Domain adaptation for ultrasound image quality enhancement with deep learning

Final Report

Autumn 2024-25

Imane Jennane

310900

Supervisor: Roser Viñals

Professor: Jean-Philippe Thiran

December 2024

Abstract

Ultrafast ultrasound imaging (UFUS) offers remarkable potential with its high frame rates, enabling advanced medical applications such as cardiac imaging for real-time assessment of heart function and in liver imaging. However, it comes with a trade-off: reduced image quality caused by noise, artifacts, and lower resolution. This project explores how deep learning can bridge that gap, focusing on adapting neural networks to work across different imaging conditions and datasets.

We investigate strategies like transfer learning and fine-tuning to make pre-trained models more versatile and effective when applied to images captured with different ultrasound probes. Among the strategies, alternating with freezing and training with a step of 4 layers was the most effective, improving image clarity by 11.69% in PSNR and 52.32% in SSIM compared to baseline models, while preserving critical details essential for diagnostic accuracy.

This work highlights how adapting tools like deep learning can make ultrafast ultrasound imaging more impactful, paving the way for better diagnostics in areas like liver imaging.

Contents

Chapter 1

Introduction

Ultrasound imaging is a widely used medical imaging technique known for its safety, real-time capabilities, and cost-effectiveness compared to modalities such as CT or MRI. Traditional ultrasound systems use focused beams to scan tissues line by line, producing high-quality images but with limitations in frame rate and penetration depth.

Ultrafast ultrasound (UFUS) imaging has emerged as an alternative, achieving much higher frame rates by emitting unfocused waves, such as plane waves. This innovation enables advanced imaging applications, including shear wave elastography and functional ultrasound imaging. However, the use of unfocused waves leads to a trade-off: while frame rates increase, image quality decreases due to lower resolution, reduced contrast, and artifacts.

Enhancing the quality of ultrafast ultrasound images is crucial to unlocking the full potential of this technique for medical diagnostics, particularly in areas like liver imaging. The challenge lies in improving image clarity without compromising the benefits of UFUS. This project addresses this challenge by leveraging CNN-based image reconstruction methods and exploring advanced strategies such as transfer learning and fine-tuning. These approaches aim to adapt neural networks to diverse imaging conditions, ultimately enabling more reliable and impactful ultrafast ultrasound imaging.

1.1 Literature Review

Deep learning has been extensively applied to improve ultrasound image quality, especially in ultrafast imaging, where traditional methods often struggle with noise and artifacts. Convolutional neural networks (CNNs) have demonstrated their ability to map low-quality images to higher-quality ones. For instance, Perdios et al. [4] used a U-Net-based CNN to reconstruct high-quality images from single plane-wave (PW) acquisitions, significantly reducing artifacts while preserving critical speckle patterns. Their work quantitatively demonstrated that the proposed method does not harm the time-coherence of consecutive frames, making it suitable for advanced imaging modalities like vector flow imaging (VFI). However, their reliance on simulated datasets limits the applicability of this method to real-world scenarios.

To address this limitation, Viñals and Thiran [2] proposed a novel loss function based on the Kullback-Leibler divergence, specifically designed for in vivo ultrafast ultrasound imaging. This loss function accounts for the wide intensity range of ultrasound images, enabling better preservation of brightness distributions. Their approach, evaluated on a large dataset of in vivo images, achieved significant quality improvements over earlier methods, particularly in preserving fine image details.

Transfer learning offers a compelling solution for bridging the gap between simulated and real-world datasets. Viñals, Beuret, and Thiran [3] demonstrated that fine-tuning pre-trained CNNs on smaller, real-world datasets leads to superior performance compared to training on either simulated or real data alone. Their method successfully improved ultrafast ultrasound image quality across varied datasets, showcasing the potential of transfer learning to adapt models across domains with minimal data requirements.

Building on these studies, this project aims to combine transfer learning and domain adaptation techniques to enhance image quality across diverse ultrafast ultrasound imaging settings. By addressing dataset variability and optimizing neural network performance, this work seeks to make ultrafast ultrasound a more versatile tool for clinical applications.

Chapter 2

Methodology

2.1 Data-set Acquisition

In this study, two distinct datasets were used, each containing 20,000 images. The datasets were collected with ethical approval from the Cantonal Commission on Ethics in Human Research (CER-VD, Vaud, Switzerland).

The first dataset was acquired from nine healthy volunteers (five males and four females) aged between 22 and 33 years. To capture these images, the GE 9L-D linear array transducer was used. This probe operates at a center frequency of 5.3 MHz, with a bandwidth of 75% at -6 dB. It has an aperture size of 43.93 mm and comprises 192 elements.

The second dataset was acquired using the GE 11L-D linear array transducer and was obtained from 16 healthy volunteers aged between 20 and 45 years. This probe operates at a higher center frequency of 7.3 MHz, with a bandwidth of 75% at -6 dB. It has an aperture size of 38.20 mm and also comprises 192 elements.

These probes provided a range of frequency and aperture characteristics that were well-suited to the requirements of the imaging tasks conducted in this project.

Two types of radiofrequency (RF) images were derived from each acquisition:

1. **Input image:** A single unfocused RF image acquired at a steering angle of 0°, representing a low-quality baseline.

2. **Target image:** A CPWC image generated by coherently combining 87 steered RF measurements. This image serves as the high-quality reference for evaluating model performance.

The use of in vivo datasets, as described in the KLD-MSLAE study [1], highlights the importance of adapting deep learning methods to realistic and diverse clinical conditions. The target images, produced through CPWC, approximate the quality needed for accurate diagnostics, while the input images emulate the challenges of single-plane wave acquisitions.

2.2 U-Net Architecture and Splitting strategy

2.2.1 U-Net

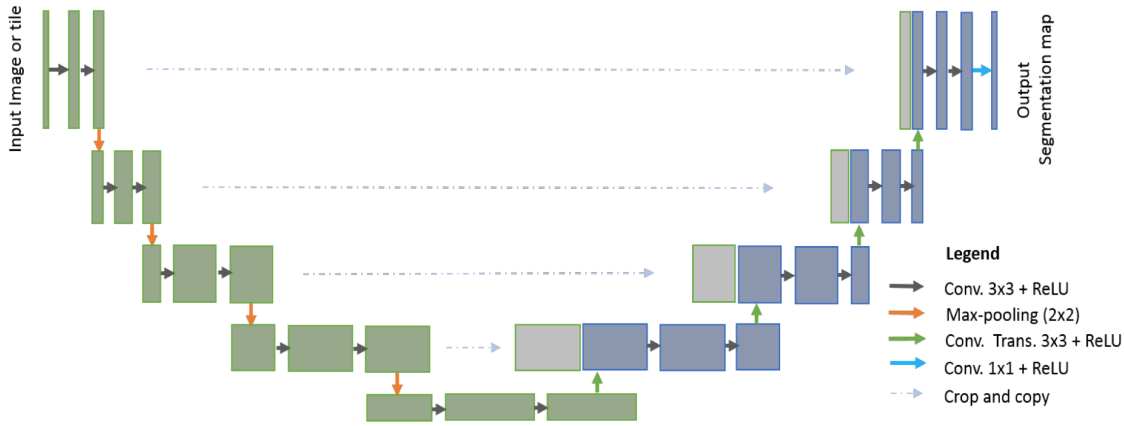

Figure 2.1: U-net Architecture.

The U-Net architecture, shown in Figure 2.1, was first introduced in the 2015 paper "*U-Net: Convolutional Networks for Biomedical Image Segmentation*" by Ronneberger, Fischer, and Brox. It was specifically designed for biomedical applications, such as segmenting cells or tissues in microscopy images. Over time, its flexibility and effectiveness have made it a popular choice for a wide range of image processing tasks, including image enhancement. This is particularly relevant for ultrasound imaging, where challenges like reducing noise and preserving critical details in low-quality scans are common.

As illustrated in the figure 2.1, the U-Net follows a straightforward yet powerful encoder-decoder structure. The encoder on the left progressively compresses the input image to extract essential features, while the decoder on the right reconstructs the image to its original resolution. A defining feature of U-Net is its skip connections, represented by the dashed arrows, which link the encoder and decoder layers at corresponding levels. These connections allow the model to retain intricate details that might otherwise be lost during compression, while also maintaining a broader contextual understanding of the image.

For this project, U-Net is used to enhance low-quality ultrasound images by improving resolution, boosting contrast, and minimizing artifacts. Its multi-scale processing, evident in the hierarchical structure of Figure 2.1, enables it to balance fine details with the overall integrity of the image. The result is not only a visually sharper output but also one that provides more accurate and reliable data for further analysis or clinical application.

2.2.2 Splitting Strategy

In order to train and evaluate our models (using U-Net or any other deep learning architecture), we adopted a **volunteer-based splitting strategy (VS)**. This means that data from each volunteer is included in exactly one subset: training, validation, or testing. By preventing overlap of volunteers across these subsets, we reduce the risk of information leakage and more closely simulate real-world usage, where newly acquired data typically comes from individuals not represented in the training set.

Implementation Details

1. **Identify Volunteers**: We compiled a complete list of all volunteers in our dataset, each providing a number of images or scans.
2. **Assign Subsets**: We allocated each volunteer exclusively to one of the following sets: training, validation, or test.
3. **Balance and Percentages**: We aimed to maintain representative proportions for each subset. Details about the exact splits used for the *9LD* and *11LD* models are explicited below.

Model	Subset (%)	Volunteers
9LD	Training (67%)	1, 2, 3, 6, 7, 9
	Validation (11%)	4
	Testing (22%)	8, 5
11LD	Training (69%)	3, 6, 7, 9, 10, 11, 13, 15, 16, 18, 19
	Validation (13%)	12, 14
	Testing (18%)	8, 5, 17

Table 2.1: Volunteer-based splitting strategy for the 9LD and 11LD models.

Benefits and Considerations

- **Consistency**: Employing the same splitting strategy for both models allows direct performance comparisons.
- **Avoiding Bias**: Randomizing volunteer IDs prior to splitting helps mitigate systematic biases.

Note: While we have chosen a volunteer-based strategy here, other approaches can also be used. For instance, one might split the data according to the scanned **body area**, ensuring that specific body regions are consistently grouped for training, validation, or testing. In our case, we rely on the volunteer-based split because it is already well-balanced in terms of body-area coverage. However, if the data distribution were not well-balanced, employing a different or more refined strategy (e.g., body-area-based or stratified splits) might be preferable.

2.3 Measurement tools

2.3.1 Performance metrics

To evaluate the performance of the proposed CNN models across different datasets, we use two widely adopted image quality metrics: ***Peak Signal-to-Noise Ratio (PSNR)*** and ***Structural Similarity Index Measure (SSIM)***.

PSNR is a metric that quantifies the reconstruction quality of an image based on the logarithmic ratio between the maximum possible pixel value and the mean squared error (MSE) between the reference and reconstructed images. Higher PSNR values indicate better fidelity to the reference image. PSNR is particularly useful in this context as it provides a straightforward numerical comparison of pixel-wise accuracy, which is essential for assessing noise reduction and resolution enhancement.

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}}{\sqrt{\text{MSE}}}\right)$$

where:

$$\begin{aligned}
\text{MSE} &= \frac{1}{N}\sum_{i=1}^{N}\left(I_1(i,j) - I_2(i,j)\right)^2, \\
I_1, I_2 &: \text{B-mode target image and output image}, \\
\text{MAX} &: \text{Maximum possible pixel value in the range -40 and 40 dB.}
\end{aligned}$$

SSIM, on the other hand, measures image quality by considering structural information, luminance, and contrast. Unlike PSNR, which focuses purely on pixel-level differences, SSIM aligns more closely with human visual perception. This makes it a complementary metric, especially valuable for evaluating the perceptual quality of images. It is sensitive to the preservation of edges and textures, which are critical in medical imaging tasks like ultrasound image reconstruction.

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where:

$$\begin{aligned}
x, y &: \text{The predicted and target images being compared}, \\
\mu_x, \mu_y &: \text{Mean intensity of images } x \text{ and } y, \\
\sigma_x^2, \sigma_y^2 &: \text{Variance of images } x \text{ and } y, \\
\sigma_{xy} &: \text{Covariance of images } x \text{ and } y, \\
C_1 = (K_1 \cdot L)^2 &\quad \text{and} \quad C_2 = (K_2 \cdot L)^2, \\
K_1, K_2 &: \text{Stabilization constants (e.g., 0.01 and 0.03)}, \\
L &: \text{Dynamic range of pixel values (-40 to 40 dB).}
\end{aligned}$$

While PSNR and SSIM are commonly used and considered well-suited for analyzing the performance of CNNs in enhancing ultrasound image quality across different datasets, they have limitations. PSNR can fail to reflect perceptual quality, while SSIM, though more aligned with human perception, may not capture certain artifacts or subtle visual details.

2.3.2 Kullback-Leibler Divergence loss

The Kullback-Leibler Divergence (KLD) loss, as introduced by R.Viñals and J-P.Thiran (2023) [2], is a powerful tool for enhancing ultrafast ultrasound images by addressing the challenges of high dynamic range and distribution alignment in echogenicity values. This loss function combines the mean signed logarithmic absolute error (MSLAE) with the KLD to preserve the probability distributions of echogenicity values, ensuring that the enhanced images closely resemble the statistical properties of high-quality target images.

In this project, the KLD component penalizes deviations between the probability distributions of the predicted and target images, reducing common issues such as changes in the shape and range of echogenicity distributions found with traditional loss functions. The MSLAE component effectively handles the high dynamic range of ultrasound data, while the addition of the KLD term ensures a closer alignment of echogenicity distributions between input and target images. The combined loss function improves metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), as demonstrated in the previously mentioned paper [2], which achieved PSNR and SSIM improvements of up to 23% and 159%, respectively, compared to baseline models.

This approach allows the model to not only enhance contrast and reduce artifacts but also to maintain the natural distribution of echogenicity values, critical for medical diagnostic applications. By adopting the KLD loss function in this project, significant improvements in image fidelity and artifact reduction might be observed.

Chapter 3

Challenges in CNN Generalization Across Datasets

The performance of two CNN models, `CNN_9LD` and `CNN_11LD`, is shown in Figures 3.1 and 3.2. Table 5.2 summarizes the corresponding quantitative results.

1. **Step 1**(Figure 3.1): A CNN (`CNN_9LD`) was trained on Dataset 9LD and tested on the same dataset. This model performs well on its own dataset.

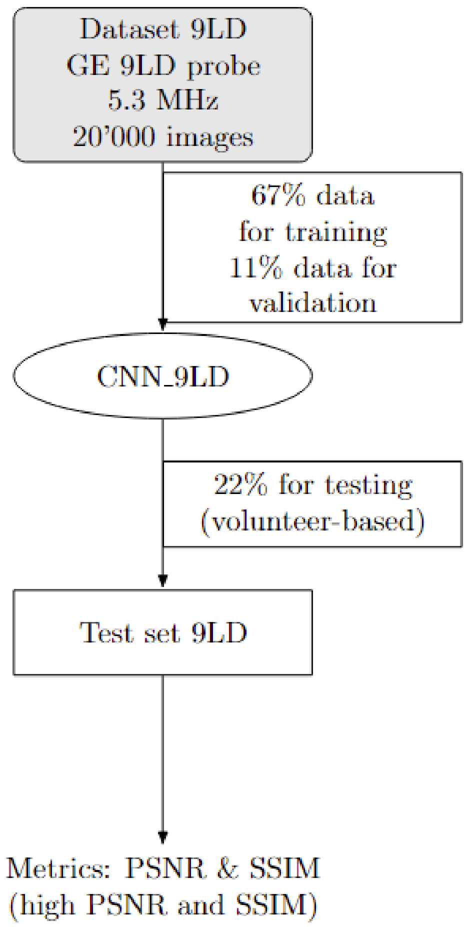

Figure 3.1: Performance of `CNN_9LD` trained on Dataset 9LD.

2. **Step 2**(Figure 3.2): A CNN (`CNN_11LD`) was trained on Dataset 11LD and tested on the same dataset. While this model achieves high performance on Dataset 11LD, it

performs poorly on Dataset 9LD.

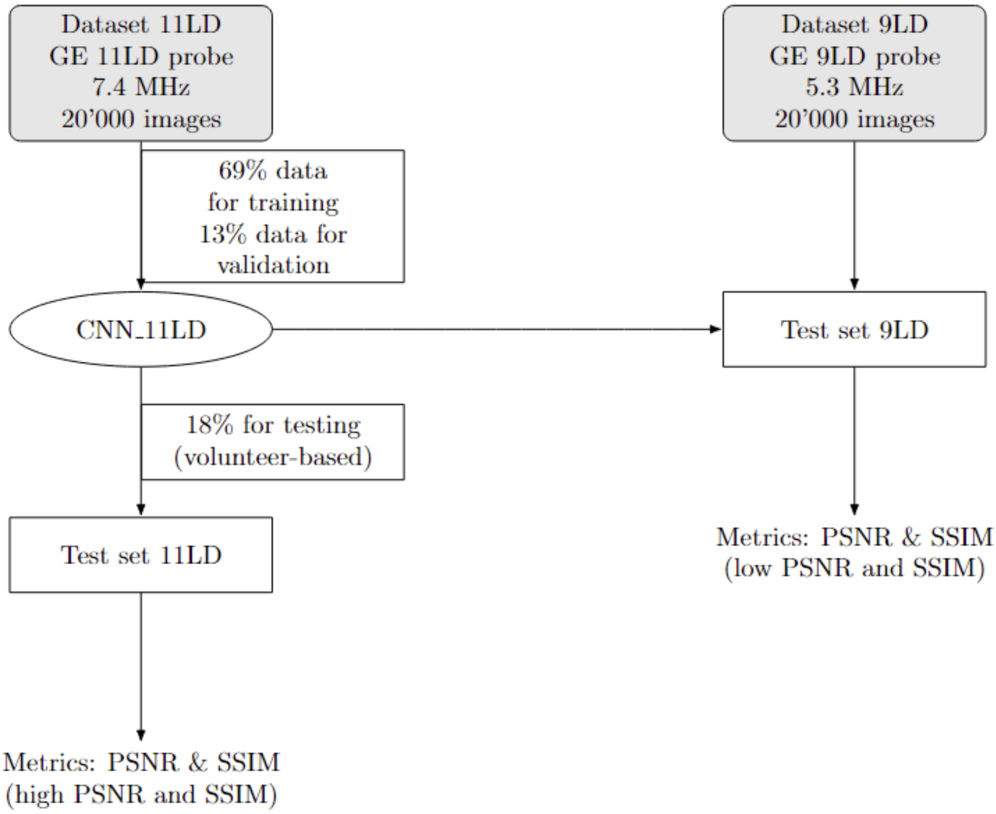

Figure 3.2: Performance of `CNN_11LD` trained on Dataset 11LD and tested on Dataset 9LD.

One major challenge in ultrafast ultrasound imaging is the variability between datasets acquired using different imaging probes and configurations. A CNN trained on a specific dataset, such as one acquired using a GE 11LD probe at 7.4 MHz, may not perform well when tested on images obtained from a different probe, such as the GE 9LD probe at 5.3 MHz. This performance gap arises because each probe produces images with different characteristics, including resolution, noise levels, and dynamic range.

Chapter 4

Transfer Learning

To address the issue discussed in Section 3, we propose a transfer learning approach, which is a technique where a model developed for a particular task is reused as the starting point for a model on a different but related task. This is particularly useful when the target task has limited data or computational resources.

This strategy adapts the knowledge learned by the CNN trained on Dataset 11LD to improve its performance on Dataset 9LD with minimal additional training. The process used to highlight the importance to use Transfer-Learning based approach, as shown in Figure 4.1.

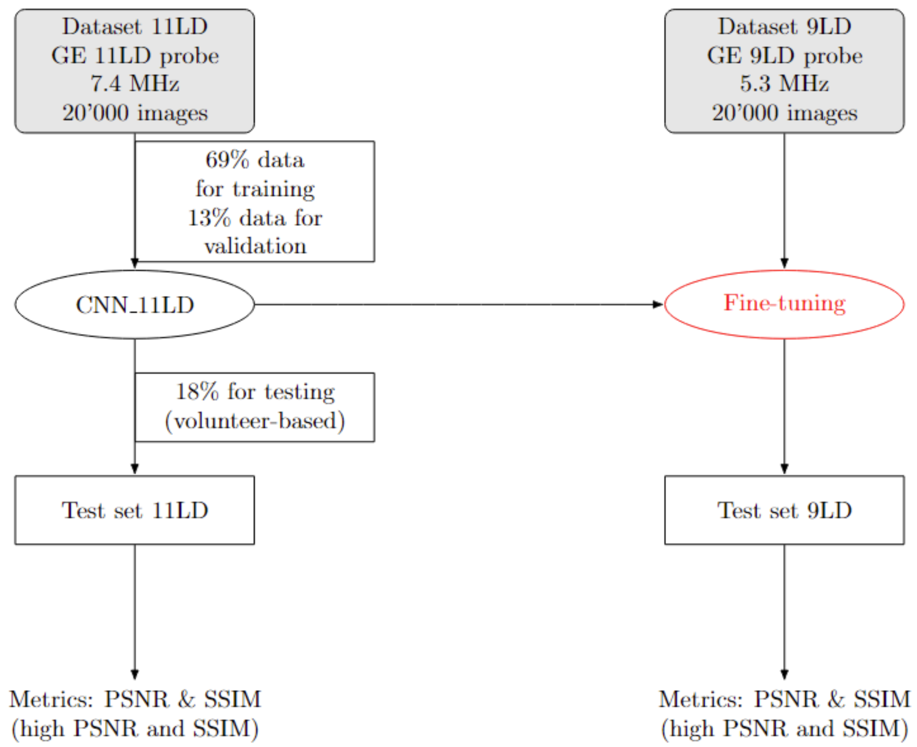

Figure 4.1: Transfer Learning: Step 3. Fine-tuning `CNN_9LD` on Dataset 11LD.

The use of transfer learning to fine-tune `CNN_11LD` on a small subset of Dataset 9LD allows the model to adapt to the new domain while reusing knowledge from Dataset 11LD. By doing so, we're expecting the PSNR and SSIM metrics on Dataset 9LD improve significantly, reducing the computational and data requirements compared to training from scratch.

4.1 Fine-tuning Strategies

In this project, we used a pretrained CNN model, specifically the 11LD model detailed in Section 2.2.2, associated with the volunteer-based splitting strategy (VS) described in Table 2.1.
To improve the model's performance, we focused on fine-tuning its parameters, particularly by varying the number of training images from the 9LD dataset. For instance, when fine-tuning is enabled and the specified number of images (e.g., 2'000, 3'000, 4'000) is less than the total available training set, they are being randomly selected from the training dataset. Additionally, 10% of this subset (400 images) is drawn from the test dataset for validation purposes. This random sampling ensures diversity and reduces potential bias by generating representative subsets, enhancing the pretrained model's generalization capabilities during the fine-tuning process.

To effectively adapt the pre-trained model, we explored various fine-tuning strategies. It's worth mentioning that the process of freezing and unfreezing layers is often guided by intuition and experimentation rather than a definitive method. Therefore, we are testing multiple strategies, as outlined below, along with the rationale behind each choice:

- **Strategy 1:** Freezing the encoder layers (the initial layers of the model) while training the downsampling layers. This approach is based on the basis that the 9LD dataset closely resembles the 11LD dataset in terms of overall image structure. Given their similarity, the encoder can preserve its pre-trained features, allowing the downsampling layers to focus on refining finer details that significantly impact the quality of UFUS images.

Note: The following diagram is a schema to illustrate the concept and does not reflect the exact architecture of our network.

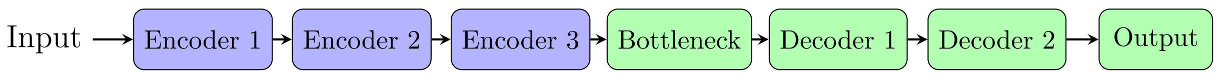

Figure 4.2: Model architecture showing frozen encoder layers (blue) and trainable decoder layers (green).

- **Strategy 2:** Freezing and training alternating layers.

It is based on the idea of selectively adapting the network to the target dataset. By freezing some layers and training others in an alternating fashion, this strategy aims to strike a balance between preserving pre-trained features and refining domain-specific details. This gradual adaptation allows the model to retain its general understanding of the 9LD dataset while fine-tuning specific layers to improve the quality of UFUS images for the 11LD dataset.

Note 1: The following diagram is a schema to illustrate the concept and does not reflect the exact architecture of our network.

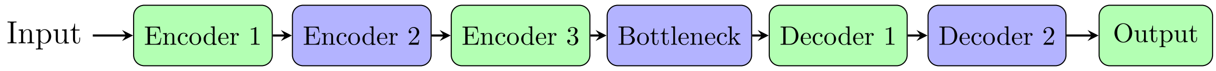

Figure 4.3: Model architecture with alterning frozen layers (purple) and alterning trainable layers (green).

Note 2: To try achieving even better results with the same strategy, we can parameterize the number of consecutive layers to freeze and the number of consecutive layers to train.

Chapter 5

Results & Discussion

5.1 Baseline comparison: Input vs. Target

5.1.1 Performance metrics

The following figures provide a visual comparison between the chosen input images (left column of each image pair), which are directly obtained from the probes involved in the study, and the corresponding target images (right column of each image pair). Figure5.1 illustrates this comparison for the 9LD dataset, while Figure5.3 presents the comparison for the 11LD dataset.

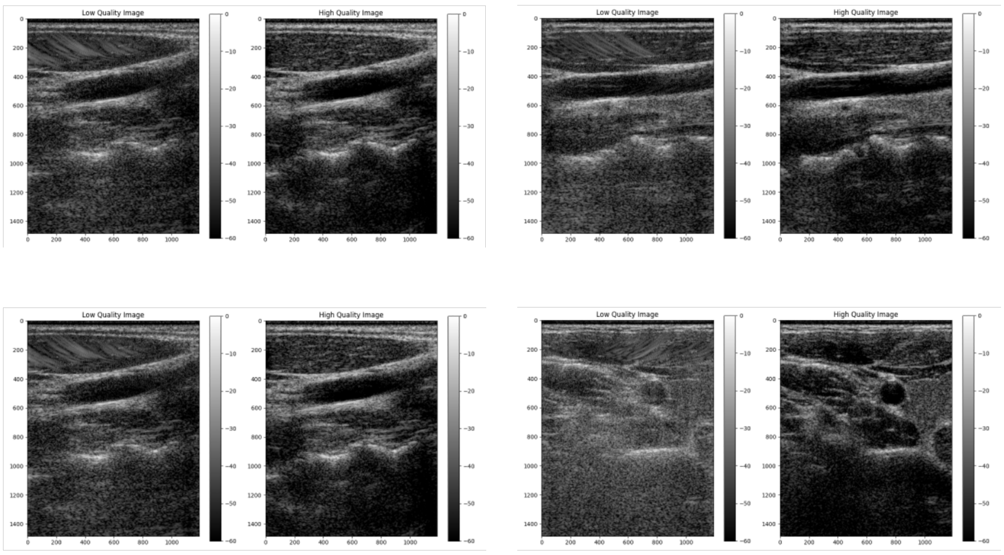

Figure 5.1: Comparison between Low Quality Input and Target (High quality) with 9LD Dateset.

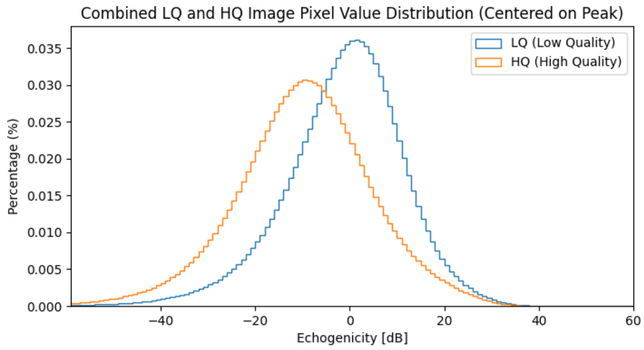

Figure 5.2: Histogram of echogenicity distributions of Low-Quality (LQ) and High-Quality (HQ) input images (Dataset 9L-D), centered on their respective peaks, with echogenicity values ranging from -60 dB to 60 dB

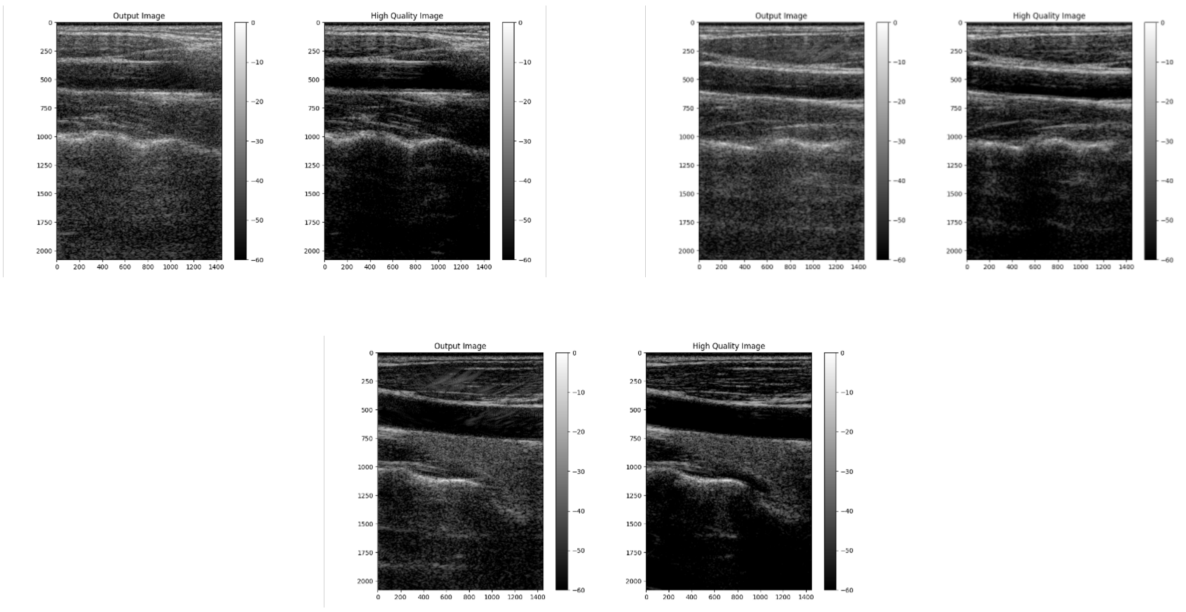

Figure 5.3: Comparison between Low Quality Input and Target (High quality) with 11LD Dateset.

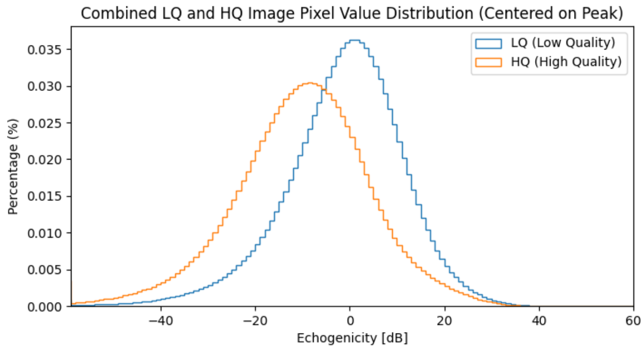

Figure 5.4: Histogram of echogenicity distributions of Low-Quality (LQ) and High-Quality (HQ) input images (Dataset 11L-D), centered on their respective peaks, with echogenicity values ranging from -60 dB to 60 dB

	Data set 9LD	**Data set 11LD**
Input vs Target	**PSNR:** *Mean*: 13.96 *Std*: 0.49 **SSIM:** *Mean*: 0.0625 *Std*: 0.230	**PSNR:** *Mean*: 13.91 *Std*: 0.84 **SSIM:** *Mean*: 0.0663 *Std*: 0.0499

Table 5.1: Performance metrics of input compared to target for each Data set

5.1.2 Metrics interpretation

The results presented in Table 5.1 provide a comprehensive overview of the baseline performance metrics, PSNR and SSIM, for the input and target images in both datasets (9LD and 11LD).

For the 9LD dataset, the PSNR mean value of 13.96 ($\pm$0.49) indicates a moderate reconstruction quality, suggesting that the input images retain some structural details but are heavily affected by noise and artifacts. The corresponding SSIM mean of 0.0625 highlights poor structural similarity and contrast fidelity between the input and target images, emphasizing the necessity for enhancement techniques.

Similarly, the 11LD dataset shows comparable baseline results with a mean PSNR of 13.91 ($\pm$0.84) and SSIM of 0.0663. These metrics suggest a slightly more consistent performance in terms of PSNR but reveal significant room for improvement in structural quality, as evidenced by the low SSIM values.

These observations highlight the challenges posed by the raw input images, which shows poor fidelity and structural integrity when compared to the target outputs. Such limitations are common in UFUS imaging due to trade-offs between frame rate and image quality.

The similarity in baseline performance across both datasets shows that the degradation patterns in the input images are consistent. This supports the use of transfer learning and fine-tuning strategies to address the quality gaps effectively, as these methods build on existing knowledge to improve the metrics. The following analysis will examine how these approaches impact PSNR and SSIM, enhancing image quality and diagnostic accuracy.

5.2 CNN Output Analysis

5.2.1 Performance metrics

Table 5.2 shows the mean and standard deviation of PSNR and SSIM for the two CNN models on both datasets. These results highlight the inability of the models to generalize across domains without additional adaptation techniques.

	CNN 9LD	**CNN 11LD**
Data Set 9LD	**PSNR:** *Mean*: 15.84 *Std*: 0.18 **SSIM:** *Mean*: 0.0983 *Std*: 0.0219	**PSNR:** *Mean*: 14.61 *Std*: 0.46 **SSIM:** *Mean*: 0.0759 *Std*: 0.0238
Data Set 11LD	**PSNR:** *Mean*: 15.45 *Std*: 0.21 **SSIM:** *Mean*: 0.0901 *Std*: 0.0179	**PSNR:** *Mean*: 14.47 *Std*: 0.52 **SSIM:** *Mean*: 0.0792 *Std*: 0.0250

Table 5.2: Performance Metrics for CNN 9LD and CNN 11LD on Data Sets 9LD and 11LD.

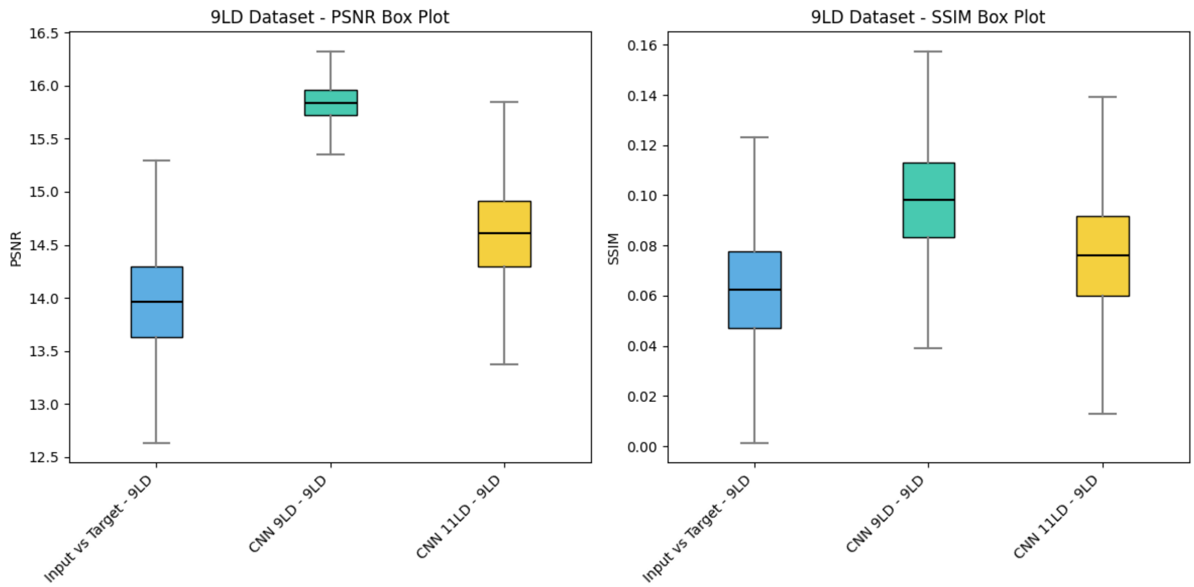

Figure 5.5: Box plot of PSNR and SSIM results comparing Dataset 9LD's input, CNN 9LD's output tested on Dataset 9LD and CNN 11LD's output tested on Dataset 9LD.

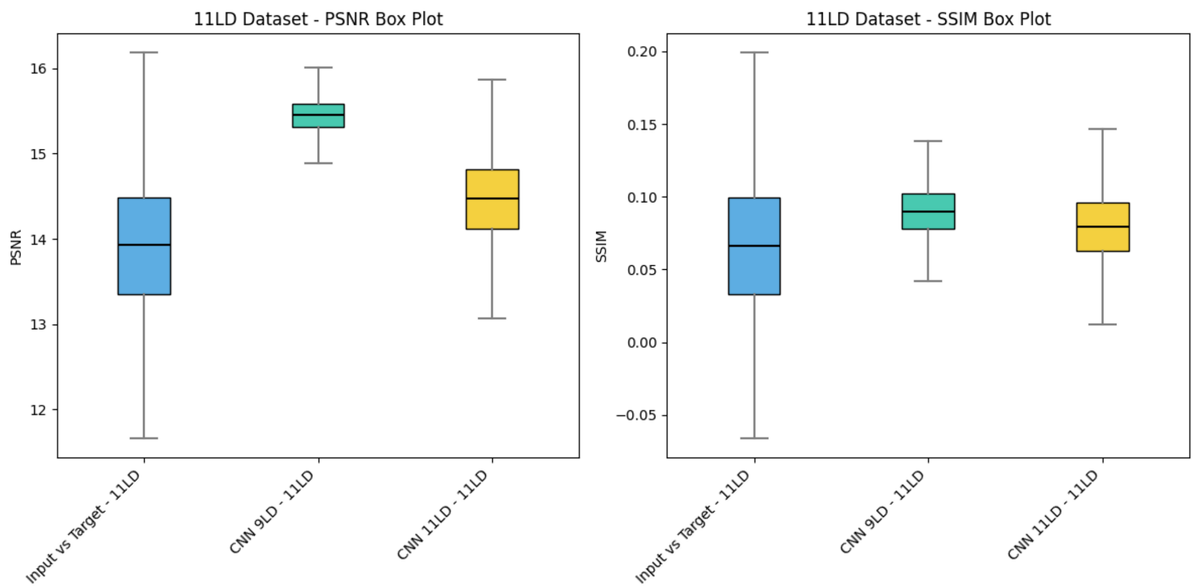

Figure 5.6: Box plot of PSNR and SSIM results comparing Dataset 11LD's input, CNN 9LD's output tested on Dataset 11LD and CNN 11LD's output tested on Dataset 11LD.

5.2.2 Metrics interpretation

The box plots in Figures 5.5 illustrate the performance of CNN models for improving ultrasound image quality on the 9LD data set.
The baseline "Dataset 9LD Target vs Input" category shows the lowest PSNR and SSIM values, representing the baseline performance of raw input images without any enhancement. In contrast, the 9LD set tested on CNN9LD ("CNN 9LD - Dataset 9LD") category

exhibits the highest PSNR and SSIM values, highlighting the effectiveness of the CNN model when trained and tested on the same dataset.

The cross-dataset evaluation ("CNN 9LD - Dataset 11LD" and "CNN 11LD - Dataset 9LD") shows a drop in PSNR and SSIM compared to the in-domain performance. This decline reflects the challenges faced by the models in generalizing across datasets, likely due to variations in imaging conditions or probe characteristics.

On the other hand, in the box plots in Figures 5.6, the 11LD dataset tells a slightly different story. The baseline "Dataset 11LD Target vs Input" again represents the unenhanced raw input, with the lowest PSNR and SSIM values. Unexpectedly, the "CNN 9LD - Dataset 11LD" configuration stands out with the highest performance, which indicates that the model trained on 9LD generalizes pretty well to the 11LD dataset, even better than CNN 11LD.

The strong performance of the 9LD-trained model on the 11LD dataset suggests that factors such as similarities between probes, dataset diversity and image distribution during training play a significant role in facilitating generalization. This aligns with the observation that the CNN trained on 9LD demonstrates robustness despite the differences in imaging conditions between the two datasets.

To keep the analysis clear and intuitive, we will focus on the 9LD dataset. This allows us to explore how transfer learning and fine-tuning can help a model adapt to new data, using PSNR and SSIM as reliable comparison and performance metrics.

5.3 Fine-tuned CNN Analysis

To effectively present multiple results, we have selected a single representative image from Figure 5.1 to display in the following figure. Specifically, we chose the lower-right carotid image, as it provides clear and detailed visual features. This image will serve as the basis for comparing the performance of various fine-tuning strategies.

Given that visual differences between processed images can often be subtle and non-visible to the naked-eye, we will focus on presenting an example that demonstrate meaningful and noticeable visual variations.

The numerical comparison results (PSNR and SSIM metrics) are displayed in table 5.3.

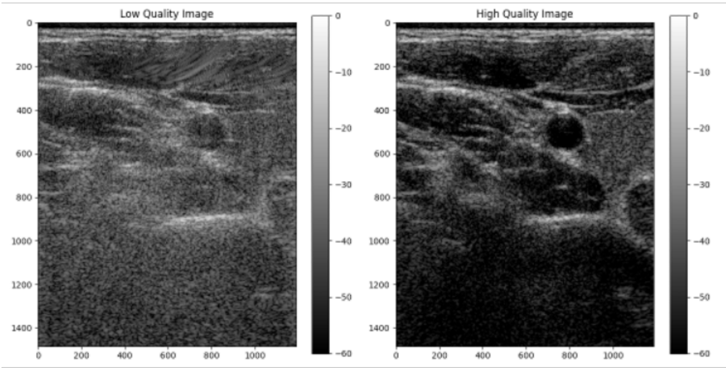

Figure 5.7: Reference Figure

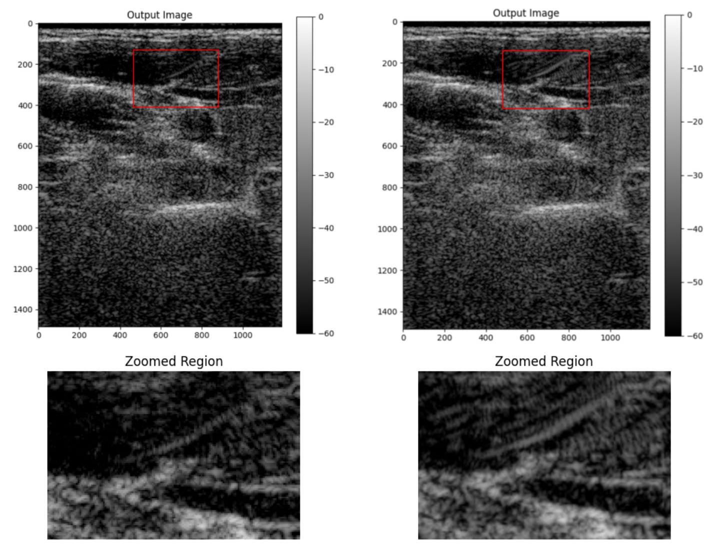

Figure 5.8: Comparison of encoder freezing (Left) vs alternative freezing strategies (Right) with respective zoomed regions.

5.3.1 Performance metrics

Added images vs strategy	**1'500**	**2'000**	**5'000**
Freezing encoder's layers	**PSNR:** *Mean*: 15.45 *Std*: 0.19 **SSIM:** *Mean*: 0.0933 *Std*: 0.0212	**PSNR:** *Mean*: 15.39 *Std*: 0.17 **SSIM:** *Mean*: 0.0914 *Std*: 0.203	**PSNR:** *Mean*: 15.56 *Std*: 0.21 **SSIM:** *Mean*: 0.0956 *Std*: 0.0219

Added images vs strategy	2'000	3'000	4'000
Alternate freezing (step of 4)	**PSNR:** *Mean*: 15.52 *Std*: 0.22 **SSIM:** *Mean*: 0.0945 *Std*: 0.0210	**PSNR:** *Mean*: 15.59 *Std*: 0.20 **SSIM:** *Mean*: 0.0952 *Std*: 0.0211	**PSNR:** *Mean*: 15.42 *Std*: 0.35 **SSIM:** *Mean*: 0.0937 *Std*: 0.0206

Added images vs strategy	1'500	3'000	4'000
Gradient based	**PSNR:** *Mean*: 15.49 *Std*: 0.19 **SSIM:** *Mean*: 0.0930 *Std*: 0.0209	**PSNR:** *Mean*: 15.29 *Std*: 0.17 **SSIM:** *Mean*: 0.0922 *Std*: 0.0203	**PSNR:** *Mean*: 15.48 *Std*: 0.24 **SSIM:** *Mean*: 0.0932 *Std*: 0.0203

Table 5.3: Performance Metrics for fine-tuned CNN 11LD tested on Data Set 9LD.

5.3.2 Discussion of Results of the fine-tuned model

The performance metrics for the fine-tuned CNN models, summarized in Table 5.3, highlight the impact of various fine-tuning strategies on the enhancement of ultrasound image quality. Among the strategies tested, alternating freezing with a step of 4 achieved the best results, with a mean PSNR of **15.59** and a mean SSIM of **0.0952** when trained with 3,000 images. This demonstrates the potential of selective layer adaptation to retain pre-trained features while effectively tuning the model to the specific characteristics of the 9LD dataset. Compared to the baseline results (PSNR: **13.96**, SSIM: **0.0625**), these results correspond to a relative improvement of **11.69%** in PSNR and **52.32%** in SSIM.

Key Observations

Impact of Training Set Size Across all strategies, the inclusion of more training images generally improves the model's performance up to a threshold. For instance, alternate freezing (step 4) showed a clear improvement in PSNR and SSIM as the training dataset size increased from 2,000 to 3,000 images. However, beyond this, the benefits appeared to plateau or diminish slightly, indicating that the model may have saturated its learning capacity for the given data distribution.

Comparison of Fine-Tuning Strategies The alternate freezing strategy consistently outperformed simple encoder freezing, achieving higher PSNR and SSIM values across all dataset sizes. This suggests that alternating freezing facilitates a more nuanced balance between feature preservation and dataset-specific learning. Gradient-based fine-tuning produced stable but slightly lower performance compared to alternating freezing, suggesting that it may be less effective in capturing the domain-specific nuances of the 9LD dataset.

Variability in Results The variability in PSNR and SSIM metrics (as indicated by the standard deviation) was relatively low across all strategies, demonstrating the robustness of the models to different fine-tuning configurations.

Notable Results The highest PSNR (**15.59**) and SSIM (**0.0952**) values were achieved using alternate freezing with a step of 4 and 3,000 training images. This marks a significant improvement compared to the baseline results (PSNR: **13.96**, SSIM: **0.0625**) and approaches the target results (PSNR: **15.84**, SSIM: **0.0983**) for the CNN 9LD model without fine-tuning.

The following figure 5.9 shows the histogram of echogenicity distribution for the best model we got (described in table 5.3). Comparing to the histograms describing the input images of 9L-D network in figure 5.2, we can clearly observe that the model has effectively narrowed this gap, as the output matches the HQ distribution, both in shape and range. This indicates the model's success in restoring or enhancing the LQ image to resemble HQ images.

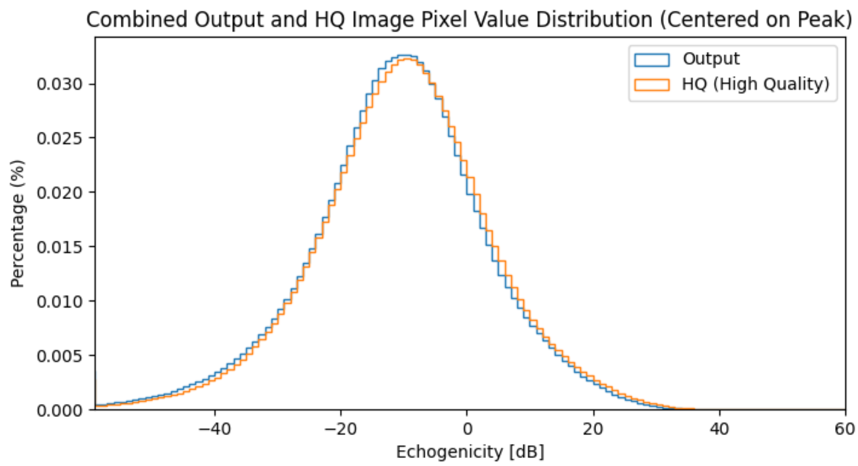

Figure 5.9: Histogram of echogenicity distributions of Output and High-Quality (HQ-Target), centered on their respective peaks, with echogenicity values ranging from -60 dB to 60 dB

Implications

These results underline the effectiveness of transfer learning and fine-tuning for domain adaptation in ultrafast ultrasound imaging. The performance improvements achieved, particularly through alternating freezing, demonstrate the feasibility of adapting pre-trained models to enhance image quality in challenging datasets. As additional results become available, further analysis will refine these conclusions and validate the trends observed in the current data.

Chapter 6

Conclusion

This project adresses the challenge of improving the quality of UFUS images, which often trade resolution and clarity for higher frame rates. By using deep learning techniques like U-Net-based convolutional neural networks (CNNs), we explored strategies like transfer learning and fine-tuning to make models perform better across different datasets.

The results showed clear improvements. Fine-tuning strategies, especially alternating layer freezing, improved image quality significantly, increasing PSNR to 15.59 and SSIM to 0.0952—up by 11.69% and 52.32%, respectively, compared to the starting point. Transfer learning made it possible to adapt model to new dataset without requiring massive retraining, proving its value in addressing variability between imaging conditions.
However, challenges remain. Differences between datasets, the trial-and-error nature of fine-tuning, and computational demands were notable challenges. Metrics like PSNR and SSIM, while useful, didn't always reflect the subtle improvements visible to the human eye or important for clinical use.

Future work can focus on testing the same CNN on additional datasets to evaluate whether the model genuinely simplifies the process of determining optimal fine-tuning strategies across domains. This step is essential to validate the robustness of the approach and its applicability to more varied scenarios.

In the end, this project demonstrated that deep learning can bridge the gap between UFUS's speed and the quality needed for accurate diagnostics. While there's room to address the challenges and validate findings further, the work sets a foundation for more reliable and adaptable UFUS imaging in healthcare.

Bibliography

[1] P. Motta, *Data Distributions Effect on CNN-Based Ultrafast Ultrasound Image Improvement*. Master Thesis, Computational Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, Academic Year: 2022-23.

[2] R. Viñals and J.-P. Thiran, ”A KL Divergence-Based Loss for In Vivo Ultrafast Ultrasound Image Enhancement with Deep Learning,” *Journal of Imaging*, vol. 9, no. 12, 2023, p. 256. `https://doi.org/10.3390/jimaging9120256`.

[3] R. Viñals, S. Beuret, and J.-P. Thiran, ”Quality Enhancement of Ultrafast Ultrasound Images with Deep Networks and Transfer Learning,” *2023 IEEE International Ultrasonics Symposium (IUS)*, 2023, pp. 1–6. `https://doi.org/10.1109/IUS51837.2023.10307893`.

[4] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, ”Deep Learning-Based Ultrasound Image Reconstruction Method: A Time Coherence Study,” *2019 IEEE International Ultrasonics Symposium (IUS)*, Glasgow, Scotland, October 6–9, 2019, pp. 448–451. `https://doi.org/10.1109/IUS45969.2019.8916202`.

[5] M. Z. Alom, C. Yakopcic, T. M. Taha, M. Hasan, and V. K. Asari, ”Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” *arXiv preprint arXiv:1802.06955*, 2018. `https://arxiv.org/abs/1802.06955`.