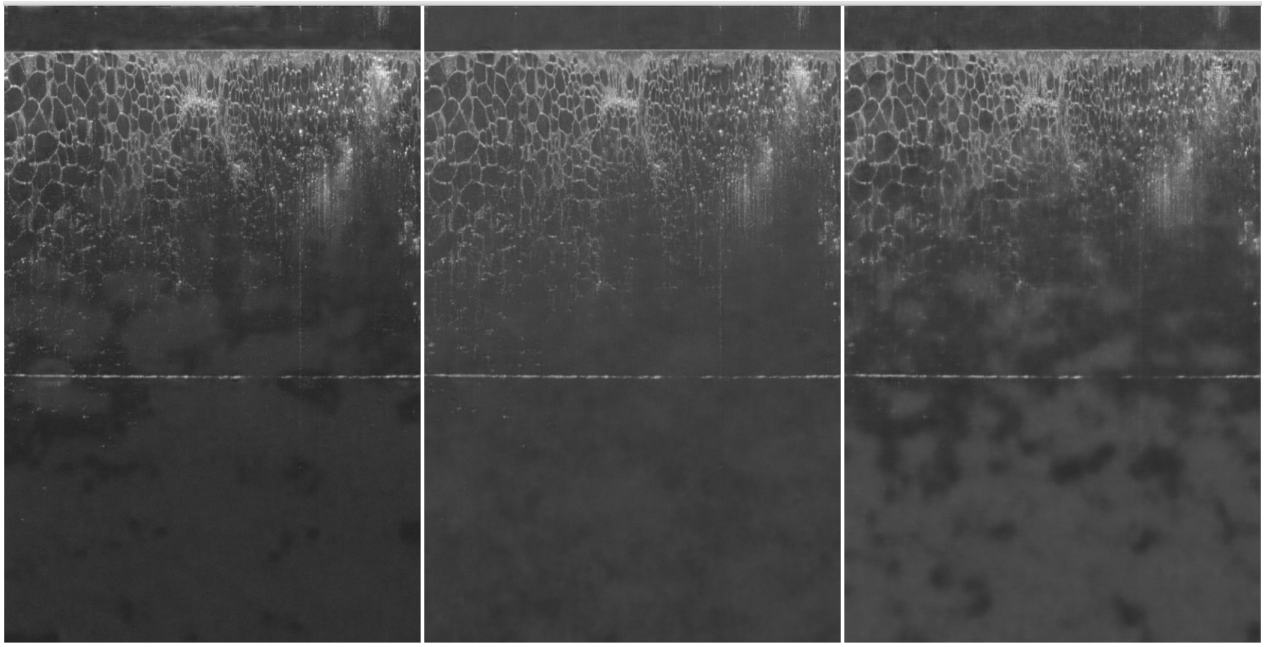

Figure 4.6: Grape Validation Performance: Comparison of models trained with 3,320 images at different learning rates with batch size = 2; all other variables remained the same. From left to right: cGAN denoised 1x image with learning rate $10^{-3}$, cGAN denoised 1x image with learning rate $10^{-4}$, and cGAN denoised 1x image with learning rate $10^{-5}$.

training loss functions have indicated convergence. Figure 4.8 provides an example of the FID score from the validation set for the model trained on a combination of two devices.

Given that the best architectures were assessed during the preliminary research phase, we conducted two experiments to evaluate the differences between devices further. The first involved sampling 10 percent of the data from each device individually. In the second experiment, we randomly sampled 5 percent of Device 1 and 2 pairs, creating a training dataset that includes data from both devices. During training, we discovered that a cGAN trained with data from an individual device did not provide superior performance compared to a cGAN trained with data from both devices. In the ablation study for specimen categories, we focused on key evaluation metrics, such as SNR changes from 1x input to 8x ground truth. Compared to organic data, we observed that non-organic tissues, such as air and glass, exhibited significantly less improvement during the frame-wise averaging process. However, removing this non-organic data did not necessarily aid the cGAN in achieving more optimal convergence. Therefore, the generative model (cGAN) prefers a larger dataset and is not merely learning to overfit specific tissues.

Regarding hyperparameters, we have observed that the cGAN's performance is optimized at a learning rate of $10^{-4}$. Table 4.1 presents various settings of the dataset and the model's performance for the human figure dataset. The table also includes 1x averaging as further evidence in the ablation study.

## 4.5 Results of Breast Cancer Test Image Data

Feasibility testing has demonstrated that reducing noise in images can achieve significant improvements in both the SNR and CNR. Furthermore, sharper-resolution images are possible. These enhancements significantly improve image quality, potentially allowing for higher resolution and/or denser measurements within clinically feasible scan times, as opposed to conventional image reconstruction of WF-OCT images.

This section evaluates the robustness of the denoising cGAN through tests on unseen clinical data to assess its performance in real clinical scenarios. Despite the clinical WF-OCT data originating from conventional processing methods, compared to the cGAN's raw 1x averaging input, notable improvements in metrics are still observed. Figure 4.9 shows a representative DCIS case with expertly defined regions using ImageJ (version 1.54g and free-hand ROI Tool) [SRE12]: the DCIS cancerous signal region (core), ductal epithelium cells

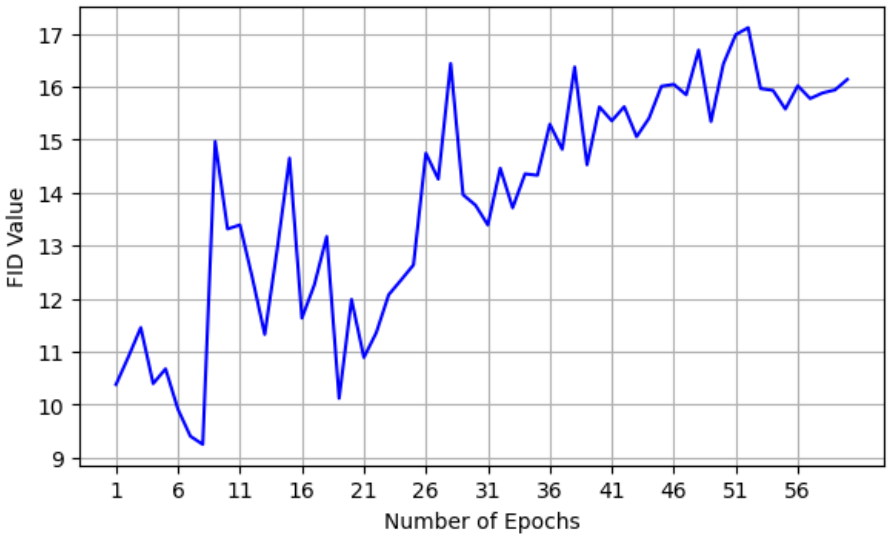


Figure 4.7: The FID value for the grape validation dataset reflects the model's performance, trained without grape images. Collected during the concept phase from the same device, the training dataset comprises a total of 3120 images, predominantly consisting of 2500 tomato images and 620 images of other types, such as chicken breast, lamb brain, surface temperature test phantoms, and air. Average Fréchet Inception Distance of images from models trained on different datasets. The figure displays the model's validation results across various epochs, demonstrating performance consistency. The validation set adheres to the same collection procedures and preprocessing methods as its training set.

(rim), and noise region for calculating CNR and SNR. Figure 4.10 below displays measurement results comparing conventionally processed images to cGAN enhanced outcomes for DCIS using labels provided in Figure 4.9. This further demonstrates the model's effectiveness in enhancing image quality for both the core and rim regions. Figure 4.11 demostrate other example clinical test images from cGAN ehancment.

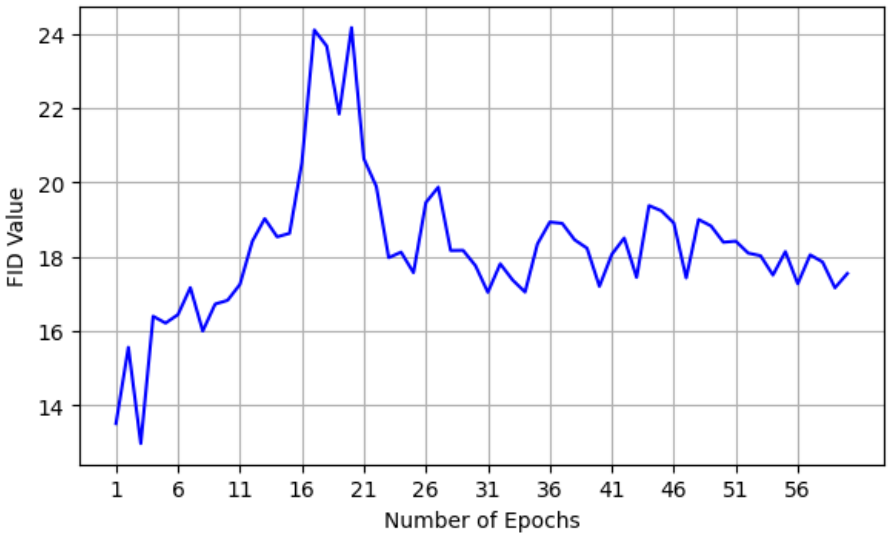


Figure 4.8: The FID value for the finger validation dataset reflects the model's performance, trained without finger images. Collected during the production phase from two devices, the training dataset includes 7840 images. It features a balanced amount of data for each organic specimen and a smaller quantity for potential transmission mediums such as air, glass, and surface temperature test phantoms(NPL). Average Fréchet Inception Distance of images from models trained on different datasets. The figure displays the model's validation results across various epochs, demonstrating performance consistency. Each validation set adheres to the same collection procedures and preprocessing methods as its training set.

Table 4.1: Performance of cGAN models trained with different data sources evaluated over human finger validation dataset

| Data Source | Model Name | FID | $SNR_{dB}$ | CNR |
|---|---|---|---|---|
| SingleDevice | cGAN | 29.67 | 32.47 | 4.10 |
| TwoDevice | cGAN | 17.04 | 35.38 | 6.20 |
| TwoDevice Cleaned | cGAN | 15.66 | 36.90 | **8.71** |
| TwoDevice Reduced | cGAN | 11.95 | **38.78** | 7.34 |
| 1x | N/A | 16.34 | 21.99 | 0.61 |
| 2x | N/A | 10.37 | 24.97 | 0.86 |
| 3x | N/A | 5.99 | 26.72 | 1.05 |
| 4x | N/A | 3.36 | 27.94 | 1.21 |
| 5x | N/A | 1.61 | 28.89 | 1.35 |
| 6x | N/A | 0.83 | 29.67 | 1.47 |
| 7x | N/A | **0.41** | 30.32 | 1.59 |

*Note:* This table compares the performance of models trained with datasets collected from two devices versus a single device during the production phase, featuring a balanced collection of specimens. The mean of evaluation metrics for the validation set are reported. The TwoDevice incorporates data from two devices, whereas the SingleDevice utilizes only one device. The 1x, serving as the baseline, employs one-time averaging without any data enhancement. Among the models trained with data from two devices, TwoDevice Cleaned omits outlier data types such as Glass, No Target (Air), and NPL. Conversely, the TwoDevice Reduced model is trained on a curated subset of 4410 images, ensuring an equal number of images across different tissue types, and still demonstrates favorable outcomes. This finding suggests that increasing tissue variety in the training dataset can significantly enhance the cGAN's performance. These metrics highlight the significance of data diversity in enhancing model performance. Region of interest (ROIs) for SNR and CNR are set using a bounding box approach based on regions categorized in Figure 4.1.

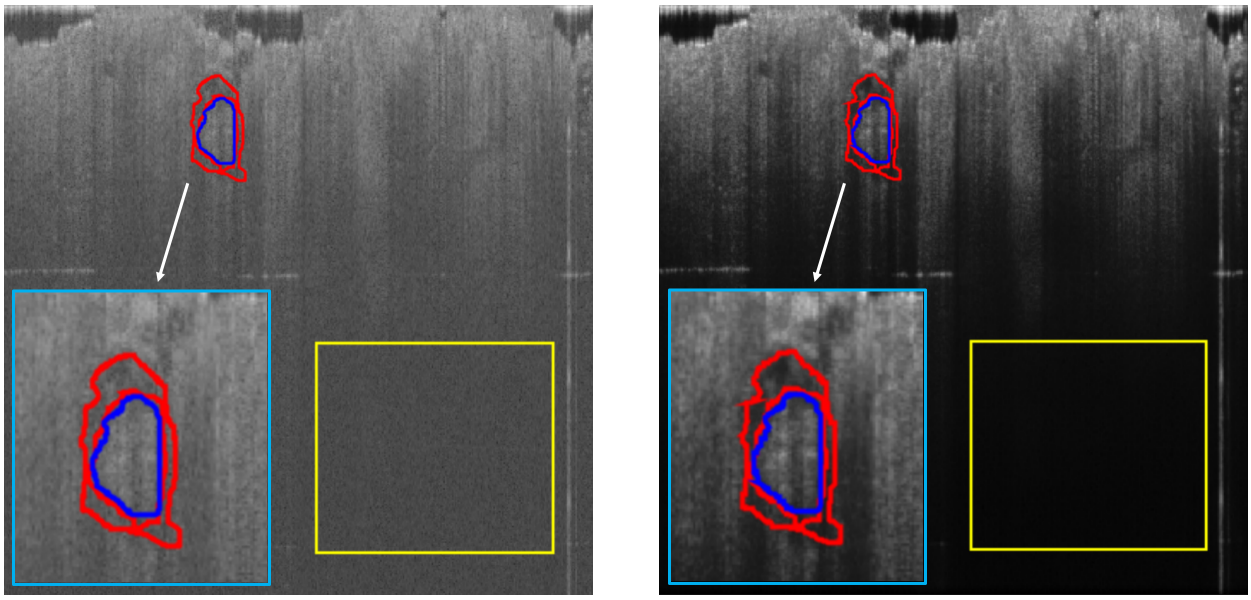

Figure 4.9: WF-OCT DCIS labels, from left to right: original DCIS and cGAN denoised image. Two contours in the images' upper part represent the two distinct regions within a DCIS. The inside core (blue contour) is the cancerous cells portion, while the outer rim (red contour) is the epithelial cell portion of a duct. Meanwhile, the rectangular area at the bottom represents the noise region.

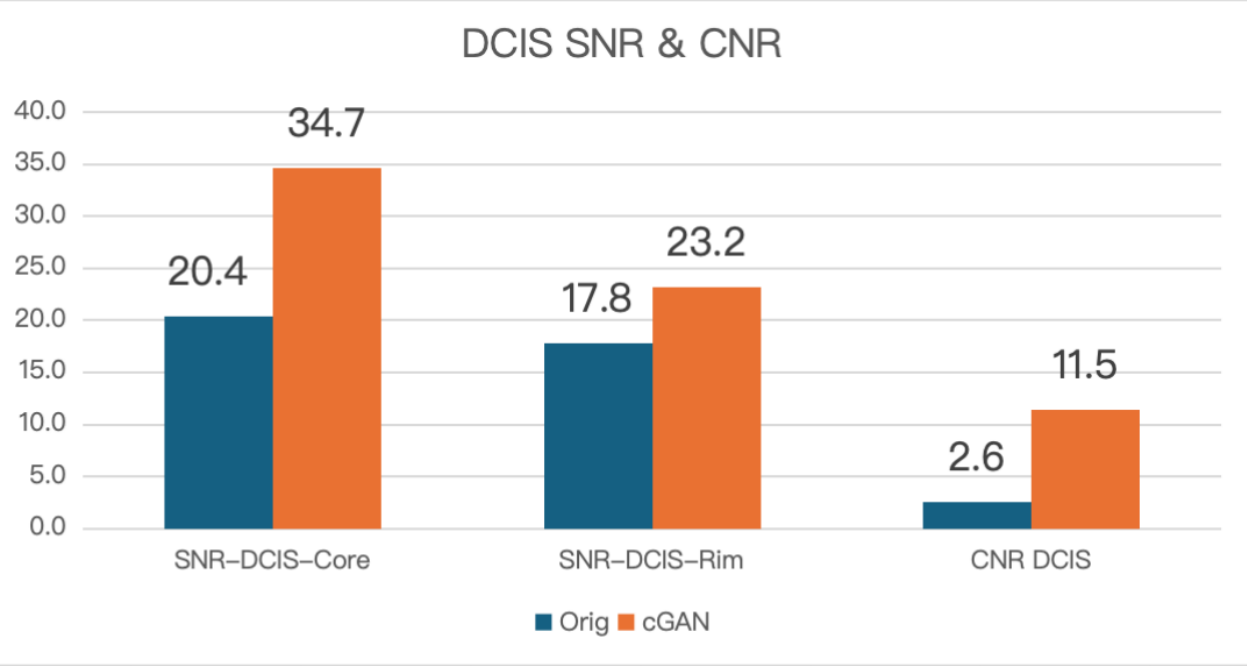


Figure 4.10: SNR and CNR analysis of the original image and cGAN denoised image, focusing on the core area (inside) and rim area (outside).

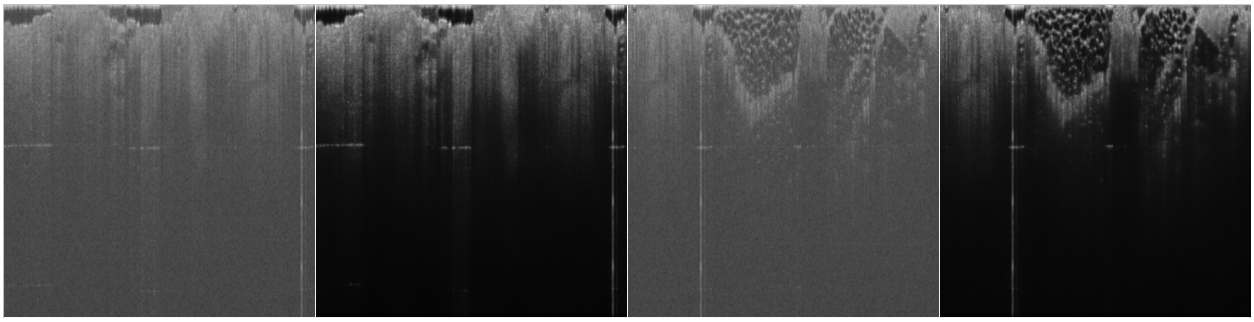

Figure 4.11: Comparison of WF-OCT DCIS versus deep learning reconstruction results, using the model 'TwoDevice Reduced' trained as described in the assessment table. From left to right: WF-OCT DCIS image, cGAN denoised 1x image, followed by another WF-OCT DCIS image and its corresponding cGAN denoised 1x image

# Chapter 5

# Result Analysis and Extension Studies

In OCT image denoising for breast cancer clinical image improvement, it is essential to reduce noise without losing crucial image features, such as the edges between tissues, adipose tissue structures, fibrous tissue structures, and other distinct structures. In this chapter, we will analyze the impact of the OCT cGAN denoising results on image quality assessment. Additionally, in the extension study, we will present methods that can be incorporated with the cGAN denoising model. These methods offer potential for various use cases and adaptability to complex real-life environments.

## 5.1 Pixel intensity profiles

An often-used analytical tool for medical imaging analysis, the intensity profile consists of a series of intensity values collected from evenly spaced points along a line or a series of lines within an image. This profile is pivotal for illustrating changes in intensity levels before and after the application of noise reduction techniques. Additionally, it enables the examination of image resolution or sharpness, often described through the line profile. In our research, despite the cGAN model being trained with an alternative processing approach, we specifically examine the Ductal Carcinoma In Situ (DCIS) and Invasive Ductal Carcinoma features (IDC) in the denoised images produced by the cGAN. In contrast to the standard collected 1x averaging, we do not have directly comparable DCIS/IDC 1x averaging images or 8x ground truth for clinical cases. Given the critical nature of these disease structures for surgical evaluation and the fact that cancerous tissues tend to infiltrate surrounding healthy

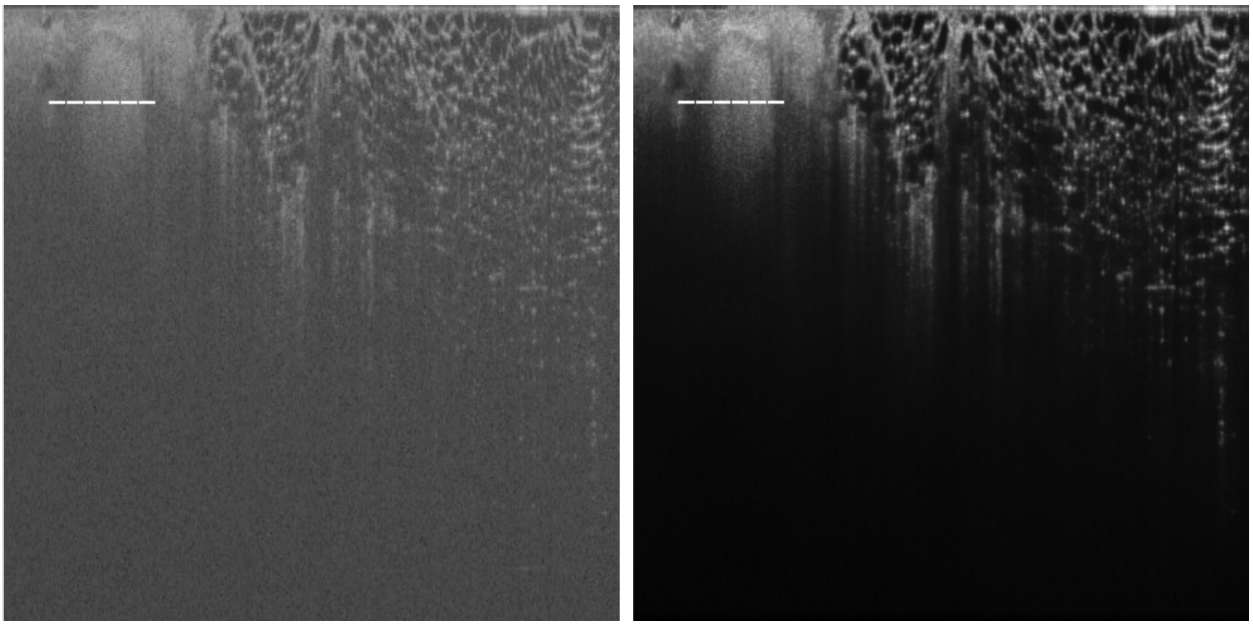

Figure 5.1: Comparison of WF-OCT DCIS and cGAN denoised result, emphasizing the horizontal DCIS line profile. This model version, referred to as 'TwoDevice Reduced' in Table 1 and trained with 4,410 images, has demonstrated a high ability to remove noise while preserving essential image features. The dashed white line indicates the line of analysis for pixel intensity.

tissues, the boundaries are more challenging to define. Thus, these tissues are more sensitive to the image quality, such as the noise level of the images. Cancerous tissues represent the most pertinent examples in our analysis.

Furthermore, we also calculated the first-order derivative of the corresponding line profiles to investigate the directional change in intensity of the deep learning-enhanced image. The gradient of the image, widely used in edge detection, highlights that an edge in an image may point in various directions. We primarily investigate the horizontal and vertical directions of the DCIS features or edges.

Examples of clinical DCIS identified by experts, along with denoised images from the 'Two Device Reduced' cGAN model presented in Table 4.1 and their labeled horizontal line profiles, are shown in Figure 5.1. Corresponding line profiles from the DCIS source image and the cGAN denoised image are provided in Figure 5.2. The first-order derivatives are displayed in Figure 5.3. Additionally, Figures 5.4, 5.5, and 5.6 present the vertical line profiles produced by the same mode.

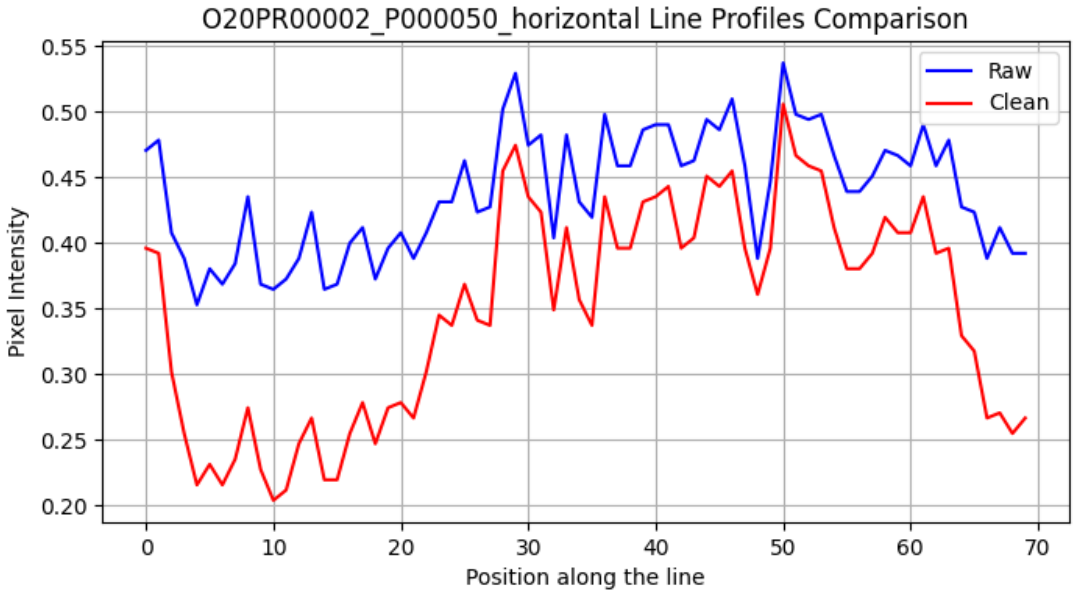


Figure 5.2: Comparison of pixel intensity values for WF-OCT DCIS versus cGAN denoised results along the horizontal line. The blue line represents the line profile of the source WF-OCT DCIS image, while the red line represents the line profile of the cGAN denoised image.

## 5.2 Perceptual loss functions in cGAN

In deep learning theory, the loss function not only serves as a method to evaluate how well an algorithm models the defined objectives through training on a dataset but also influences the direction in which a model converges through backpropagation. The loss function calculates the difference between the network's output and its expected output after a training example propagates through the network. Loss functions for image restoration tasks have been studied specifically to evaluate human visual quality, as a deep learning model's converged minimal point may not necessarily correspond to high-quality images from a human visual perspective, [JAFF16], [ZGFK16].

Unlike natural image translation tasks, the speckle noise reduction task in OCT presents unique challenges in finding optimal loss functions for the generator, as speckle noise does not form a uniform distribution across images. Different from the nature image translation task, this thesis explores different perceptual loss functions for OCT denoising, drawing on the original approaches for image restoration tasks described by Zhao et al. [ZGFK16].

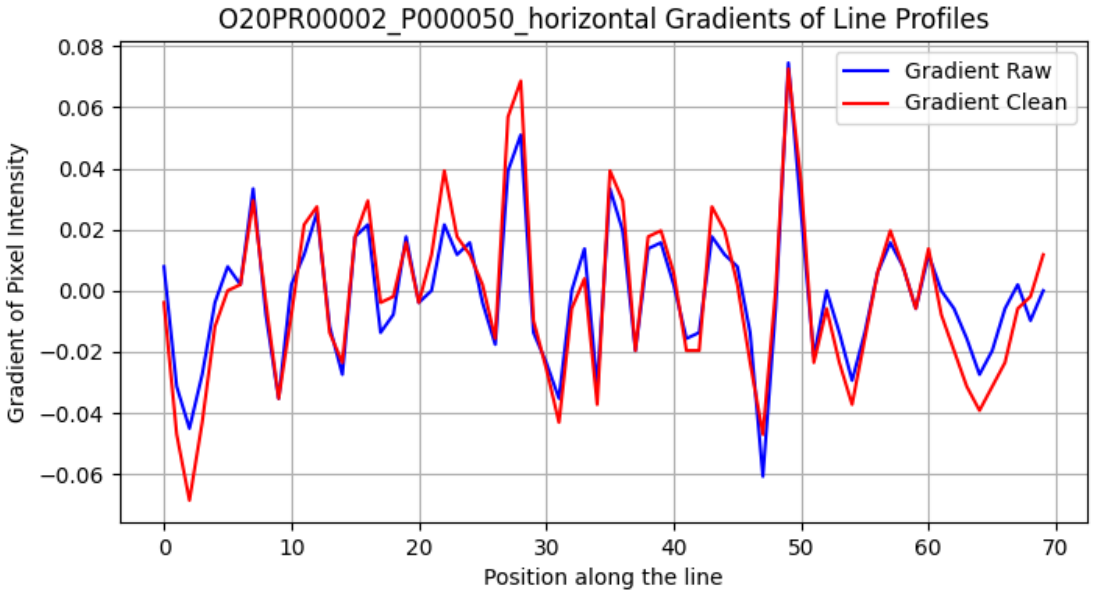


Figure 5.3: Comparison of first-order derivatives of pixel intensity values for WF-OCT DCIS versus cGAN denoised results along the horizontal line. The blue line represents the first-order derivative of the pixel intensity line profile for the source WF-OCT DCIS image, and the red line represents the same for the cGAN denoised image, indicating edge preservation.

### 5.2.1 Using L1 and L2 norms as loss function

L1 loss uses the absolute value of the difference between the predicted and actual values to measure the loss (or error) made by the model. L1 is described in Equation 5.1. L2 loss, which calculates the mean squared error, is sensitive to differences in the image due to the squared term. Thus, L2 loss can efficiently update the discriminator according to generated samples in LSGAN and CycleGAN [MLX+17], [ZPIE17]. It could potentially converge faster than L1. Due to the nature of L2, which penalizes large errors such as sharp changes in the images, it does so regardless of whether these are structures underlying the image. L2 is described in Equation 5.2. Thus, using L1 rather than L2 would result in sharper images in natural image translation tasks [ZPIE17]. In this work, we expand upon this idea to test L2 loss with the generator.

Where $x$ is the target image, $y$ is the generated image, and $N$ is the total number of pixels in the image.

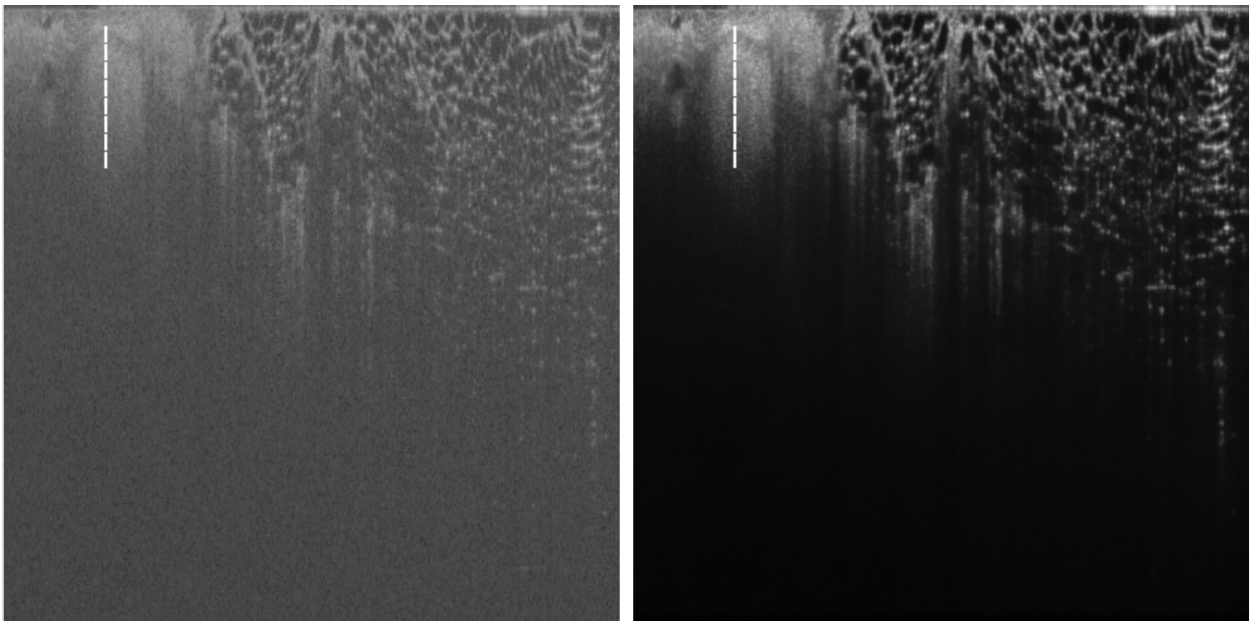

Figure 5.4: Comparison of WF-OCT DCIS and cGAN denoised result, emphasizing the vertical DCIS line profile. This model version, referred to as 'TwoDevice Reduced' in Table 1 and trained with 4,410 images, has demonstrated a high ability to remove noise while still preserving essential image features. The dashed white line indicates the line of analysis for pixel intensity.

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{5.1}$$

$$L_{L2} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{5.2}$$

### 5.2.2 Using SSIM and MS-SSIM as loss function

The SSIM loss function is structured to measure the perceptual difference between two images $x$ and $y$, focusing on structure similarity, contrast similarity, and luminance similarity. These include features such as step edges and speckles [ZGFK16]. The SSIM for a pixel $p$ from two

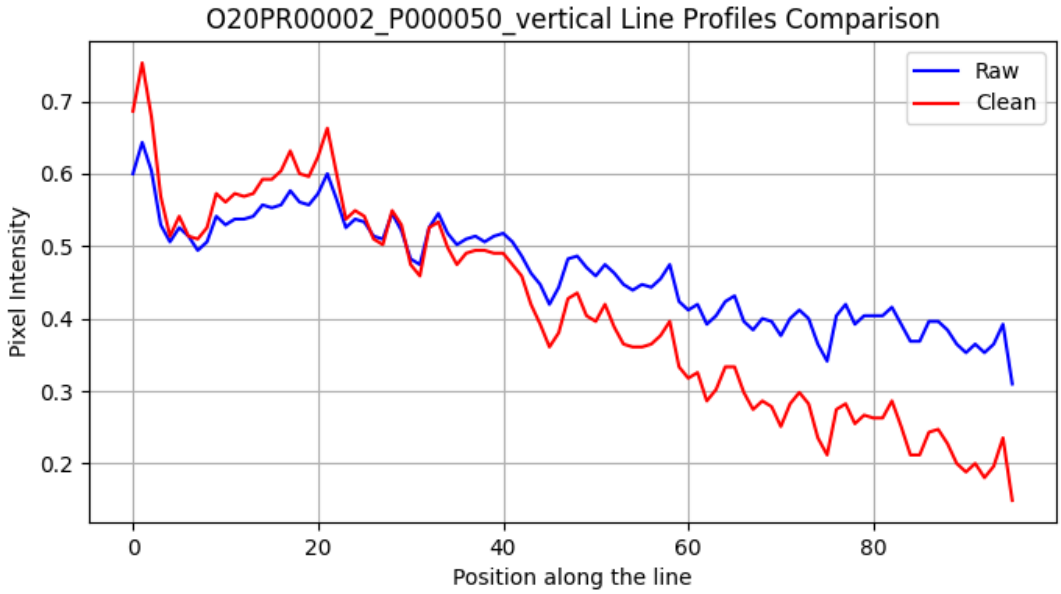


Figure 5.5: Comparison of pixel intensity values for WF-OCT DCIS versus cGAN denoised results along the vertical line. The blue line represents the line profile of the source WF-OCT DCIS image, while the red line represents the line profile of the cGAN denoised image.

images $x$ and $y$ is given by Equation 5.3.

$$\begin{aligned} SSIM(p) &= \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \\ &= l(x,y) \cdot cs(x,y) \\ &= l(p) \cdot cs(p) \end{aligned} \tag{5.3}$$

$\mu_x$ and $\mu_y$ are the average intensities, $\sigma_x$ and $\sigma_y$ are the variances, and $\sigma_{xy}$ is the covariance between $x$ and $y$. Constants $c_1$ and $c_2$ are small numbers added to stabilize the division. The term $l(x,y)$ represents the luminance comparison function, capturing the perceived brightness, while the product term, $l(x,y) \cdot cs(x,y)$, aggregates the contrast and structure measures.

Thus, the loss function for SSIM can be written as Equation 5.4.

$$L_{SSIM}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(1 - SSIM(p)) \tag{5.4}$$

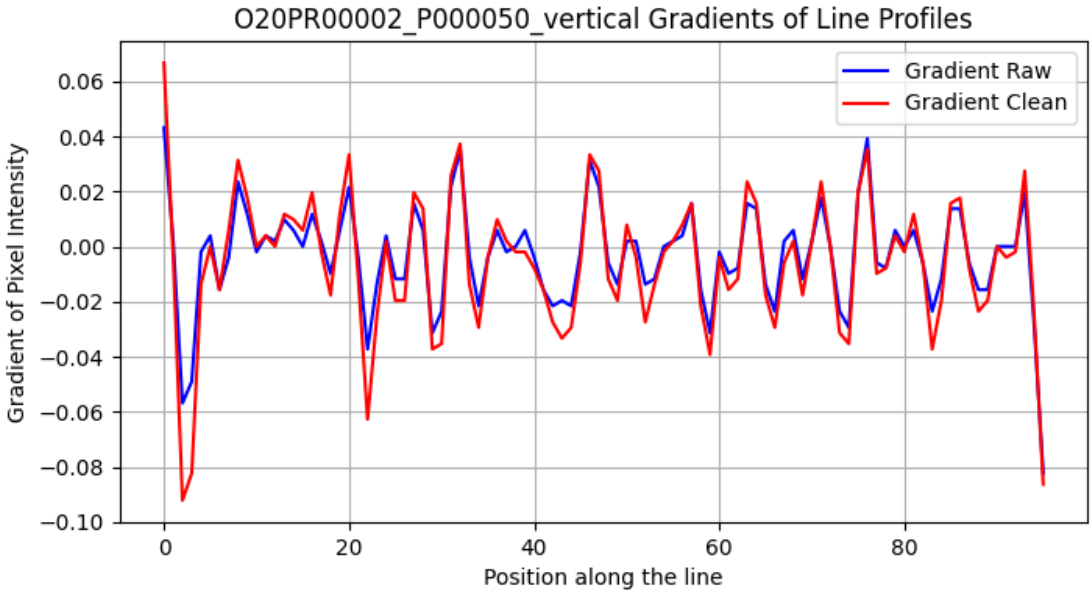


Figure 5.6: Comparison of first-order derivatives of pixel intensity values for WF-OCT DCIS versus cGAN denoised results along the vertical line. The blue line represents the first-order derivative of the pixel intensity line profile for the source WF-OCT DCIS image, and the red line represents the same for the cGAN denoised image, indicating edge preservation.

Similar to SSIM loss, the Multi-scale structure similarity index (MS-SSIM) is simply an SSIM applied at different pyramid scales to improve the evaluation of structure information under the human perception system [WSB03]. Equation 5.5 shows the MS-SSIM loss function.

$$L_{MS-SSIM}(x,y) = 1 - l_M(x,y)^{\alpha} \prod_{j=1}^{M} cs_j(x,y)^{\beta_j} \tag{5.5}$$

where $x$ and $y$ are the two images being compared. The term $l_M(x,y)$ represents the luminance comparison function at the $M$-th scale. The product term, $\prod_{j=1}^{M} cs_j(x,y)^{\beta_j}$, aggregates the contrast and structure comparison across $M$ scales, where $cs_j(x,y)$ combines both contrast and structural similarity at each scale. The exponents $\alpha$ and $\beta_j$ are parameters that adjust the relative importance of the luminance, contrast, and structure components. We assume that all three components are equally important in the denoising task; we set $\alpha$ and $\beta_j$ to 1 as suggested from the original paper [ZGFK16].

### 5.2.3 Using VGG loss as loss function

To utilize the prior knowledge in pre-trained models to get an estimation of image perception difference, we use the VGG loss from early activation layers of the pre-trained VGG network from pytorch implementations [PGM+19], [Alp21]. We experimented with feature reconstruction loss using a 16-layer VGG model (VGG16) and a 19-layer VGG model (VGG19). In VGG16 model, we use the first four blocks' Relu activation, whereas in VGG19, we use the first five blocks' Relu activation layer. We use L1 rather than L2 norms to compute the image differences as L1 loss shows sharpened output during the training of the OCT denoising task. The VGG loss function is defined in Equation 5.6 from the feature reconstruction loss function [JAFF16].

$$L_{\text{VGG}}(Y, \hat{Y}) = \sum_{l \in L} \frac{1}{N_l} \| F_l(Y) - F_l(\hat{Y}) \|_1 \tag{5.6}$$

Where:

- $Y$ is the target image.
- $\hat{Y}$ is the generated image.
- $L$ is the set of layers used for extracting features.
- $F_l$ represents the feature map extracted from layer $l$.
- $N_l$ is the number of elements in the feature maps from layer $l$.
- $\| \cdot \|_1$ denotes the sum of absolute differences.

### 5.2.4 Result for loss functions in cGAN

In this work, We try to combine perceptual loss with GAN loss. More specifically, we aimed to utilize a pre-trained VGG network to help the cGAN converge toward results that also consider human perception. The cGAN was experimented with by combining L1 or L2 loss with various types of perceptual loss, such as VGG loss, Multi-Scale Similarity (MS-SSIM) loss, or Structural Similarity (SSIM) loss, following the ratios suggested in image restoration tasks [ZGFK16]. More specifically, we have tested setting $\alpha$ to 0.84, a typical value used in

image restoration tasks [ZGFK16], or to 1, which eliminates the L1 or L2 terms in the loss function to explicitly observe the impact of perceptual losses. The overall equation is given in Equation 5.7. However, incorporating a perceptual loss term often results in artifacts in the referenced image. We believe such a loss function could help the cGAN converge more quickly, but further design and experimentation are required to find the optimal ratio for combining these loss functions. Since GANs are designed to find the Nash equilibrium between the generator and discriminator, this introduces increased complexity compared to a simple CNN.

Figure 5.9 shows an example of a cGAN denoised image using the generator's loss function. This function combines GAN loss and VGG16 loss, with 1x and 8x as the input pair.

$$L_{Mix_G} = l_{cGAN}(G(x,z),1.0) + \alpha \cdot L_{\text{perceptual loss}} + (1-\alpha) \cdot l_{\text{L1/L2}}(G(x,z),y) \tag{5.7}$$

## 5.3 Noise adding experiment

Noise is distributed into two basic modes: additive or multiplicative. Additive noise, being systematic, can be easily interpreted and modeled using statistical distributions, thus it can be reduced or removed straightforwardly. Multiplicative noise depends on pixel intensity and is image-dependent, making it difficult to modify [ESA12]. As we do not have direct information on the mean or variance of speckle noise in OCT images, it is challenging to derive a mathematical model for the speckle noise. Therefore, we simply use subtraction from the frame-vise averaging 8× image and the original 1× image to test model robustness. The equation is given in Equation 5.8, and visual results can be observed in Figures 5.7 and 5.8. Despite 25% and 50% noise being injected into the image, the cGAN model can still visually enhance the image.

$$\text{new image} = (\text{noisy\_image} - \text{clean\_image}) \times \text{noise percentage} + \text{noisy\_image} \tag{5.8}$$

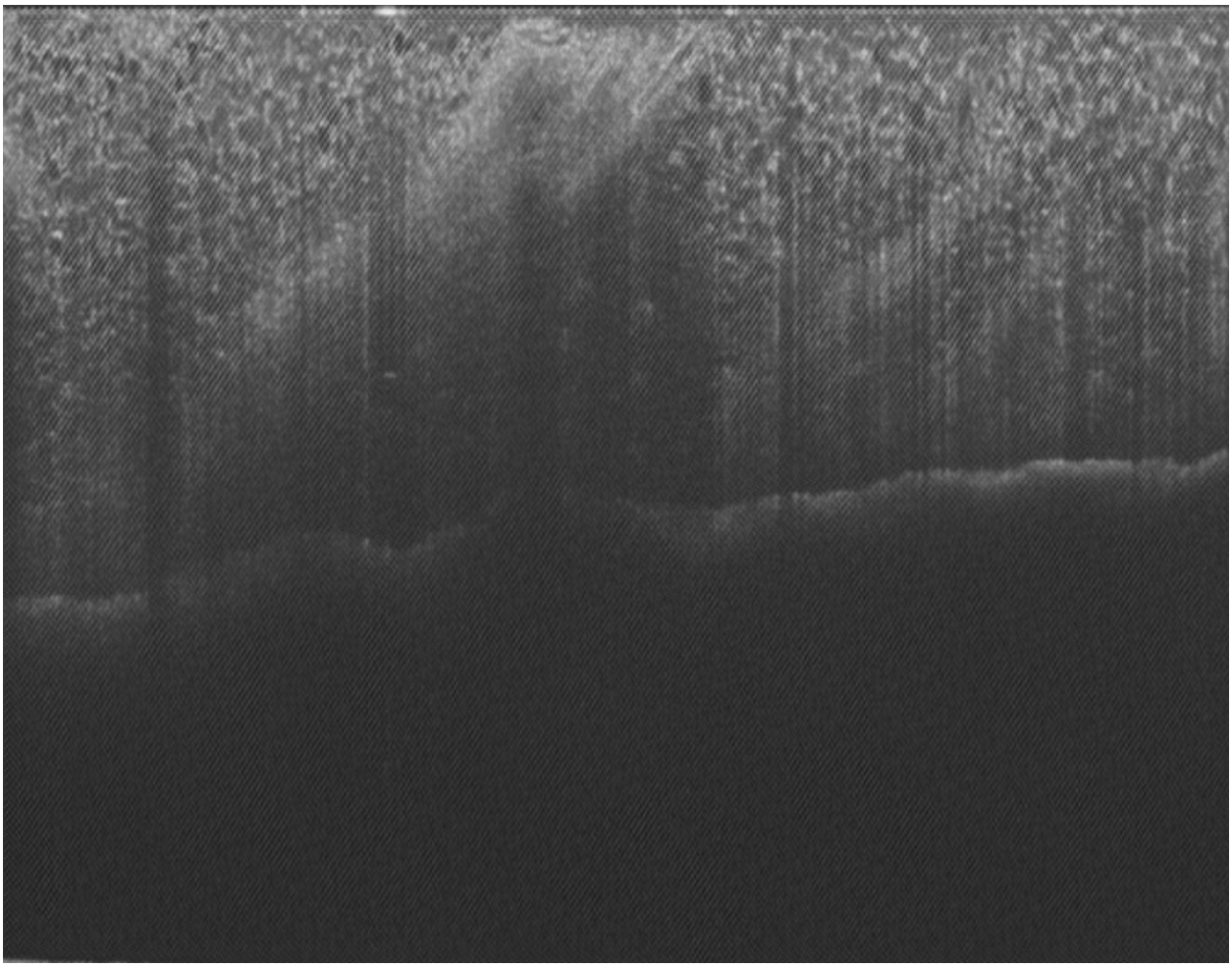

Figure 5.7: cGAN model trained with VGG16 loss showing the denoised result of a 1x OCT ginger image.

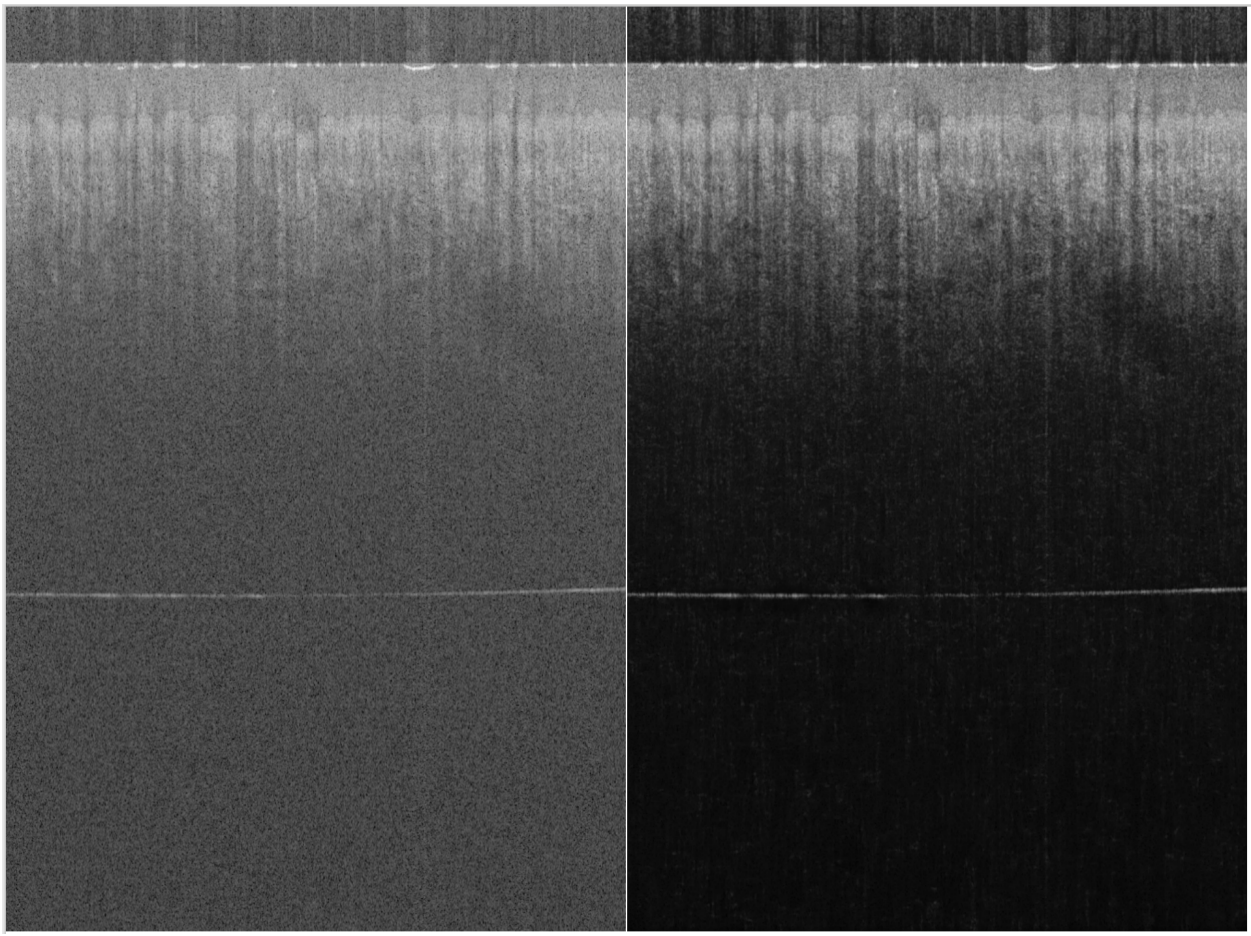

Figure 5.8: From left to right: 1x averaging image with 25 percent of noise injected into the image, cGAN model mentioned in Table 4.1 denoised result.

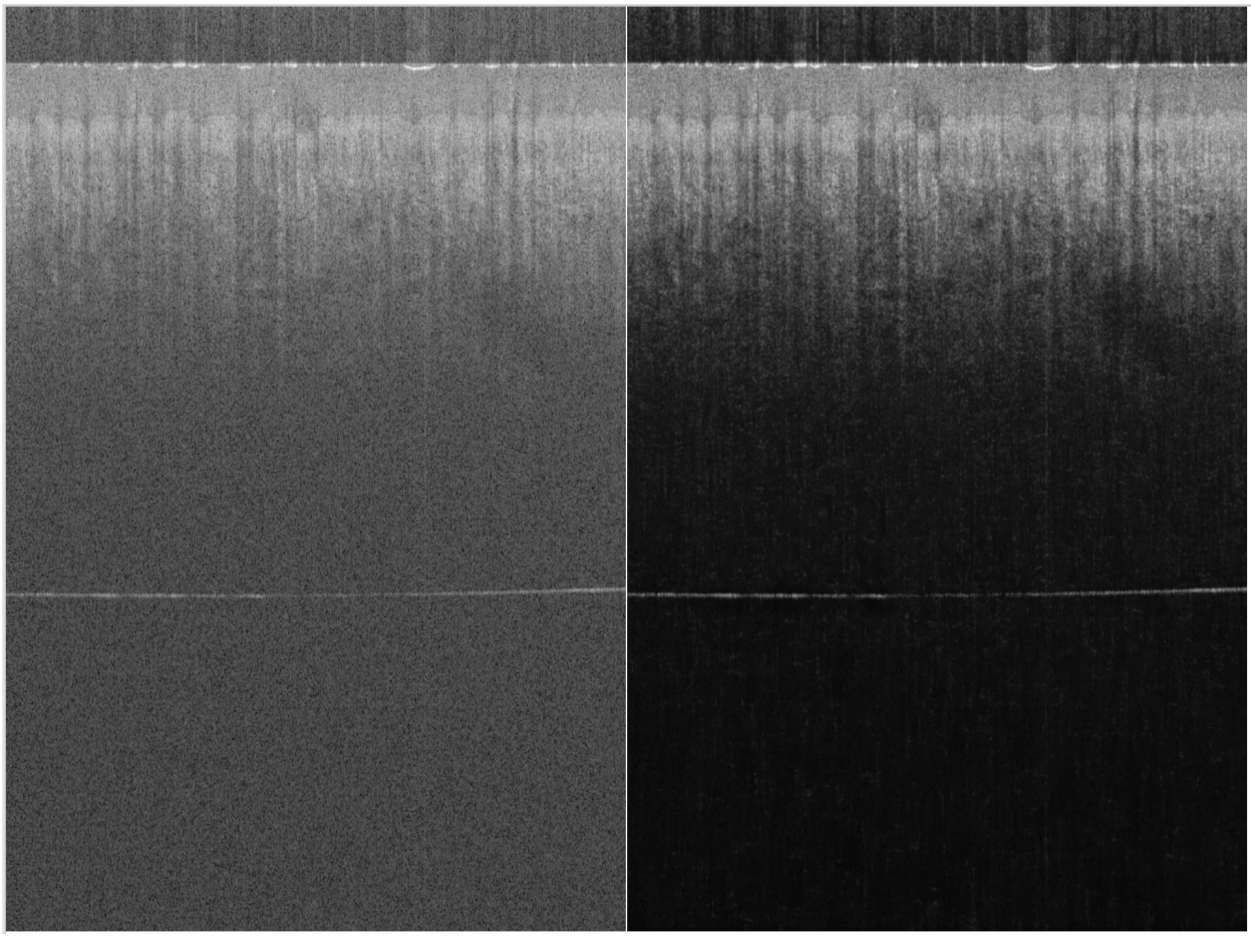

Figure 5.9: From left to right: 1x averaging image with 50 percent of noise injected into the image, cGAN model mentioned in Table 4.1 denoised result.

# Chapter 6

# Conclusion, Discussion, and Future Work

## 6.1 Conclusion

This work has explored deep learning-based OCT image enhancement to improve key image quality metrics, such as SNR and CNR. A cGAN approach was implemented with ResNet as a generator and PatchGAN as a discriminator. The proposed method was tested with OCT images acquired from breast cancer patients. The evaluations demonstrated that the proposed method significantly improved the image quality measured in terms of SNR and CNR. In validating our approach on a set of collected OCT images, our method consistently achieved higher SNR and CNR than the results from 1x, 2x, up to 7x averaging, suggesting it is a potential replacement for traditional frame averaging approaches.

## 6.2 Discussion and Limitation

The experimental results have shown that a pre-trained model could significantly facilitate model training, as evidenced by using FID scores. During the hyper-parameter search in the ablation study, as shown in Table 4.1, we experimented with both single-device and two-device dataset scenarios.

In addition, Table 4.1 lists different levels of conventional frame-wise averaging ap-

proaches. Sampling strategies for 2x, 3x, 4x, 5x, and 6x are employed because the pairing process could lead to very large validation sets, yet the results are relatively similar across these different sampling strategy-based sets. It is important to note that SNR and CNR are key evaluation metrics in the OCT denoising field. Other metrics such as FID, Root mean squared error, peak SNR, and SSIM are biased toward the target 8x 'noise-free' images, which favors the frame averaging approach. These evaluations can improve fairness by collecting more simultaneous scans to generate higher-level frame-averaging images that at least match the SNR or CNR levels of our deep learning-based method.

Also, we do not have a 'frame averaging generated cleaned ground truth' from the actual clinical cases used in our test set; thus, we evaluated only by SNR and CNR in conventional processed clinical tests. This can be improved by collecting actual 1x breast cancer scans. Additionally, while we have multiple conventional clinical cases for testing model performance, as shown in Figure 4.9, we do not have clinically defined regions for every case to serve as ground truths that require manual labeling. Therefore, we only provide a typical labeled DCIS case as an example to calculate CNR and SNR.

Our cGAN model has 11.4 million parameters. Because on-device evaluation requires a lot of extra engineering work to draw more precise conclusions, for example, different samples might have different image sizes that complicate the test. The cGAN model has been evaluated locally on a GPU Nvidia RTX 3070 with an average inference time of 157.990 ms using Python, loading one large image at a time. This time could be further improved by loading multiple images in real time and applying other engineering optimizations, such as quantization and parallel computation.

## 6.3 Future work

With the increasing availability of data and the continual development of applications in medical AI, future studies in healthcare AI are expected to bring substantial benefits. Our work has transferability potential and can be extended to different tasks or embedded as part of a system. With appropriate parameter settings and considering clinical demands, our method can be applied to other modalities suffering from noise, such as low-dose CT, synthetic-aperture radar, and ultrasound imaging. In the real clinical application of the cGAN network, only the generator part would be deployed. To shorten the model inference time, the generator could be altered to a more lightweight model depending on specific task

requirements, such as using fewer residual blocks in the ResNet. Or, only selected frames rather than all frames during the surgical operations would be applied since the model takes frames as input rather than being built as a 3D model that requires sequence information. Furthermore, our work can systematically generate new data, enhance clinical training, and serve as the foundation for other artificial intelligence (AI) systems, such as classification and segmentation tasks. The application and research of AI technologies continue to progress, fueled by hardware advancements that provide more computational resources available for training and inference. This progress would make more problems approachable and open new avenues for future innovation.

# References


[AASA22] Waqar Ahmad, Hazrat Ali, Zubair Shah, and Shoaib Azmat. A new generative adversarial network for medical images super resolution. *Scientific Reports*, 12(1):9533, 2022.

[ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[ADFM19] Silke Aumann, Sabine Donner, Jörg Fischer, and Frank Müller. Optical coherence tomography (oct): principle and technical realization. *High resolution imaging in microscopy and ophthalmology: new frontiers in biomedical optics*, pages 59–85, 2019.

[Alp21] Alper. Vgg perceptual loss, 2021. Accessed on 05 06, 2024.

[AM22] Evelina Arzanova and Harvey N Mayrovitz. The epidemiology of breast cancer. *Exon Publications*, pages 1–19, 2022.

[BBR+09] J Quincy Brown, Torre M Bydlon, Lisa M Richards, Bing Yu, Stephanie A Kennedy, Joseph Geradts, Lee G Wilke, Marlee K Junker, Jennifer Gallagher, William T Barry, et al. Optical assesssment of tumor resection margins in the breast. *IEEE Journal of selected topics in Quantum Electronics*, 16(3):530–544, 2009.

[Bie20] Lukas Biewald. Experiment tracking with weights and biases, 2020. Software available from wandb.com.

[BSL+23] Arvind K Badhey, Julia S Schwarz, Benjamin M Laitman, Brandon M Veremis, William H Westra, Mike Yao, Marita S Teng, Eric M Genden, and Brett A

Miles. Intraoperative use of wide-field optical coherence tomography to evaluate tissue microstructure in the oral cavity and oropharynx. *JAMA Otolaryngology–Head & Neck Surgery*, 149(1):71–78, 2023.

[BW12] Archie Bleyer and H Gilbert Welch. Effect of three decades of screening mammography on breast-cancer incidence. *New England Journal of Medicine*, 367(21):1998–2005, 2012.

[CJ19] Hyunho Choi and Jechang Jeong. Speckle noise reduction technique for sar images using statistical characteristics of speckle noise and discrete wavelet transform. *Remote Sensing*, 11(10):1184, 2019.

[ESA12] Shaimaa A El-Said and Ahmad Taher Azar. Speckles suppression techniques for ultrasound images. *Journal of medical imaging and radiation sciences*, 43(4):200–213, 2012.

[FPBB00] James G Fujimoto, Costas Pitris, Stephen A Boppart, and Mark E Brezinski. Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. *Neoplasia*, 2(1-2):9–25, 2000.

[GB10] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[Goo76] Joseph W Goodman. Some fundamental properties of speckle. *JOSA*, 66(11):1145–1150, 1976.

[GPAM+14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[GW16] Sam Gross and Michael Wilber. Training and investigating residual nets. http://torch.ch/blog/2016/02/04/resnets.html, 2016. Accessed: 2024-04-18.

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[HJKSN17] Olaf Johan Hartmann-Johnsen, Rolf Kåresen, Ellen Schlichting, and Jan F Nygård. Better survival after breast-conserving therapy compared to mastectomy when axillary node status is positive in early-stage breast cancer: a registry-based follow-up study of 6387 norwegian women participating in screening, primarily operated between 1998 and 2009. *World Journal of Surgical Oncology*, 15:1–10, 2017.

[HMLMM14] Nehmat Houssami, Petra Macaskill, M Luke Marinovich, and Monica Morrow. The association of surgical margins and local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy: a meta-analysis. *Annals of surgical oncology*, 21:717–730, 2014.

[HRU+17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[IZZE17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[JAFF16] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.

[JO19] Rebecca M Jordan and Jacqueline Oxenberg. Breast cancer conservation therapy. 2019.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KFH21] Stephen Keelan, Michael Flanagan, and Arnold DK Hill. Evolving trends in surgical management of breast cancer: an analysis of 30 years of practice changing papers. *Frontiers in Oncology*, 11:622621, 2021.

[KW13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[ŁCF+21] Sergiusz Łukasiewicz, Marcin Czeczelewski, Alicja Forma, Jacek Baj, Robert Sitarz, and Andrzej Stanisławek. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. *Cancers*, 13(17):4287, 2021.

[LLS+17] Orly Liba, Matthew D Lew, Elliott D SoRelle, Rebecca Dutta, Debasish Sen, Darius M Moshfeghi, Steven Chu, and Adam de La Zerda. Speckle-modulating optical coherence tomography in living mice and humans. *Nature communications*, 8(1):15845, 2017.

[LRN+23] Yanir Levy, David Rempel, Mark Nguyen, Ali Yassine, Maggie Sanati-Burns, Payal Salgia, Bryant Lim, Sarah L Butler, Andrew Berkeley, and Ersin Bayram. The fusion of wide field optical coherence tomography and ai: Advancing breast cancer surgical margin visualization. *Life*, 13(12):2340, 2023.

[LTH+17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[Mas12] Shahla Masood. Expanded role of cytopathology in breast cancer diagnosis, therapy and research: the impact of fine needle aspiration biopsy and imprint cytology, 2012.

[MCZ+18] Yuhui Ma, Xinjian Chen, Weifang Zhu, Xuena Cheng, Dehui Xiang, and Fei Shi. Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan. *Biomedical optics express*, 9(11):5129–5146, 2018.

[MLX+17] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[MMX+21] Maryam Mehdizadeh, Cara MacNish, Di Xiao, David Alonso-Caneiro, Jason Kugelman, and Mohammed Bennamoun. Deep feature loss to denoise oct images using deep neural networks. *Journal of Biomedical Optics*, 26(4):046003–046003, 2021.

[MO14] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[OTC+18] Ryan K Orosco, Viridiana J Tapia, Joseph A Califano, Bryan Clary, Ezra EW Cohen, Christopher Kane, Scott M Lippman, Karen Messer, Alfredo Molinolo, James D Murphy, et al. Positive surgical margins in the 10 most common solid cancers. *Scientific reports*, 8(1):5686, 2018.

[PGM+19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[PN21] S Pradeep and P Nirmaladevi. A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and cnn methods. In *IOP conference series: materials science and engineering*, volume 1055, page 012116. IOP Publishing, 2021.

[QHL+20] Bin Qiu, Zhiyu Huang, Xi Liu, Xiangxi Meng, Yunfei You, Gangjun Liu, Kun Yang, Andreas Maier, Qiushi Ren, and Yanye Lu. Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function. *Biomedical optics express*, 11(2):817–830, 2020.

[RBDS+22] David Rempel, Andrew Berkeley, Allison A DiPasquale Sr, Maryam Elmi, Richard E Fine, Marie C Lee, Bridget O'Brien, Lee Gravatt Wilke, and Alastair M Thompson. A prospective, multicenter, randomized, double-arm trial to determine the impact of the perimeter b-series optical coherence tomography and artificial intelligence system on positive margin rates in breast conservation surgery. *Journal of the American College of Surgeons*, 235(5):S4, 2022.

[RC23] Beryl Rabindran and Adriana D Corben. Wide-field optical coherence tomography for microstructural analysis of key tissue types: a proof-of-concept evaluation. *Pathology and Oncology Research*, 29:1611167, 2023.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[RJHT+22] Jose J Rico-Jimenez, Dewei Hu, Eric M Tang, Ipek Oguz, and Yuankai K Tao. Real-time oct image denoising using a self-fusion neural network. *Biomedical Optics Express*, 13(3):1398–1409, 2022.

[RMC15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[SADJK+22] Vania B Silva, Danilo Andrade De Jesus, Stefan Klein, Theo van Walsum, João Cardoso, Luisa Sánchez Brea, and Pedro G Vaz. Signal-carrying speckle in optical coherence tomography: a methodological review on biomedical applications. *Journal of Biomedical Optics*, 27(3):030901–030901, 2022.

[SCJ+20] Hank Schmidt, Courtney Connolly, Shabnam Jaffer, Twisha Oza, Christina R Weltz, Elisa R Port, and Adriana Corben. Evaluation of surgically excised breast tissue microstructure using wide-field optical coherence tomography. *The Breast Journal*, 26(5):917–923, 2020.

[SGJ24] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1):12–49, 2024.

[SJL23] Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. Gans for medical image synthesis: An empirical study. *Journal of Imaging*, 9(3):69, 2023.

[SMWJ23] Rebecca L. Siegel, Kimberly D. Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *Ca Cancer J Clin*, 73(1):17–48, 2023.

[SRE12] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

[SVI+16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[SWP+23] Julia Schottenhamml, Tobias Würfl, Stefan B Ploner, Lennart Husvogt, Bettina Hohberger, James G Fujimoto, and Andreas Maier. Ssn2v: unsupervised oct denoising using speckle split. *Scientific Reports*, 13(1):10382, 2023.

[SXY99] Joseph M Schmitt, SH Xiang, and Kin Man Yung. Speckle in optical coherence tomography. *Journal of biomedical optics*, 4(1):95–105, 1999.

[VFJB12] Benjamin J Vakoc, Dai Fukumura, Rakesh K Jain, and Brett E Bouma. Cancer imaging by optical coherence tomography: preclinical progress and clinical potential. *Nature Reviews Cancer*, 12(5):363–368, 2012.

[WLVI17] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, 36(12):2536–2545, 2017.

[WSB03] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[YZR+19] Kiran S Yemul, Adam M Zysk, Andrea L Richardson, Krishnarao V Tangella, and Lisa K Jacobs. Interpretation of optical coherence tomography images for breast tissue assessment. *Surgical innovation*, 26(1):50–56, 2019.

[ZGFK16] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

[ZPIE17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.