



Skolkovo Institute of Science and Technology

MASTER'S THESIS

**Intelligent Frequency-space Image Filtering for Computer  
Vision Problems in Medicine**

Master's Educational Program: Data Science

Student:\_\_\_\_\_

Viktor Shipitsin

Research Advisor:\_\_\_\_\_

Dmitry V. Dylov  
Associate Professor, Skoltech

Moscow 2021

All rights reserved.©

The author hereby grants to Skoltech permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole and in part in any medium now known or hereafter created.



Skolkovo Institute of Science and Technology

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**Интеллектуальная фильтрация изображений в частотном пространстве для задач компьютерного зрения в медицине**

Магистерская образовательная программа: Наука о данных

Студент: \_\_\_\_\_  
Виктор В. Шипицин

Научный руководитель: \_\_\_\_\_  
Дмитрий В. Дылов  
Доцент, Сколтех

Москва 2021

Все права защищены. ©

Автор настоящим дает Сколковскому институту науки и технологий разрешение на воспроизводство и свободное распространение бумажных и электронных копий настоящей диссертации в целом или частично на любом ныне существующем или созданном в будущем носителе.

# **Intelligent Frequency-space Image Filtering for Computer Vision Problems in Medicine**

Viktor Shipitsin

Submitted to the Skolkovo Institute of Science and Technology  
on June 2, 2021

## **Abstract**

This work is devoted to topical problems in ultrasound studies of various human organs and tissues that arise when analyzing images obtained with ultrasound sensors. Automation and elimination of human factors such as fatigue and inexperience are achieved by developing computer vision techniques that can proceed as an assistant to the medical specialist. In Computer Vision (CV) field, the tasks of segmentation, classification and denoising are quite successfully solved at the moment, thanks to the invention of convolutional neural networks. However, the accuracy of the solution and the achievement of high-quality metrics is an essential requirement in medicine.

This study proposes an adaptive layer algorithm based on the frequency representation of the image, which can be obtained using the Fourier transform. Depending on the base model, CV problem and dataset, global filtering is performed, which consists of adjusting the frequencies of the obtained image spectrum.

This work presents several configurations of the global filtering layer, which are trained along with selected popular models and are compared in terms of the number of parameters, computational complexity and metrics with models without a layer on ultrasound datasets for various CV problems.

The proposed algorithm in this research can be considered as an adaptive data preprocessing method, which achieves higher quality, performance rate and a smaller memory size on a computing device or the last two *ceteris paribus*.

Research Advisor:

Name: Dmitry V. Dylov

Degree: PhD

Title: Associate Professor, Skoltech

# **Интеллектуальная фильтрация изображений в частотном пространстве для задач компьютерного зрения в медицине**

**Виктор В. Шипицин**

Представлено в Сколковский институт науки и технологий  
Июнь 2, 2021

## **Реферат**

Данная работа посвящена актуальным задачам в ультразвуковых исследованиях (УЗИ) различных органов и тканей человека, которые возникают при анализе полученных с помощью датчиков УЗИ изображений. Автоматизация и исключение таких человеческих факторов как усталость и неопытность достигается с помощью разработки методов компьютерного зрения, которые могут выступать в роли ассистента медицинского специалиста. Задачи сегментации, классификации и шумоподавления довольно успешно решаются на данный момент, благодаря изобретению свёрточных нейронных сетей. Однако, точность решения и достижение высоких метрик качества является неотъемлемым требованием в области медицины.

В данном исследовании предлагается алгоритм адаптивного слоя, основанный на частотном представлении изображения, которое можно получить с помощью преобразования Фурье. В зависимости от базовой модели, задачи и набора данных производится глобальная фильтрация, которая заключается в настройке частот полученного спектра изображения.

В работе приводятся несколько конфигураций слоя глобальной фильтрации, которые обучаются вместе с выбранными популярными моделями и сравниваются по количеству параметров, сложности вычислений и метрикам с моделями без слоя на наборах ультразвуковых данных для различных задач компьютерного зрения.

Предлагаемый алгоритм в данном исследовании может рассматриваться как адаптивный метод предобработки данных, с использованием которого достигаются более высокие оценки качества и/или производительности и меньший занимаемый объём памяти *ceteris paribus* на электронно-вычислительной машине.

Научный руководитель:

Имя: Дмитрий В. Дылов

Ученое звание, степень: PhD

Должность: Доцент, Сколтех

# Contents

<b>Introduction</b>	<b>5</b>
<b>1 Theoretical Background</b>	<b>7</b>
1.1 Vector Spaces . . . . .	7
1.2 Fourier Transform . . . . .	9
1.3 Image Processing . . . . .	11
<b>2 Methodology</b>	<b>14</b>
2.1 Related Work . . . . .	14
2.2 Datasets Description . . . . .	16
2.3 Method . . . . .	18
<b>3 Results and Discussions</b>	<b>24</b>
3.1 Experiments . . . . .	24
3.2 Discussions . . . . .	29
<b>Conclusion</b>	<b>31</b>
<b>Bibliography</b>	<b>32</b>
<b>Appendix A Adjustment Algorithms</b>	<b>38</b>
<b>Appendix B Segmentation Results</b>	<b>39</b>
<b>Appendix C Classification Results</b>	<b>42</b>
<b>Appendix D Denoising Results</b>	<b>43</b>

# Introduction

The appearance and development of a neural network approach and an increase in the performance and power of computing resources have significantly improved computer vision algorithms in recent years. Large amounts of data attracted by researchers led to the construction of models with a huge number of parameters [41], which increased the training time, the required storage space and the performance rate. All this prompted the origin of a trend to develop efficient modules in data processing and optimization to obtain more informative representations with minimal required resources [12, 39].

Transforming images before transferring them to a specific neural network is often associated with image preprocessing. For example, data augmentation [5, 58] makes the dataset generally more robust and prevent overfitting, especially in the case of many application domains that do not have access to big data [43], such as medical image analysis.

Data preprocessing is an essential part of any computer vision algorithms [47]. The most common methods are based on the representation of an image in a computer. There are different representations of the image, which are achieved using a transformation to another space, for example, the frequency domain. This representation helps analyse the spectrum, energy, and phase components of each frequency, one of the fundamental and powerful analysis tools in signal processing [24, 51].

This study is motivated by typical images taken during ultrasonic examinations in the clinic. Their grainy appearance, low resolution, and poor contrast demand an eye of a very qualified expert to discern patterns and to spot pathologies. Automating any computer vision task on such data always involves excessive preprocessing and batch filtering, with an accumulation of the error emerging due to the annotation uncertainty, the digital post-filtering artefacts, and the amplified noise. Each patient case generally requires an individually tuned frequency filter to obtain optimal image contrast and to yield the desired outcome in a given computer vision problem. Considering the image in frequency space, we try to account for the physics of ultrasound examination. However, developing the necessary frequency filter that will take into account the specifics of the task and data manually is a rather tricky task.

This research is based on a deep learning approach. The main contribution of this work is to provide the concept of a trainable layer, which will be trained together with the selected model to solve different computer vision problems for ultrasound modality of images. By placing this

layer in front of the model's baseline, the algorithm will automatically select weights for frequency filtering of images at the input to the model, which allows the algorithm to take into account the specifics of the problem presented in the training dataset.

In this work, to better understand the proposed approach's foundations, the text will be structured as follows. In Chapter 1, theoretical information will be given, revealing the analysis methods and existing techniques to the investigated problem. Chapter 2 will discuss the methodology of the proposed approach and its application to the described problems after covering the works related to different computer vision tasks and frequency approaches to solving them. And finally, Chapter 3 will be devoted to the obtained results and a discussion of the details of the study.

The results corresponding to the methodology of this research work are essentially highlighted in the paper [42].

# Chapter 1

## Theoretical Background

Computer vision tasks are related to the processing of discrete digital signals. A large number of new ideas and approaches that arise in the search and analysis of solutions to problems in machine learning are associated with a view from a different angle. The mentioned method is quite often performed by mapping to another space, in which the data acquires a new structure. The current approach to signal processing of different modalities is associated with receiving, transmitting, converting and analysing analogue and digital signals. We will focus on interpreting a two-dimensional finite digital signal already received and presented in discrete form - an image, which in this field of study is often performed by considering the frequency spectrum.

### 1.1 Vector Spaces

We will use linear algebra notation to describe theoretical concepts. Let us recall the basic understandings related to vector spaces.

A set of vectors  $\mathcal{V}$  with operations

- addition:  $\mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$
- multiplication by a scalar from a field  $\mathbb{F}$ :  $\mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$

will be called a **vector space over the field  $\mathbb{F}$**  if the following axioms is satisfied:

1. commutativity of addition:  $x + y = y + x \quad \forall x, y \in \mathcal{V}$
2. associativity of addition:  $(x + y) + z = x + (y + z) \quad \forall x, y, z \in \mathcal{V}$
3. identity element of addition:  $\exists \mathbf{0} \in \mathcal{V} : x + \mathbf{0} = x \quad \forall x \in \mathcal{V}$
4. inverse element of addition:  $\forall x \in \mathcal{V} \quad \exists (-x) \in \mathcal{V} : x + (-x) = \mathbf{0}$
5. distributivity of scalar multiplication with respect to vector addition:  
$$\alpha(x + y) = \alpha x + \alpha y \quad \forall x, y \in \mathcal{V}, \forall \alpha \in \mathbb{F}$$
6. distributivity of scalar multiplication with respect to field addition:  
$$(\alpha + \beta)x = \alpha x + \beta x \quad \forall x \in \mathcal{V}, \forall \alpha, \beta \in \mathbb{F}$$



7. associativity of scalar multiplication:  $(\alpha\beta)x = \alpha(\beta x) \quad \forall x \in \mathcal{V}, \forall \alpha, \beta \in \mathbb{F}$
8. identity element of scalar multiplication:  $\mathbf{1}x = x \quad \forall x \in \mathcal{V}$

In what follows, our attention will be mainly focused on vector spaces over the field of complex numbers  $\mathbb{C}$ , that is, complex linear spaces.

The transition to discrete signals will be carried out from the initial consideration of signals as functions in *Hilbert* space, which expands the concept of *Euclidean* space. For this, we recall a few more definitions.

Let  $\mathcal{V}$  be a complex linear space. The **inner product** in  $\mathcal{V}$  is a mapping  $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$  satisfying the properties:

1. linearity in the first argument:  $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \quad \forall x, y, z \in \mathcal{V}, \forall \alpha, \beta \in \mathbb{C}$
2. conjugate (Hermitian) symmetry:  $\langle x, y \rangle = \langle y, x \rangle^* \quad \forall x, y \in \mathcal{V}$
3. positive definiteness:  $\langle x, x \rangle \geq 0$  and  $\langle x, x \rangle = 0 \Leftrightarrow x = \mathbf{0} \quad \forall x \in \mathcal{V}$

A **norm** on  $\mathcal{V}$  is a function  $\| \cdot \| : \mathcal{V} \rightarrow \mathbb{R}$  with the following properties:

1. triangle inequality (subadditivity):  $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathcal{V}$
2. absolute scalability (homogeneity):  $\|\alpha x\| = |\alpha| \|x\| \quad \forall x \in \mathcal{V}, \forall \alpha \in \mathbb{C}$
3. positive definiteness:  $\|x\| \geq 0$  and  $\|x\| = 0 \Leftrightarrow x = \mathbf{0} \quad \forall x \in \mathcal{V}$

In the space  $\mathcal{V}$  there is a norm induced by the inner product:  $\|x\| = \sqrt{\langle x, x \rangle}$ .

**Hilbert Space.** Let us consider the  $\mathcal{L}_2(E)$  space of measurable (by Lebesgue) functions  $f(x)$ , where  $x \in E$ , which satisfy the property:

$$\left( \int_E |f(x)|^2 dx \right)^{1/2} < \infty$$

This property is important, in view of the possibility of the existence of a norm  $\|f(x)\| \quad \forall f \in \mathcal{L}_2(E)$  generated by the inner product:

$$\langle f, g \rangle = \left( \int_E f(x)g(x)^* dx \right)^{1/2}, \quad \text{where } x \in E \text{ and } f, g \in \mathcal{L}_2(E) \quad (1.1.1)$$

Thus,  $\mathcal{L}_2(E)$  is a normed vector space (where the norm is finite) of square-integrable functions.  $\mathcal{L}_2(E)$  is also a complete metric space (in which each fundamental sequence converges

to an element of this space) with respect to the **distance function** induced by the inner product:  $d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle} \quad \forall x, y \in \mathcal{V}$ , i.e a *Hilbert space*.

Each function (vector) in such space can be represented as a linear combination of a set of **basis** vectors  $\{\phi_k\}_{k \in \mathcal{I}}$ , which must be linearly independent (or any subset of it must be linearly independent in the case of infinite-dimensional space):  $\sum_{k \in \mathcal{I}} \alpha_k \phi_k = 0 \Leftrightarrow \alpha_k = 0 \quad \forall k$ .

Further, we will also use the properties of **orthogonality** and **orthonormality** of vectors. Let  $S = \{\phi_k\}_{k \in \mathcal{I}}$  be a set of vectors. Then

- S is orthogonal, if  $\langle \phi_k, \phi_m \rangle = 0 \quad \forall k, m \in \mathcal{I}, k \neq m$
- S is orthonormal, if S is orthogonal and  $\langle \phi_k, \phi_k \rangle = 1 \quad \forall k \in \mathcal{I}$

To analyse the frequency spectrum of a signal, we need to expand the function (vector) on some orthogonal basis, that is, to represent it as a linear combination of basis vectors. When the basis is orthogonal, it is simple to find the coefficients in the expansion. Actually,

$$x = \sum_{k \in \mathcal{I}} \alpha_k \phi_k = \sum_{k \in \mathcal{I}} \left\langle x, \frac{\phi_k}{\|\phi_k\|^2} \right\rangle \phi_k, \quad (1.1.2)$$

where  $x \in \mathcal{V}$ . Because  $\langle x, \phi_m \rangle = \sum_{k \in \mathcal{I}} \alpha_k \langle \phi_k, \phi_m \rangle = \alpha_m \langle \phi_m, \phi_m \rangle = \alpha_m \|\phi_m\|^2 \quad \forall m \in \mathcal{I}$

Frequency space is nothing more than the consideration of functions on another basis, into which one can go using the linear operator **Fourier transform**, which we will consider below.

## 1.2 Fourier Transform

Let us consider  $\mathcal{L}_2([-L; L])$  space, where there is an infinite-dimensional set of orthogonal vectors  $\{\cos(\frac{\pi k}{L}x), \sin(\frac{\pi k}{L}x)\}_{k=0}^{\infty}$ . A fundamental result in Fourier analysis is that if  $f(x)$  is real-valued periodic and piece-wise smooth, then it can be expanded on a cosine and sine basis as (1.1.2):

$$f(x) = \sum_{k=0}^{\infty} \left\langle f, \frac{\cos(\frac{\pi k}{L}x)}{\|\cos(\frac{\pi k}{L}x)\|} \right\rangle \frac{\cos(\frac{\pi k}{L}x)}{\|\cos(\frac{\pi k}{L}x)\|} + \left\langle f, \frac{\sin(\frac{\pi k}{L}x)}{\|\sin(\frac{\pi k}{L}x)\|} \right\rangle \frac{\sin(\frac{\pi k}{L}x)}{\|\sin(\frac{\pi k}{L}x)\|} \quad (1.2.1)$$

or can be written in terms of a **Fourier series**:

$$f(x) = \frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos\left(\frac{\pi k}{L}x\right) + B_k \sin\left(\frac{\pi k}{L}x\right) \quad (1.2.2)$$

For complex-valued functions we can use Euler's formula to write a Fourier series in complex form with complex coefficients:

$$f(x) = \sum_{k=-\infty}^{\infty} (\alpha_k + j\beta_k) \left[ \cos\left(\frac{\pi k}{L}x\right) + j \sin\left(\frac{\pi k}{L}x\right) \right] = \sum_{k=-\infty}^{\infty} c_k e^{j\frac{\pi k}{L}x} \quad (1.2.3)$$

Similarly, in this case we have an orthogonal set of vectors  $\{e^{j\frac{\pi k}{L}x}\}_{k \in \mathbb{Z}}$ , which provide the basis for complex-valued periodic functions.

The **Fourier transform** occurs when we want to consider functions on an unbounded domain  $[-L; L]$  for  $L \rightarrow \infty$ . Define  $\omega_k = \frac{\pi k}{L}$  and  $\Delta\omega = \frac{\pi}{L}$ . Consider  $\mathcal{L}_2((-\infty, \infty))$ , that is, take the limit  $L \rightarrow \infty$  (or  $\Delta\omega \rightarrow 0$ ):

$$\begin{aligned} f(x) &= \lim_{L \rightarrow \infty} \sum_{k=-\infty}^{\infty} c_k e^{j\frac{\pi k}{L}x} = \lim_{L \rightarrow \infty} \sum_{k=-\infty}^{\infty} \left\langle f, \frac{e^{j\frac{\pi k}{L}x}}{\|e^{j\frac{\pi k}{L}x}\|} \right\rangle \frac{e^{j\frac{\pi k}{L}x}}{\|e^{j\frac{\pi k}{L}x}\|} = \\ &= \lim_{\Delta\omega \rightarrow 0} \sum_{k=-\infty}^{\infty} \frac{\Delta\omega}{2\pi} \left[ \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} f(x) e^{-j\omega_k x} dx \right] e^{j\omega_k x} \end{aligned}$$

Passing from the sum of infinitesimal quantities to its limit by integration, we obtain

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi} \left[ \int_{-\infty}^{\infty} f(x) e^{-j\omega x} dx \right] e^{j\omega x} d\omega \quad (1.2.4)$$

In this expression, the **direct** and **inverse** Fourier transforms are hidden:

$$\mathcal{F}(f(x)) = \hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-j\omega x} dx \quad (1.2.5)$$

$$\mathcal{F}^{-1}(\hat{f}(\omega)) = f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{j\omega x} d\omega \quad (1.2.6)$$

We list the important **properties** of this transformation, which perform a principal role in signal processing:

- The direct and inverse Fourier transforms are **linear** operators, so:

$$\mathcal{F}(\alpha f(x) + \beta g(x)) = \alpha \mathcal{F}(f(x)) + \beta \mathcal{F}(g(x))$$

$$\mathcal{F}^{-1}(\alpha \hat{f}(\omega) + \beta \hat{g}(\omega)) = \alpha \mathcal{F}^{-1}(\hat{f}(\omega)) + \beta \mathcal{F}^{-1}(\hat{g}(\omega))$$

- **Parseval's theorem** is an important result in Fourier analysis that can be used to put guarantees on the accuracy of signal approximation in the Fourier domain:

$$\int_{-\infty}^{\infty} \|\widehat{f}(\omega)\|^2 d\omega = 2L \int_{-\infty}^{\infty} \|f(x)\|^2 dx, \text{ for } 2L\text{-periodic functions.}$$

- In functional analysis **convolution** is a mathematical operation on two functions that produces a third function expressing how the shape of one is modified by the other.

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-t)g(t) dt = \int_{-\infty}^{\infty} f(t)g(x-t) dt$$

The following **convolution theorem** holds

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$$

## 1.3 Image Processing

**Discrete Fourier transform (DFT).** To analyze data on computers, it is necessary to approximate continuous signals by discrete vectors. Initially, we able to obtain discrete data sample  $\mathbf{f} = [f(x_0), \dots, f(x_{n-1})]^T = [f_0, \dots, f_{n-1}]^T$ . Then we can numerically compute the direct and inverse Fourier Transform (assuming that the function is defined on  $[0; 2L]$  and  $x_i = \frac{i}{n}2L$ ):

$$\widehat{f}_k = \sum_{i=0}^{n-1} f(x_i) e^{-j\omega_k i} = \sum_{i=0}^{n-1} f_i e^{-j\frac{2\pi k}{n} i}$$

$$f_k = \frac{1}{n} \sum_{i=0}^{n-1} \widehat{f}(x_i) e^{j\omega_i k} = \sum_{i=0}^{n-1} \widehat{f}_i e^{j\frac{2\pi i}{n} k}$$

Let  $w_n = e^{-j\frac{2\pi}{n}}$ . Then

$$\widehat{\mathbf{f}} = \begin{bmatrix} \widehat{f}_0 \\ \widehat{f}_1 \\ \widehat{f}_2 \\ \vdots \\ \widehat{f}_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w_n & w_n^2 & \dots & w_n^{n-1} \\ 1 & w_n^2 & w_n^4 & \dots & w_n^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w_n^{n-1} & w_n^{(n-1)2} & \dots & w_n^{(n-1)(n-1)} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{n-1} \end{bmatrix} = \mathbf{F}_n \mathbf{f} \quad (1.3.1)$$

**Fast Fourier Transform (FFT).** For standard matrix-vector multiplication  $O(n^2)$  operations are required. But DFT (1.3.1) may be implemented much more efficiently if the number of data points  $n$  is a power of 2.

$$\hat{\mathbf{f}} = \mathbf{F}_n \mathbf{f} = \begin{bmatrix} \mathbf{I}_{n/2} & -\mathbf{D}_{n/2} \\ \mathbf{I}_{n/2} & -\mathbf{D}_{n/2} \end{bmatrix} \begin{bmatrix} \mathbf{F}_{n/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{n/2} \end{bmatrix} \begin{bmatrix} \mathbf{F}_{\text{even}} \\ \mathbf{f}_{\text{odd}} \end{bmatrix}, \quad (1.3.2)$$

where

$$\mathbf{D}_{n/2} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & w_n & 0 & \cdots & 0 \\ 0 & 0 & w_n^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n^{(n/2-1)} \end{bmatrix}$$

**2D Fourier Transform.** The Fourier transform of one-dimensional signals is easily generalized to the high-dimensional case. If we represent the image as a two-dimensional signal, then making an alternating one-dimensional Fourier transform, first by rows, then by columns (or vice versa), we obtain a description of the image in the frequency domain [4].

$$\hat{\mathbf{I}}(u, v) = \frac{1}{N_{\text{cols}} N_{\text{rows}}} \sum_{x=0}^{N_{\text{cols}}-1} \sum_{y=0}^{N_{\text{rows}}-1} \mathbf{I}(x, y) e^{-j \frac{2\pi u}{N_{\text{cols}}} x} e^{-j \frac{2\pi v}{N_{\text{rows}}} y}, \quad (1.3.3)$$

where  $\mathbf{I}(x, y)$  - original image (spatial description),  $\hat{\mathbf{I}}(u, v)$  - frequency domain image. Denoting  $W_n = e^{j \frac{2\pi}{n}}$ , we can rewrite the expression as follows:

$$\hat{\mathbf{I}}(u, v) = \frac{1}{N_{\text{cols}} N_{\text{rows}}} \sum_{x=0}^{N_{\text{cols}}-1} \sum_{y=0}^{N_{\text{rows}}-1} \mathbf{I}(x, y) W_{N_{\text{cols}}}^{-xu} W_{N_{\text{rows}}}^{-yv} \quad (1.3.4)$$

Similarly, the formula for the inverse transformation is obtained:

$$\mathbf{I}(x, y) = \sum_{u=0}^{N_{\text{cols}}-1} \sum_{v=0}^{N_{\text{rows}}-1} \hat{\mathbf{I}}(u, v) e^{j \frac{2\pi x}{N_{\text{cols}}} u} e^{j \frac{2\pi y}{N_{\text{rows}}} v}, \quad (1.3.5)$$

Initially, we believe that the entire plane is periodically paved (with periods  $N_{\text{cols}}$  and  $N_{\text{rows}}$ ) with the original image (Fig. 1.1).

In the frequency spectrum, specific frequencies are responsible for different properties of the image, which can be extracted with filtering (Fig. 1.2). Further research work is based on the study of filtering as image preprocessing. The existing approaches and methodology will be presented in the following chapters, together with discussing the details and difficulties encountered.

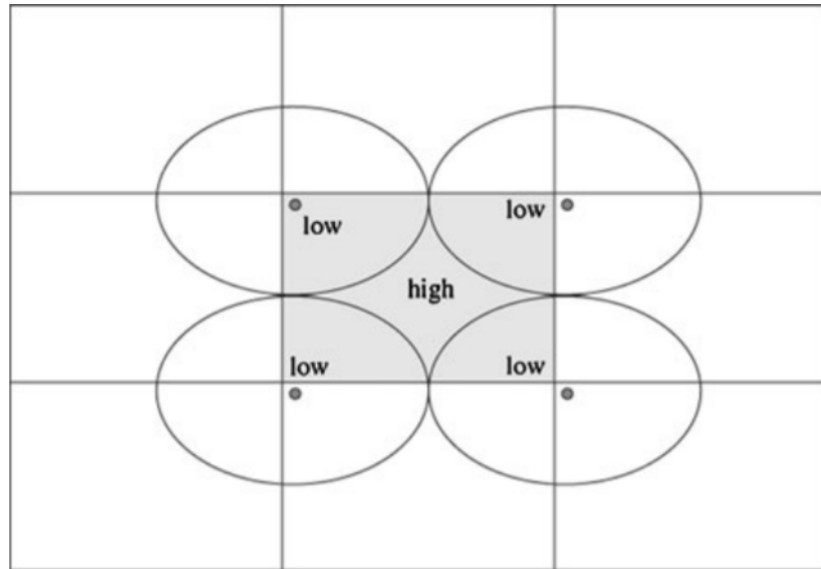


Figure 1.1. Due to the periodicity, low frequencies are in the shown ellipses and thus in the four corners of the matrix; the highest frequencies are at the centre of the matrix [23].

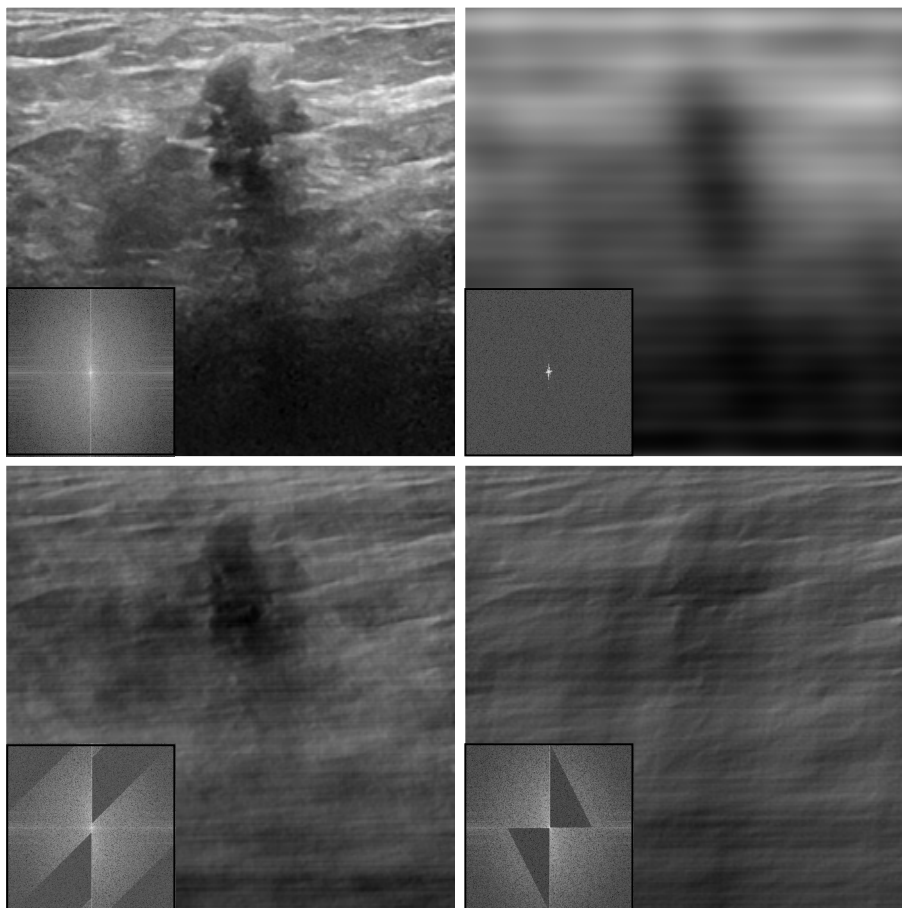


Figure 1.2. Motivation elaborated. Different frequency filters applied to the initial ultrasonic image (top left) and their Fourier spectra. Notice the effect of high-pass/low-pass components on morphology and texture.

## Chapter 2

# Methodology

### 2.1 Related Work

Ultrasound scanning is a widely used method in the medical field due to the fact that it is inexpensive, non-invasive, safe, accessible, without any ionization [6]. Internal organs and tissues of a human are able to reflect an echo, for example, to return a signal during ultrasound examination, that is, they have echogenicity. Depending on the high or low echogenicity in the ultrasound image, we will see lighter or darker colors, respectively [19]. With ultrasound, a number of soft-tissue (parenchymal) formations can be completely visualized, the state of which has important diagnostic value:

- Thyroid;
- Salivary glands (parotid and submandibular, paired);
- Lymph nodes (according to anatomical levels - from I to VI).

In these organs\structures, the pathology can be diffuse (affecting the entire organ/structure), focal or mixed. Today, analysis of various formations is carried out by each doctor personally and its information content is extremely operator-dependent. With the development of computer vision technologies, implementation of those algorithms to assist doctors in analyzing ultrasound is a perspective trend in the area of biomedicine [46, 52]. Such a combination of deep learning and biomedicine can solve a very serious problem, such as automatic image segmentation, detection or classification (Fig. 2.1).

**Image Segmentation.** Image segmentation is a very vast domain of computer vision problems. When speaking about medical image segmentation we cannot omit mentioning the U-Net encoder-decoder like model [40]. At the moment there are various modifications, each of which brings improvements and represents a state-of-the-art solutions: Attention U-Net [35], UNet++ [59, 60], UNet 3+ [18], etc. At the same time, the skeleton of models such as ResNet [13] and DenseNet [17], created for classification, are often used for segmentation problems. Despite the larger number of parameters and slower training, they represent better results in some cases. One of the natural

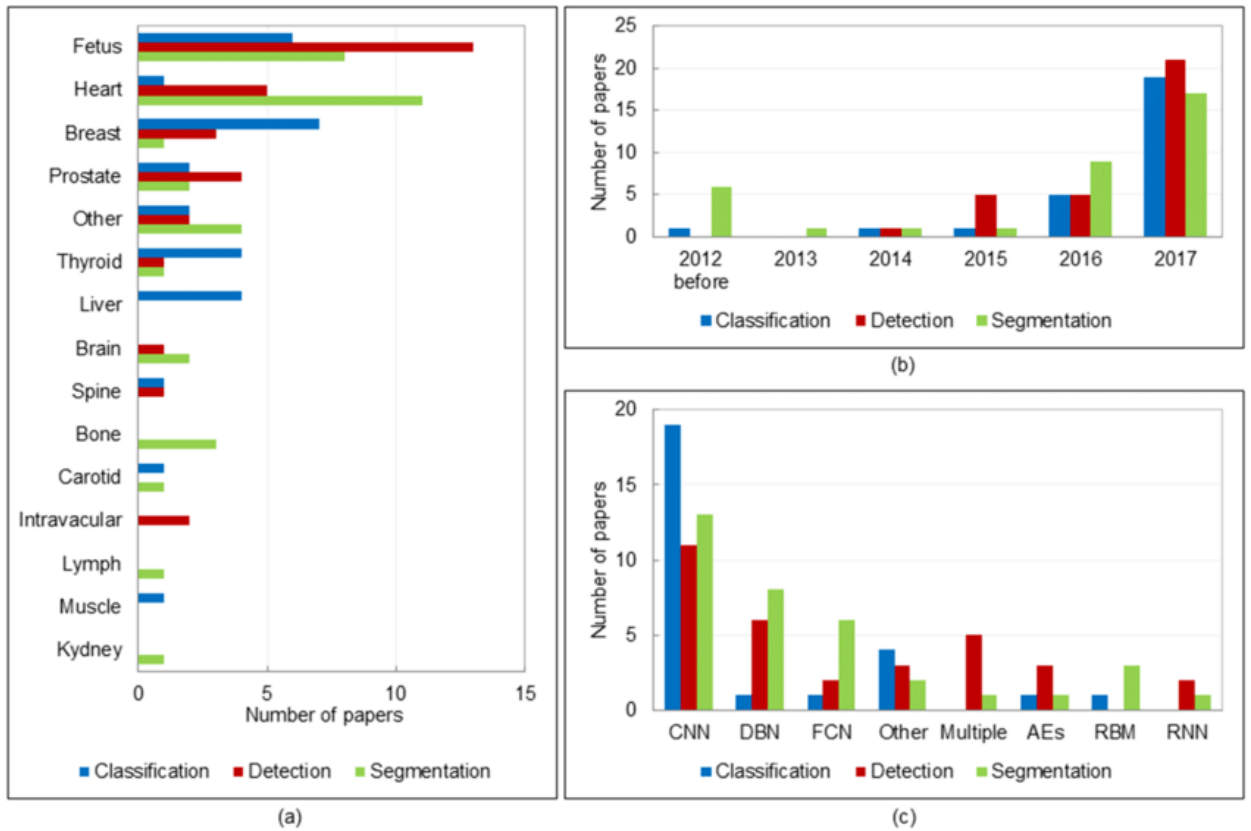


Figure 2.1. Rapid development of ultrasound analysis using deep learning [30].

approaches is to look for lightweight models that give comparable quality to heavier models [13, 47].

**Image Classification.** This problem is devoted to splitting images into classes, each of which has its own distinctive properties or physical interpretation [38, 47]. The segmentation task into two areas can be considered as a pixel-wise binary classification problem. So by constructed embeddings using an encoder used for the segmentation task, you can get a solution that is a precursor for segmentation [38].

**Image Denoising.** Another popular task is the restoration of an image after exposure to various types of noise or effects such as blur [31]. Such a proven approach as DnCNN [56, 61] copes well with the classic problems of deblurring and denoising. The problem of frequency filtering for denoising has been studied very thoroughly in the signal processing and in the imaging physics communities [6, 7]. The quality of data strongly influences ultrasound image segmentation. There are characteristic artefacts that make the segmentation task complicated, such as attenuation, reverberation, speckles, shadows, excessive absorbance, reflection and signal dropout; due to the orientation dependence of acquisition that can result in missing boundaries. Further complications



arise as the contrast between areas of interest is often low. So, ultrasound's typical high\mid-range (5-15 MHz) and low (2-5 MHz) frequency probes provide either good resolution or good penetration, but not both at once. The resulting images, therefore, are extremely susceptible to frequency tuning.

**Frequency Analysis.** Using the direct and inverse Fourier transform, we can go from the spatial domain of the image description to the frequency domain and vice versa. Fourier analysis has been successfully used for dynamic structure segmentation problems, where dynamic structures were distinguished using only the phase spectrum [28], semantic segmentation using domain adaptation, where the frequency amplitudes of the source and target images are combined [55], and various high-frequency low-dimensional regression problems, where Fourier features improved the results of coordinate-based multilayer perceptrons for image regression, 3D shape regression, MRI reconstruction, inverse rendering tasks [48].

So, the frequency domain carries a lot of information. Existing algorithms do not consider the selection of the required spectrum as data preprocessing. The difficulty is that manually selecting the correct frequency filter is not a simple task. Therefore, the suggested method described further is for the computer vision model to automatically select the weights for transforming the frequency spectrum for each specific task. As a result, the invention automatically performs the denoising process since it results in removing «unnecessary» noise frequencies depending on the provided dataset.

## 2.2 Datasets Description

The ultrasound method is a principally manual method of arbitrary positioning by the operator's hand and, unlike X-ray Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) (recording on a disc in DICOM format), it is not standardized from a tomographic (full organ review) point of view. Simultaneously, on modern ultrasonic devices, it is possible to save individual images (ultrasound slices) or a series of consecutive images.

**Endocrinology Dataset.** This dataset represents raw tagged data provided by the Center for Endocrinology<sup>1</sup>. Specialists form a sets of ultrasound frames and corresponding frames with highlighted areas responsible for various nodal formations for each of 15 patients (Fig. 2.2). Each slice set for all patients is classified as Thyroid image reporting and data system (TIRADS) level, developed by the American College of Radiology (ACR) [49]. Each level, which is determined by echostructure, echogenicity, shape, contour and inclusions, has a different risk of malignancy and

---

<sup>1</sup><https://www.endocrinentr.ru/>

further tactics for treating patients, for example, whether it is necessary to perform a fine needle aspiration (FNA) biopsy or additional ultrasound follow-up of suspicious nodules.

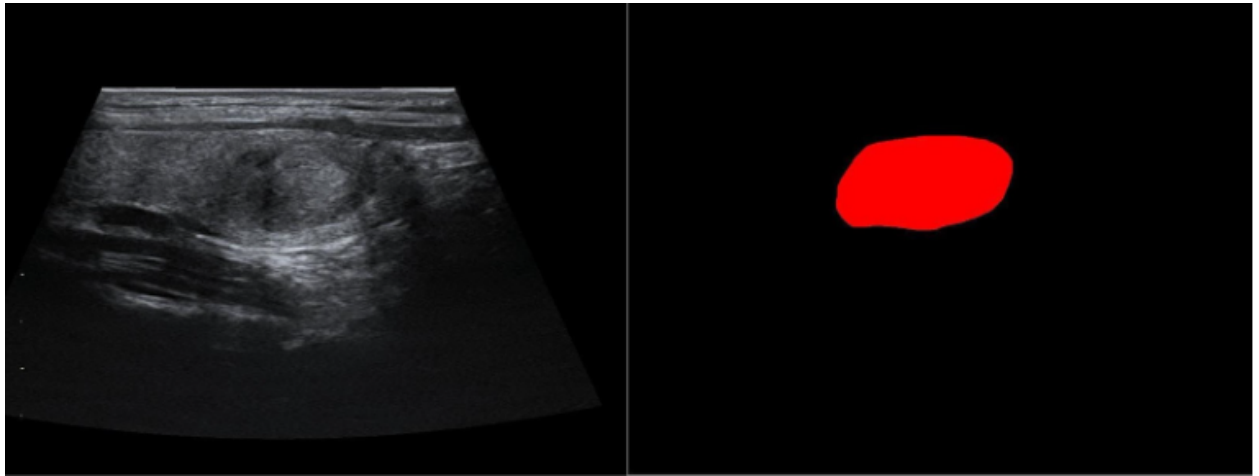


Figure 2.2. Example of Endocrinology dataset with highlighted formations areas of different types TIRADS nodules.

**Breast Ultrasound Images (BUSI) Dataset [1].** This data was collected to reduce the number of early deaths due to breast cancer, one of the most common causes of death in women worldwide, for different female patients. The dataset is represented using 1578 ultrasound scans of breast cancer along with the ground truth masks, which are categorized into three classes: with normal (266), benign (891) and malignant (421) tumours (Fig. 2.3).

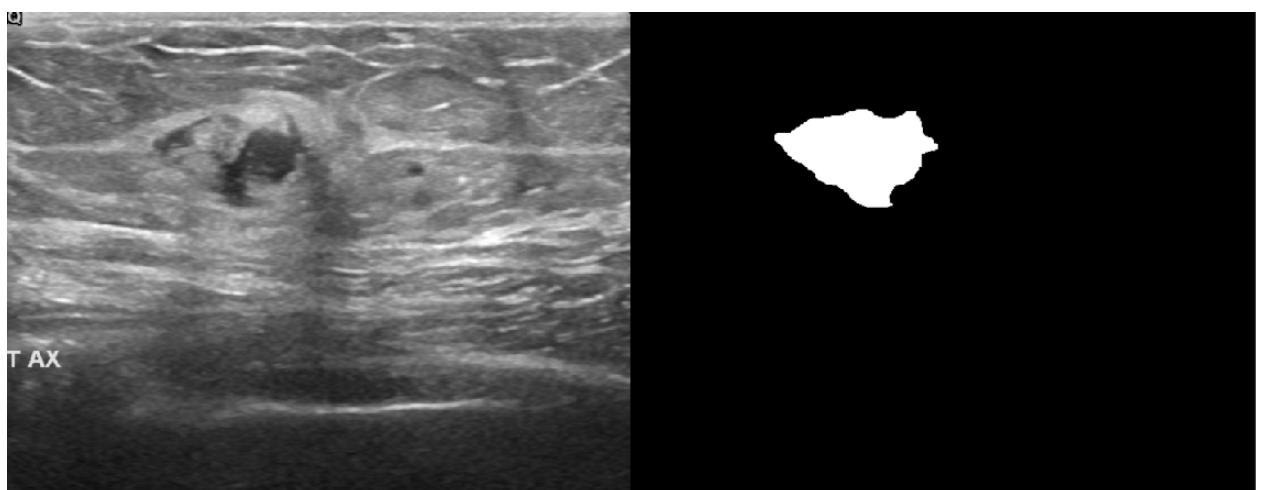


Figure 2.3. Example of BUSI dataset which is categorized into three classes: normal, benign, and malignant images.

**Brachial Plexus Ultrasound Images (BPUI) Dataset.** This data (5635 images and masks) at the kaggle<sup>2</sup> is formed to build a model that can identify nerve structures formed by the anterior branches of the four lower cervical nerves and the first thoracic nerve, called the brachial plexus, which is extended from the spinal cord through the cervicoaxillary canal in the neck, in ultrasound images for the segmentation task.

Localization of the nerve structures on the ultrasound scans is an important step in the clinical practice. The dataset contains both the images with the plexus and without, as well as the ground truth masks. However, as with all human-tagged data, one should expect artefacts and potential errors in the markings: for example, this data (Fig. 2.4) have been annotated by pseudo-experts (people who have been trained and instructed by experts).

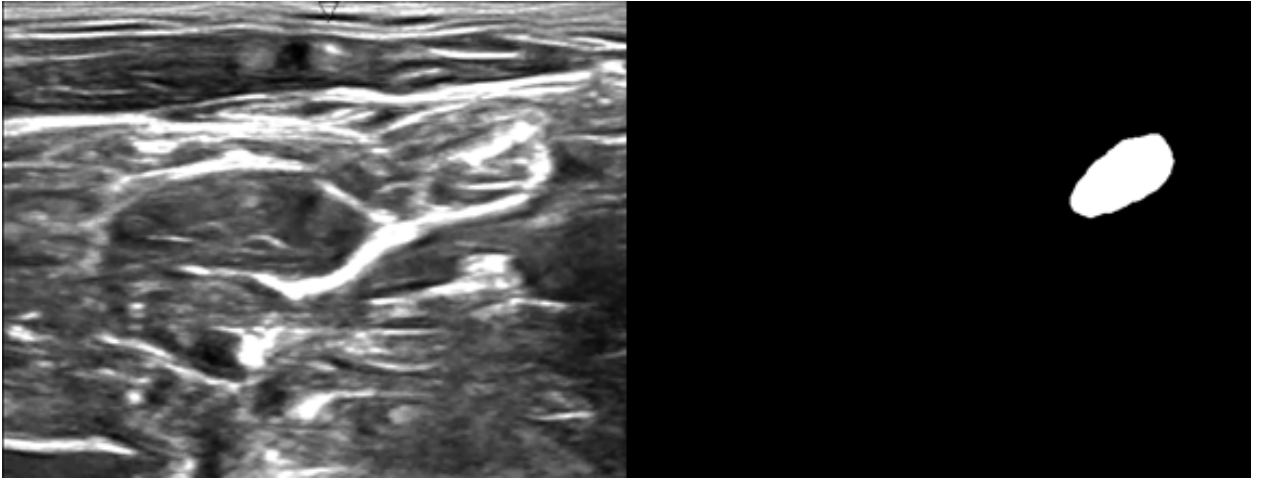


Figure 2.4. Collection of nerves called the Brachial Plexus (BP) in ultrasound images, where the nerve has been manually annotated by humans.

## 2.3 Method

The Fourier transform has the properties of linearity, preserving the accuracy of signal approximation (Parseval's theorem), which are preserved when the signal is represented using discrete vectors. The main method of visual analysis of Fourier transformation is to calculate its spectrum, that is coordinate-wise absolute value  $\|\hat{\mathbf{I}}(u, v)\|_2$ , or energy spectrum  $\|\hat{\mathbf{I}}(u, v)\|_2^2$ . To filter image in the frequency domain we need to take a function that modifies spectrum in a specific way. Another important property of the Fourier transform is described by the convolution theorem [3]: we can select both the kernel filter in the spatial domain and the global (kernel with the same dimensions as the image) filter in the frequency domain, which will do the same transformations, however, in the

<sup>2</sup><https://www.kaggle.com/c/ultrasound-nerve-segmentation/data>

first case we deprive ourselves of the possibility of applying direct frequency-dependent filtering functions as in the second case [23].

It is necessary to find a global filter, which will leave only those frequencies that improve the segmentation, classification or denoising metrics as a whole. We may decide on a high-pass filter, in which high frequencies are passed and low frequencies are significantly reduced, for example for edge detection, visualizing details or a low-pass filter for reduction of outliers and of contrast, that is a smoothing effect (Figures 1.2 and 2.5).

For each task, you need your own filter, which will not be easy to select manually. Therefore, we can give the task of filter selection to the computer by integrating the appropriate trainable denoising layer, which will be trained along with the entire model. The proposed design of the filtering layer shown in Figure 2.6 is capable of automatically forming a more sophisticated and task-specific filter (basic Fast Fourier Transform (FFT) and the element-wise multiplication functions in modern software packages are suitable).

**Spectrum Adjustment.** So, the proposed preprocessing method should automatically highlight significant frequencies, taking into account the model (since training occurs together with the selected base model), the dataset and the type of computer vision task. Spectrum adjustment will be done by adjusting the amplitude of each frequency separately, which is achieved using element-wise multiplication

$$\|\hat{\mathbf{I}}(u, v)\|_2 \leftarrow W(u, v) * \|\hat{\mathbf{I}}(u, v)\|_2 \quad \forall u, v, \quad (2.3.1)$$

where  $W$  is a weight matrix with non-negative elements, which is adaptive to each frequency, and  $*$  - element-wise multiplication.

**Handling Small Frequency Values.** Sometimes a lot of values in  $\|\hat{\mathbf{I}}(u, v)\|_2$  are rather small. For better visibility, the spectrum is normally log-transformed into  $\ln(1 + \|\hat{\mathbf{I}}(u, v)\|_2)$ . Thus, another idea is to learn the passing and/or amplification of certain frequencies as follows:

$$\|\hat{\mathbf{I}}(u, v)\|_2 \leftarrow \exp \left[ W(u, v) * \ln(1 + \|\hat{\mathbf{I}}(u, v)\|_2) \right] - 1 \quad \forall u, v, \quad (2.3.2)$$

where  $W$  - trained weight matrix (non-negative) and  $*$  - element-wise multiplication. This adjustment is designed to amplify low frequencies in the spectrum, the intensity of which is much less than the centre high frequencies (Fig. 1.1).

**Phase Adjustment.** We also looked at the phase adjustment, namely, the rotation of each frequency by the trainable angle, which can help to highlight the boundaries of significant areas in

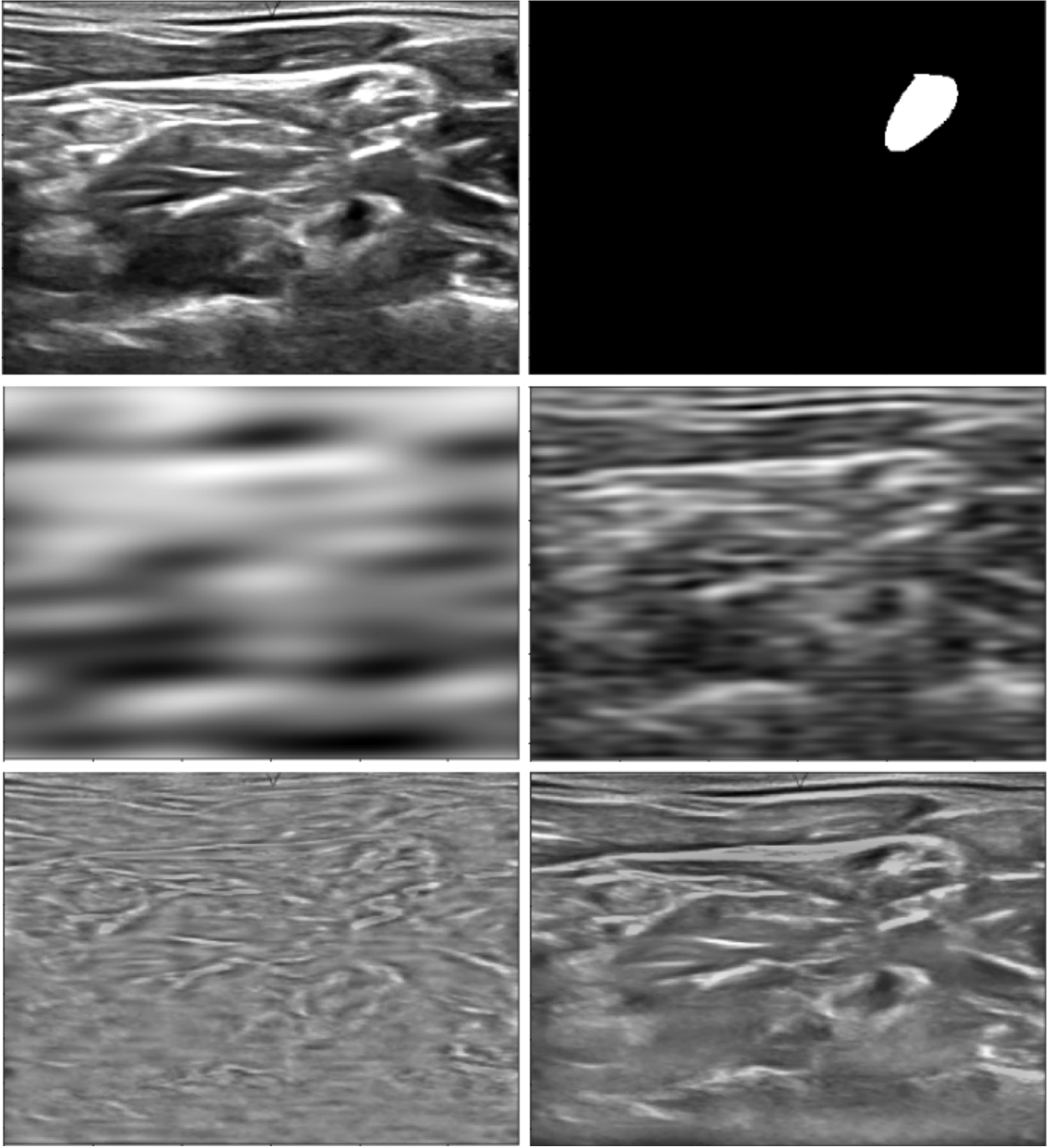


Figure 2.5. Top row: example of ultrasound image and corresponding mask. Middle row: filtered by low-pass filter with different thresholds image. Bottom row: filtered by high-pass filter with different thresholds image.

images [24, 51]. The weight matrix remains exactly the same size as in the configurations above, but now each weight encodes the angle for which the rotation operator is constructed for each frequency:

$$\begin{bmatrix} f \\ g \end{bmatrix}_{kl} \leftarrow \mathbf{R}_{w_{kl}} \begin{bmatrix} f \\ g \end{bmatrix}_{kl} = \begin{bmatrix} \cos w_{kl} & -\sin w_{kl} \\ \sin w_{kl} & \cos w_{kl} \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}_{kl} \quad (2.3.3)$$

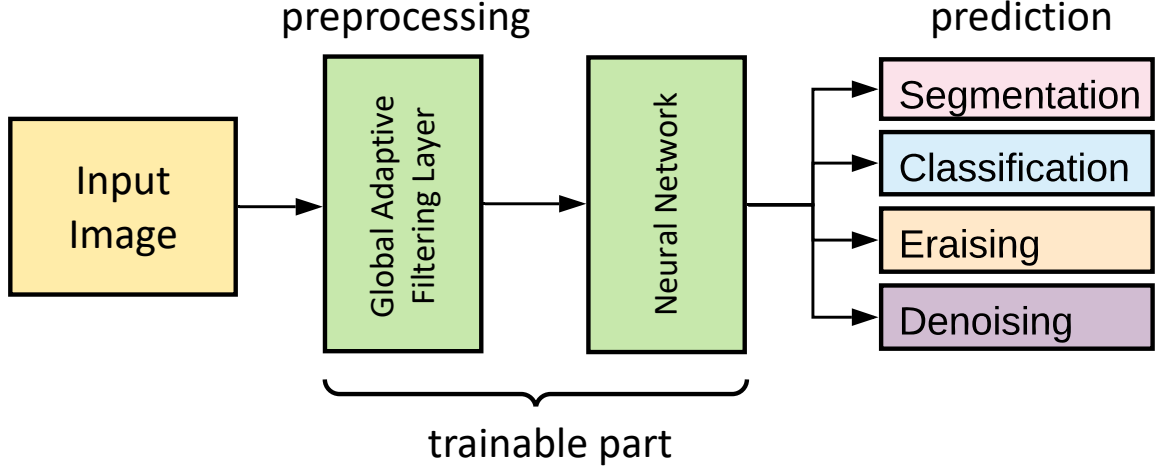


Figure 2.6. Diagram of the proposed method. Global Adaptive Filtering Layer is trained together with the weights of the main neural network until the prediction is maximized in a given CV task [42].

where  $f_{kl} + jg_{kl}$  is  $kl$ -th frequency in the Fourier transformed image,  $w_{kl}$  is  $kl$ -th weight in trainable weight matrix  $W$ .

A general algorithm for constructing a trainable adaptive layer based on the above provisions is described below:

---

**Algorithm 1:** Global Adaptive Filtering Layer

---

**Input:**  $I$  – Initial image  
 $\mathcal{F}$  – Real Fast Fourier Transform operator  
 $W$  – Weight Matrix

- 1:  $F = \mathcal{F}I$
- 2:  $I_{\text{adj}} = \text{adjustment}(F, W)$
- 3:  $I' = \mathcal{F}^{-1}I_{\text{adj}}$

**Output:**  $I'$  – Image after global frequency filtering

---

In the Appendix A you can find a more detailed description of the *adjustment* function for three different configurations: *spectrum* (Algorithm 2), *spectrum\_log* (Algorithm 3) and *phase* (Algorithm 4).

**Number of parameters.** Even if the image is real, its Fourier transform is usually complex. But there is an important symmetry property (Fig. 2.7):

$$\widehat{\mathbf{I}}(N_{\text{cols}} - u, N_{\text{rows}} - v) = \widehat{\mathbf{I}}(-u, -v) = \widehat{\mathbf{I}}(u, v)^* \quad (2.3.4)$$

which allows us to reduce the size of the image in the frequency space (and hence a global filter)



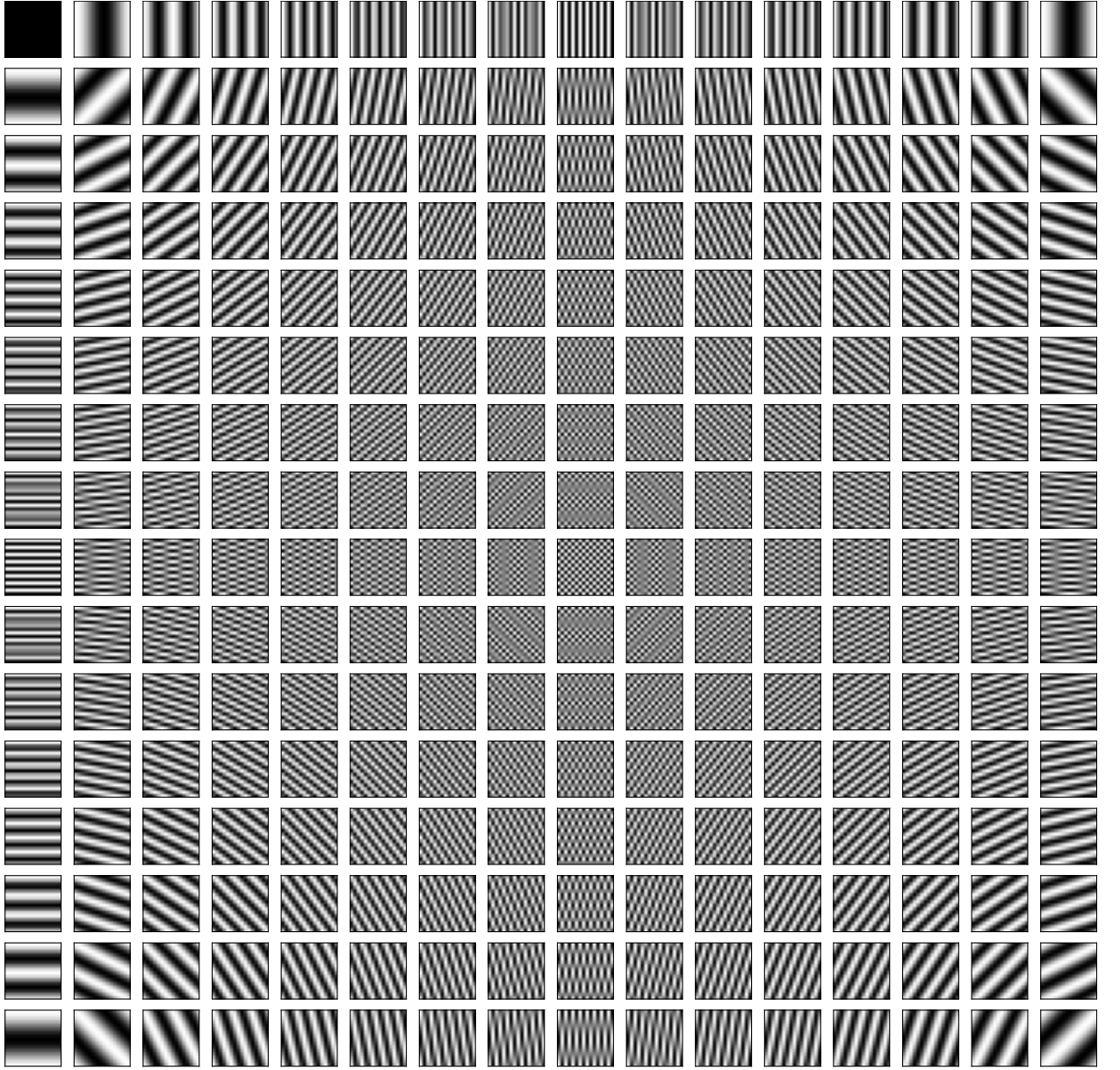


Figure 2.7. Example of the real part of the frequency basis for  $16 \times 16$  image.

and computation complexity by approximately 2 times. So, in Algorithm 1 each weight matrix is a tensor of size  $(C, N_{\text{cols}}, \lfloor N_{\text{rows}}/2 \rfloor + 1)$ , where  $C$  is the number of channels,  $(N_{\text{cols}}, N_{\text{rows}})$  is the image size. Thus, the number of learnable parameters of the suggested adaptive layer, considering that operations in the frequency space are element-wise, is equal to the  $C \cdot N_{\text{cols}} \cdot (\lfloor N_{\text{rows}}/2 \rfloor + 1)$ .

**General Spectrum Adjustment.** The above spectrum adjustment strategies are linear at each frequency. Theoretically, a neural network with just one hidden layer is enough to approximate an arbitrary nonlinear function [9, 15, 16, 27]. Therefore, the second idea is that the frequency filter can be any arbitrary function:

$$\|\hat{\mathbf{I}}(u, v)\|_2 \leftarrow W_2(u, v) * \sigma\left(W_1(u, v) * \|\hat{\mathbf{I}}(u, v)\|_2 + B_1(u, v)\right) + B_2(u, v) \quad \forall u, v, \quad (2.3.5)$$

where  $W_1, W_2$  - trained weight matrices (non-negative),  $B_1, B_2$  - trained bias matrices (non-negative),  $\sigma(\cdot)$  - some nonlinear activation function and  $*$  - element-wise multiplication.

The activation function used in the global filtering layer is a hyperparameter that needs to be selected depending on the problem being solved and the dataset. The function receives a non-negative matrix as input; it is since the frequency with negative weight has no physical interpretation. So, several popular activation functions have been selected and investigated: *ReLU* [29], *Sigmoid* [44], *ReLU6* [25], *Softplus* [34], *Tanh* [20], *Swish* (with  $\beta = 1.0$ ) [37] and *Mish* [33].

Further, the experiments' results on the above datasets will be presented, indicating the using models and described here proposed method.



## Chapter 3

# Results and Discussions

### 3.1 Experiments

To estimate the performance of the proposed method (of its various configurations), comparisons will be made as follows:

1. One of the previously described ultrasonic datasets is selected;
2. One of the computer vision problems is selected to be solved on this dataset: the segmentation, classification or denoising task;
3. Base model is selected and the number of parameters to be trained is calculated;
4. The same model in terms of architecture with the added proposed adaptive layer, which will have a similar or fewer number of parameters, is selected;
5. The two models are compared using the selected metric, number of parameters and computational complexity appropriate for the given task.

**Computational Complexity.** The computational complexity for base models is calculated using the software package. In the *spectrum* adaptive proposed layer, discrete Fourier transforms, and element-wise multiplications are used. To compute the number of operations in direct and inverse Fourier transforms, Split-radix [10, 11] algorithm estimation is used. Totally, *MAC operations* are equal to 2D DFT of size  $(N_{\text{cols}}, N_{\text{rows}})$  plus element-wise multiplication of weight matrix (global filter) of size  $(N_{\text{cols}}, \lfloor N_{\text{rows}}/2 \rfloor + 1)$  and frequency image:

$$\begin{aligned}
 \text{MACs} &\simeq \left[ N_{\text{cols}}(4N_{\text{rows}} \log_2 N_{\text{rows}} - 6N_{\text{rows}} + 8) + N_{\text{rows}}(4N_{\text{cols}} \log_2 N_{\text{cols}} - 6N_{\text{cols}} + 8) \right] + \\
 &\quad + N_{\text{cols}} \cdot (\lfloor N_{\text{rows}}/2 \rfloor + 1) = \\
 &= 4N_{\text{cols}}N_{\text{rows}} \log_2 N_{\text{cols}}N_{\text{rows}} - 12N_{\text{cols}}N_{\text{rows}} + 8(N_{\text{cols}} + N_{\text{rows}}) + \\
 &\quad + N_{\text{cols}} \cdot (\lfloor N_{\text{rows}}/2 \rfloor + 1)
 \end{aligned} \tag{3.1.1}$$

The formula 3.1.1 can be easily generalized to more multidimensional images, for example, with  $C > 1$  is the number of channels, taking into account the linearity in dimensions. Also, for other configurations (*spectrum log* or *phase*) where there are exponent or sine operations, an upper estimation is used to calculate the complexity.

To compare the models, a cross-validation technique [2] was used with disjoint train and test sets. Further, for all models, statistical averages are mainly compared (the learning curves plots also show the dispersion).

**Hyperparameters.** All models are trained setting the *batch size* = 4, and using *Adam optimizer* [22] with *learning rate* = 0.001. The technique of adaptive step reduction on the plateau was also used (if the value of the loss function on the validation set did not reduce for some epochs).

**Segmentation.** To test the proposed method, U-Net [40] - the first successful architecture in biomedical image segmentation was taken as a basis. The core idea behind U-Net is to use the advantage of rich feature representation from convolutional networks and combine it with lower-level image representation (that is solved by using skip connections). Models with DenseNet [17], which contains shorter connections between layers close to the input and those close to the output for more accurate and efficient teaching, and ResNet [13], where the layers are reformulated as learning residual functions regarding the layer inputs, instead of learning the unreferenced functions, encoders quite fully considered in the paper [42].

For the learning process, the *Combined Loss* function of *Dice* and *Cross Entropy*, weighted as 0.6 and 0.4 respectively, is used. The quality of segmentation is evaluated with the *Dice coefficient* [32], which, in essence, measures the overlap between the predicted and the ground-truth masks.

For the segmentation task, U-Net with *init\_features* = 32 was compared against U-Net with *init\_features* = 16 along with adaptive layer for Endocrinology, BUSI and BPUI datasets with *image\_size* = (512, 512), yielding 467,266 parameters in base model vs. 248,994 parameters in proposed model. The segmentation results are presented in Table 3.1 and Figures 3.1, 3.2 (Figures B.1, B.2 in the Appendix B).

To test the difference with the *general* configuration U-Net with *init\_features* = 64 was compared against U-Net with *init\_features* = 32 along with general adaptive layer for Endocrinology, BUSI and BPUI datasets. All of the experiments were carried out using *ReLU* activation function (see section 3.2 for details). These results are presented in Table 3.2 and Figure 3.3 (Figures B.3, B.4 in the Appendix B).

Table 3.1. Segmentation Results I. The number of parameters, MAC operations and Dice metrics for different models on Endocrinology, BUSI and BPUI validation sets.

Model	# params	MACs	Endocrinology	BUSI	BPUI
U-Net [40]	0.467 M	28.79 G	0.69	0.64	0.59
U-Net + <i>spectrum</i> adjustment [42]	0.249 M	7.3 G	0.73	0.68	0.63
U-Net + <i>spectrum log</i> adjustment [42]	0.249 M	7.31 G	0.74	0.68	0.63
U-Net + <i>phase</i> adjustment [42]	0.249 M	7.31 G	0.69	0.65	0.62

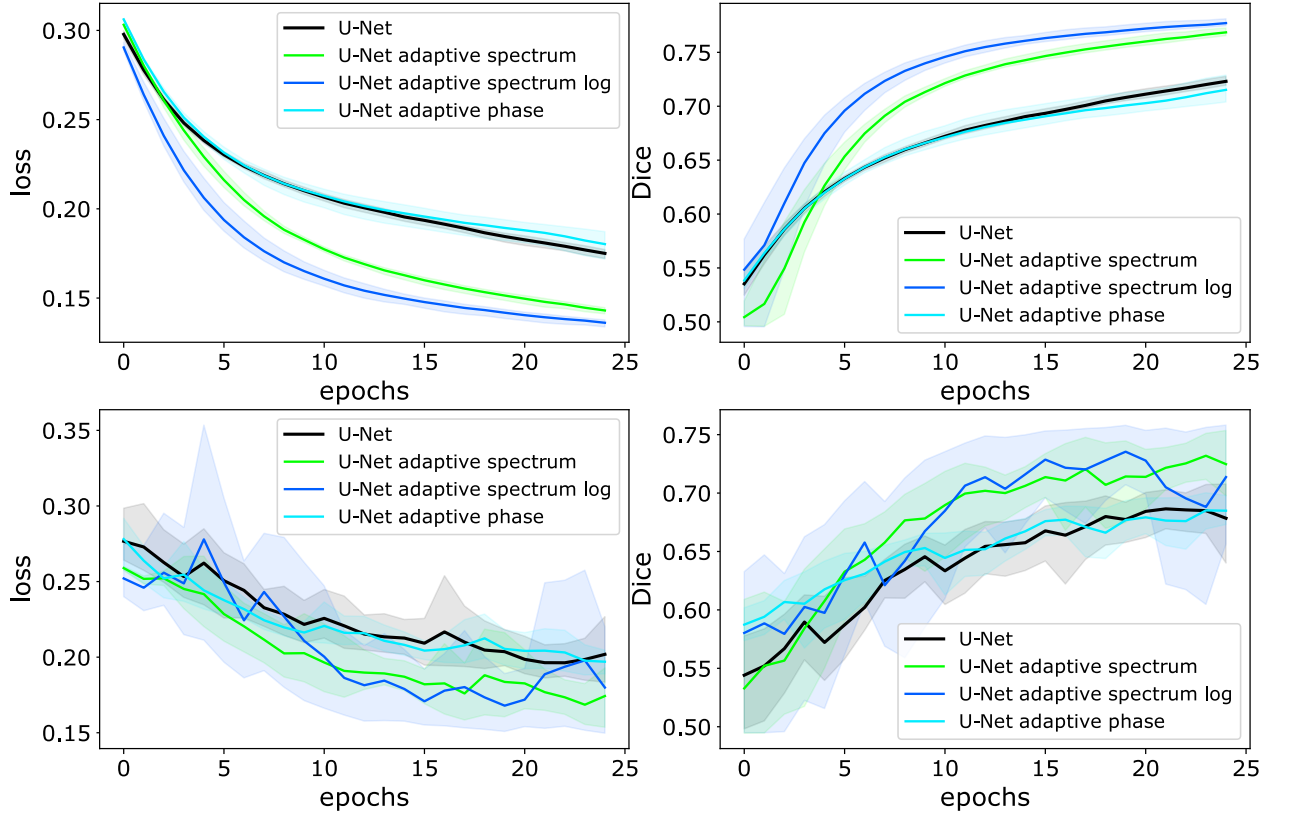


Figure 3.1. Segmentation Results I. Top row: Loss function values and Dice metrics on Endocrinology train sets. Bottom row: the same on validation sets.

Table 3.2. Segmentation Results II. The number of parameters, MAC operations and Dice metrics for different models on Endocrinology, BUSI and BPUI validation sets.

Model	# params	MACs	Endocrinology	BUSI	BPUI
U-Net [40]	1.864 M	114.49 G	0.72	0.64	0.60
U-Net + <i>general spectrum</i> adjustment [42]	0.994 M	28.82 G	0.75	0.69	0.63

**Classification.** To verify the suggested algorithm for the classification problem, ResNet (as resnet-20) with *blocks* = 3, 3, 3 (numbers of blocks stacked at the building blocks) and *filters* = 16, 32, 64 (numbers of filters) was compared with the model with *filters* = 8, 32, 64, pre-pended by the adaptive layer with *image\_size* = (128, 128), yielding 272,566 vs. 267,454 parameters for

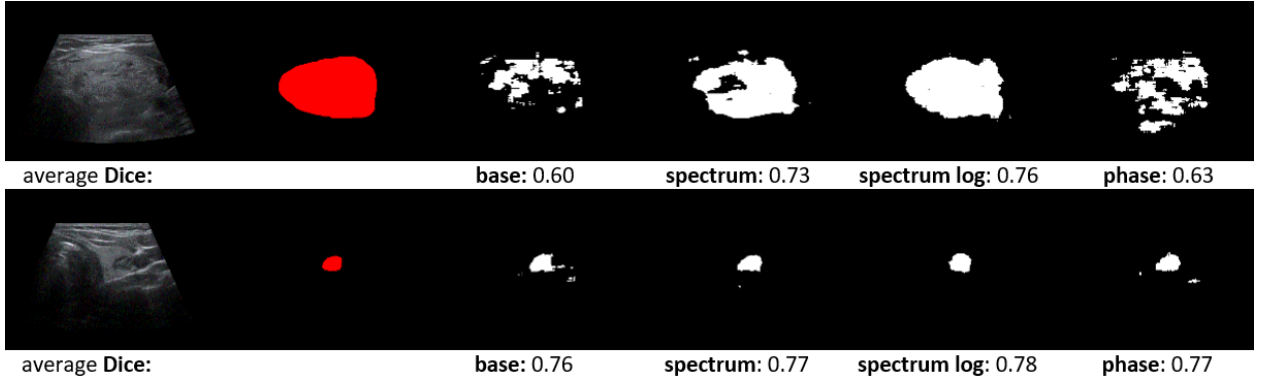


Figure 3.2. Segmentation Results I (Visual). From left to right: ultrasound slice from Endocrinology dataset, ground truth mask, segmentation result performed by base model and linear adaptive layer configurations.

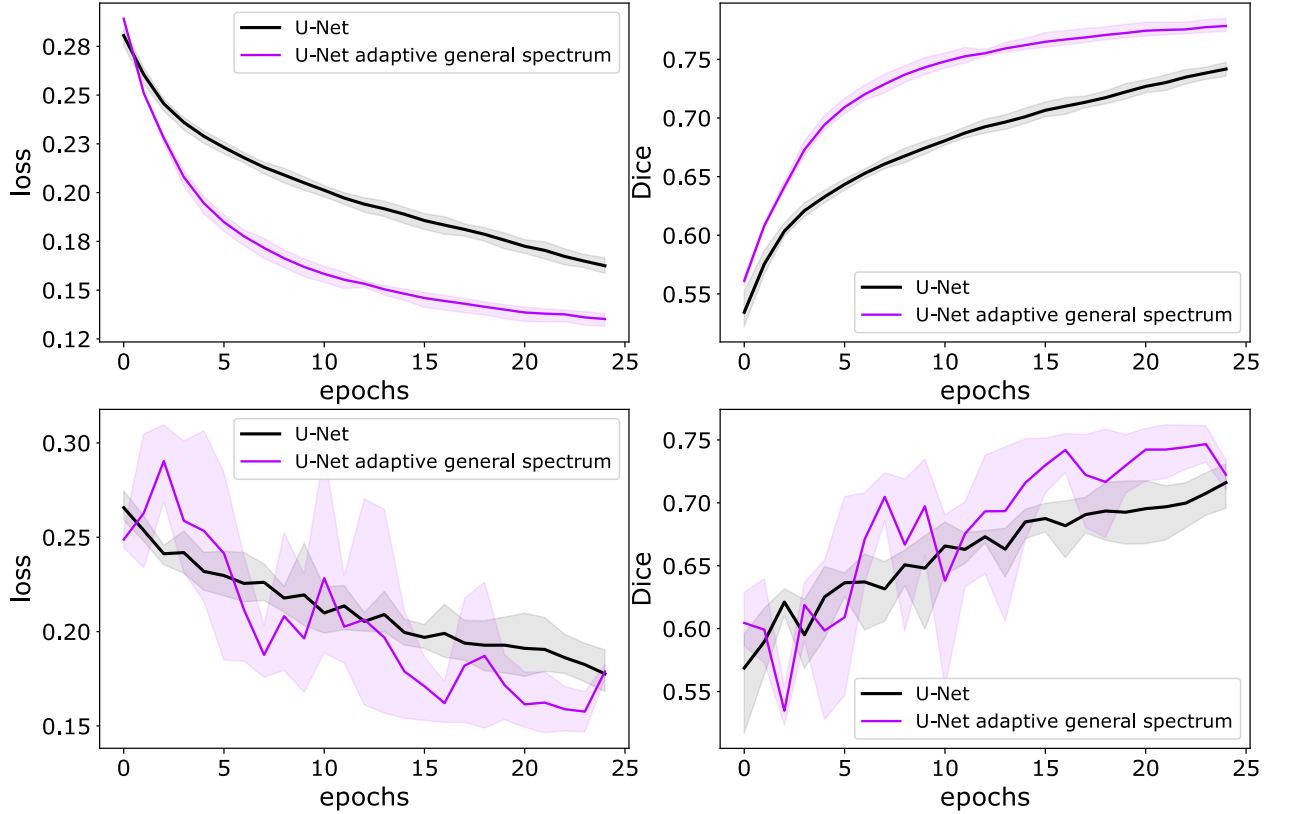


Figure 3.3. Segmentation Results II. Top row: Loss function values and Dice metrics on Endocrinology train sets. Bottom row: the same on validation sets.

Endocrinology dataset and 272,371 vs. 267,259 for BUSI dataset respectively.

The training process is similar to the one above, with using the *weighted Cross Entropy Loss* [14] combined with the  $F_1$ -score evaluation, which is the harmonic mean of *Precision* and *Recall* and better takes into account class imbalance and misclassified objects by the model than the *Accuracy* metric [45]. The results for this task are presented in Table 3.3 and Figure 3.4 (Figure C.1 in the Appendix C).

Table 3.3. Classification Results. The number of parameters, MAC operations and  $F_1$ -scores for different models on Endocrinology and BUSI validation sets.

Model	# params	MACs	Endocrinology	BUSI
resnet-20 [13]	0.272 M	44.4 M	0.95	0.67
ResNet + <i>spectrum</i> adjustment [42]	0.267 M	32 M	0.95	0.64
ResNet + <i>spectrum log</i> adjustment [42]	0.267 M	32.5 M	0.95	0.65
ResNet + <i>phase</i> adjustment [42]	0.267 M	32.5 M	0.95	0.68

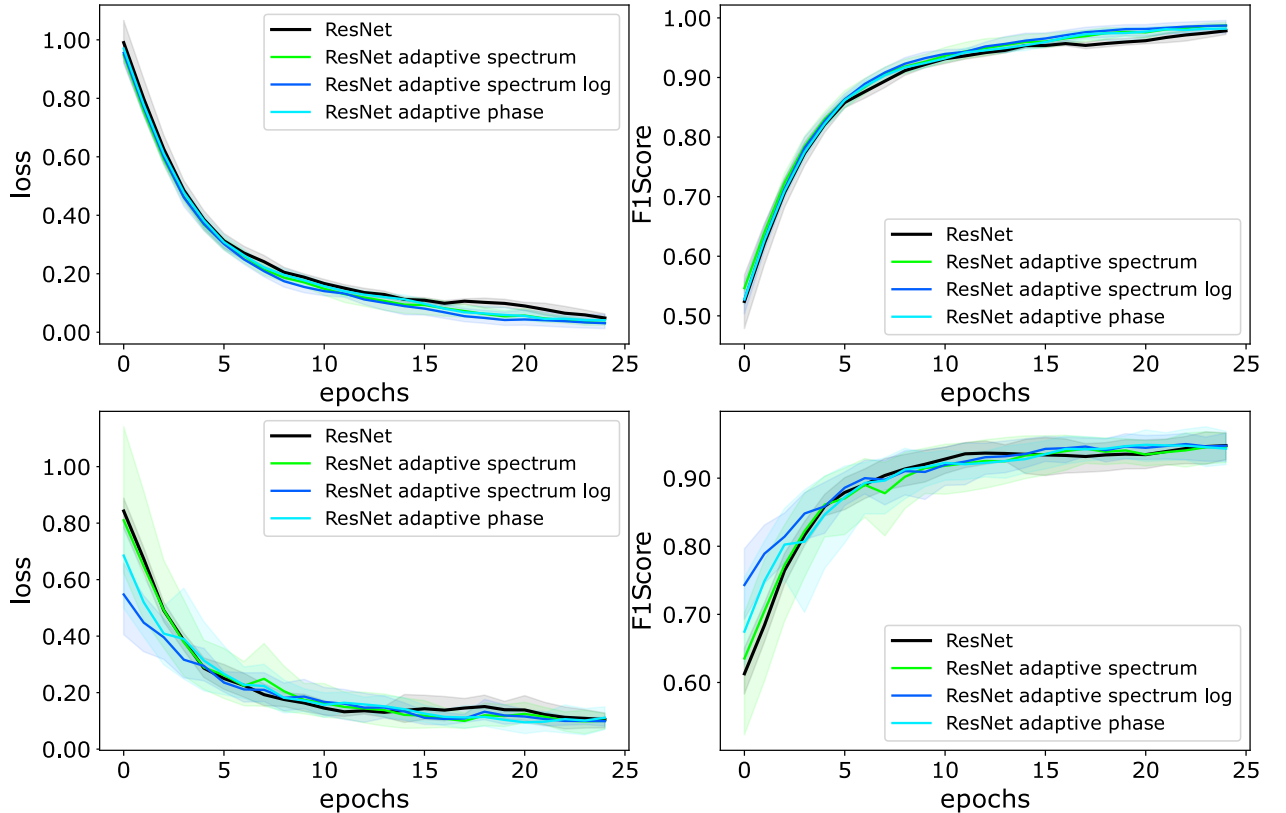


Figure 3.4. Classification Results I. Top row: Loss function values and  $F_1$ -scores on Endocrinology train sets. Bottom row: the same on validation sets.

**Denoising and Erasing.** For the problem of denoising and erasing, the popular model DnCNN [56] is considered as the baseline, the main task of which is to restore the noise, in contrast to the standard feedforward models, which restore the image. To assess the success of denoising, *Combined Loss* function of *MS-SSIM* [53] and  $L_1$  Loss with weights 0.8 and 0.2 respectively [57] is used along with the *SSIM* [54], which is structural similarity index, and *PSNR*, which is essentially the ratio of the maximum possible signal value over the noise level, metrics [21].

For these tasks, DnCNN with *init\_features* = 32 and *num\_layers* = 20 was compared to itself with *init\_features* = 16 and *num\_layers* = 17 and with the pre-pended adaptive layer for BUSI dataset with *image\_size* = (512, 512), yielding 168,225 vs. 167,169 parameters.

Two different modes of noise pollution were chosen to verify efficiency: regular Gaus-

sian noise and rectangular erasing corruption [58]. The former simulated typical noise present in sampling-based medical scanners (both ultrasound and others, for example, MRI); the latter mimicked the corruption on an image when part of the image may be lost or blocked and is used to test how properly and similarly the model reconstructs the structure based on the area around. The denoising results are highlighted in Table 3.4 and Figure 3.5 (Figures D.1, D.2 in the Appendix D).

Table 3.4. Denoising and Erasing Results. The number of parameters, MAC operations, PSNR and SSIM metrics for different models on BUSI validation set for Gaussian denoising and rectangular erasing tasks.

Model	# params	MACs	Denoising		Erasing	
			PSNR	SSIM	PSNR	SSIM
DnCNN [56]	0.168 M	44.26 G	30.92	0.83	23.61	0.88
DnCNN + <i>spectrum</i> adjustment [42]	0.167 M	9.41 G	30.79	0.83	23.59	0.88
DnCNN + <i>spectrum log</i> adjustment [42]	0.167 M	9.42 G	30.89	0.83	24.71	0.87
DnCNN + <i>phase</i> adjustment [42]	0.167 M	9.42 G	30.78	0.82	22.96	0.88

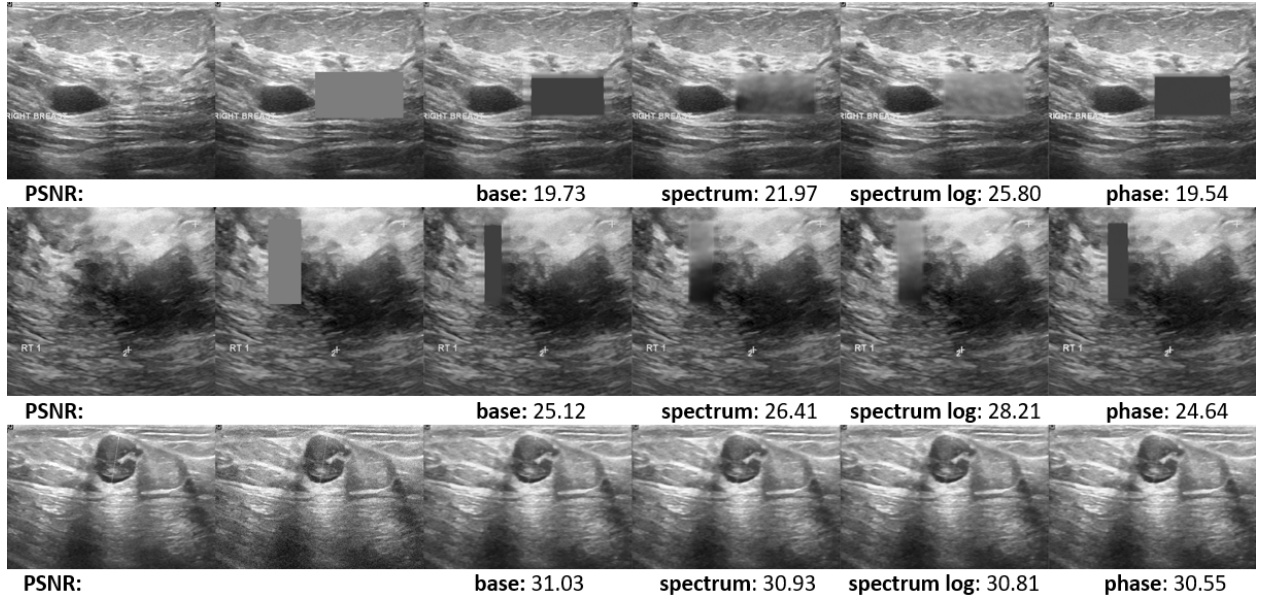


Figure 3.5. Denoising and Erasing Results. From left to right: ultrasound slice from BUSI dataset, corrupted image, denoising/erasing result performed by base model and linear adaptive layer configurations.

## 3.2 Discussions

**Medical Data.** Such data require more focus, dedication, and expertise to annotate them with labels. But, as with all human-tagged data, one should expect artefacts and potential errors in the markings: for example, the BPUI data have been annotated by pseudo-experts (people who have been trained and instructed by experts).



**Weights Initialization.** The initialization of weight matrices in *linear* configurations is a hyperparameter that should also be investigated. In this work, such an initialization was chosen to obtain an identical transformation when applied (see Algorithms in the Appendix A for details). Therefore, a matrix of **ones** was chosen for *spectrum*, *spectrum log* adjustments and a matrix of **zeros** was chosen for *phase* adjustment.

**Activation Functions.** The Table 3.5 contains the results of the study of the activation function influence in the *general spectrum* configuration of the adaptive layer for BUSI and BPUI datasets. Models are trained using the *image size* = (256, 256). As can be seen from the results, the activation function plays an important role. The average difference between the best and worst activation function in some cases is about 10% by metric. It should be noted that the activation functions *Mish* and *ReLU* have shown themselves well on all datasets. So *Mish* and *ReLU* are good for using the algorithm out of the box.

Table 3.5. Validation Dice score for different activation functions on the BUSI and BPUI datasets in the *general* configuration of the adaptive layer. Red color font corresponds to the worst score, and green font corresponds to the best score.

Activation Function	BUSI			BPUI		
	U-Net	DenseNet	ResNet	U-Net	DenseNet	ResNet
Sigmoid	<b>0.70</b>	0.82	0.83	<b>0.68</b>	<b>0.75</b>	<b>0.72</b>
<b>ReLU</b>	0.77	<b>0.84</b>	<b>0.83</b>	<b>0.75</b>	<b>0.77</b>	<b>0.75</b>
ReLU6	0.76	0.83	0.81	0.73	0.76	0.74
Softplus	0.74	0.83	0.82	0.70	<b>0.75</b>	0.74
Tanh	0.75	0.83	<b>0.83</b>	0.73	0.76	0.74
Swish ( $\beta = 1.0$ )	<b>0.78</b>	<b>0.81</b>	<b>0.80</b>	0.74	0.76	<b>0.75</b>
<b>Mish</b>	<b>0.78</b>	0.83	0.82	<b>0.75</b>	<b>0.77</b>	<b>0.75</b>

**Denoising and Erasing.** There was not observe a stunning improvement in the denoising problem built around the DnCNN model. Additional studies are required with the DnCNN to understand the enhancement dynamics; however, the same very model is well improved by our layer when there is a notable corruption (the erasing problem). Erasing is a common task for MRI studies. MRI examination can take about an hour, and the patient must fix his body position at this time; it is very inconvenient, especially when the patient is a child. Therefore, there are ways to speed up the MRI procedure by lowering the resolution of the images, forming broken pixies (erased pixels), a kind of problem where can be a useful proposed adaptive Fourier layer. We can visually confirm (Figure 3.5) that our adaptive configuration denoises and «heals» the corruptions better than the base DnCNN model alone.

# Conclusion

The method proposed in this research work proved to be efficient on all datasets and all classical model architectures that there have considered. A model with an added adaptive layer has fewer parameters and lower computational complexity but at the same time gives a comparable or better result in terms of the metrics considered for a specific CV problem. This expands the possible areas of application of the proposed approach and opens up a new direction of research of adaptive layers (*e.g.*, in more complex multi-layer architectures [7], in generative and image translation models [36], in learnable frequency kernels [26], in iterative anomaly detection models [50], *etc.*). Lightweighting these models via a simple adaptive filtering layer holds the potential for seamless integration in various CV applications.

A rather important finding is the increase of Dice score for the case of simple U-Net by around 4 – 5% when the adaptive global filter is added. This promises an opportunity for the areas, such as medicine, where getting marked data is an acknowledged challenge, causing one to attempt learning on small datasets. In these cases, the use of heavy models with a large number of parameters is one possible solution that frequently leads to a fast overfit; whereas, the addition of simple adaptive filtering layers «trims» unnecessary frequencies in the Fourier domain and makes the model learn only the vital frequencies along with the weights of the main neural network. All of this is accomplished while optimizing the targeted advantage function of interest to a particular application (*e.g.*, Dice score or  $F_1$ -score).

The proposed layer can be a good add-on for several powerful modern preprocessing tools, including those that exist in various AutoML pipelines [8]. In fact, there are many recently devised methods of augmentation and data preprocessing (*e.g.*, see [5]) that still await frequency filtering capability. There could find no other articles reporting Fourier filter preprocessing trained alongside the main network.



# Bibliography

- [1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 11 2019.
- [2] Sylvain Arlot, Alain Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.
- [3] J.M. Blackledge. *Digital Image Processing: Mathematical and Computational Methods*. 11 2005.
- [4] Steven Brunton and J. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. 01 2019.
- [5] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2):125, Feb 2020.
- [6] Vikas Chaudhary and Shahina Bano. Thyroid ultrasound. *Indian journal of endocrinology and metabolism*, 17:219–27, 03 2013.
- [7] Aritra Chowdhury, Dmitry V Dylov, Qing Li, Michael MacDonald, Dan E Meyer, Michael Marino, and Alberto Santamaria-Pang. Blood vessel characterization using virtual 3d models and convolutional neural networks in fluorescence microscopy. pages 629–632. IEEE, 2017.
- [8] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data. 2019.
- [9] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [10] Pierre Duhamel and Henk Hollmann. Split radix’fft algorithm. *Electronics letters*, 20(1):14–16, 1984.
- [11] Pierre Duhamel and Martin Vetterli. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.

- [12] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27 – 48, 2016. Recent Developments on Deep Big Vision.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [14] Yaoshiang Ho and Samuel Wookey. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, PP:1–1, 12 2019.
- [15] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.
- [16] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [17] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.
- [18] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Zhang Qiaowei, Yutaro Iwamoto, Xian-Hua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. pages 1055–1059, 05 2020.
- [19] Barys Ihnatsenka and Andre Boezaart. Ultrasound: Basic understanding and learning the language. *International journal of shoulder surgery*, 4:55–62, 07 2010.
- [20] Barry L Kalman and Stan C Kwasny. Why tanh: choosing a sigmoidal function. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 4, pages 578–581. IEEE, 1992.
- [21] Sergey Kastyulin, Dzhamil Zakirov, and Denis Prokopenko. PyTorch Image Quality: Metrics and measure for image quality assessment, 2019. Open-source software available at <https://github.com/photosynthesis-team/piq>.
- [22] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.
- [23] Reinhard Klette. *Concise Computer Vision - An Introduction into Theory and Algorithms*. 01 2014.
- [24] Jelena Kovacevic, Vivek K Goyal, and Martin Vetterli. Fourier and wavelet signal processing. *Fourier and Wavelets. org*, pages 1–294, 2013.

- [25] Alex Krizhevsky and Geoff Hinton. Convolutional deep belief networks on cifar-10. *Unpublished manuscript*, 40(7):1–9, 2010.
- [26] Elizaveta Lazareva, Oleg Rogov, Olga Shegai, Denis Larionov, and Dmitry V Dylov. Learnable hollow kernels for anatomical segmentation. *arXiv preprint arXiv:2007.05103*, 2020.
- [27] Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867, 1993.
- [28] Jianghong Li, Liang Chen, and Yuanhu Cai. Dynamic texture segmentation using fourier transform. *Modern Applied Science*, 3, 08 2009.
- [29] Yuanzhi Li and Yang Yuan. Convergence analysis of two-layer neural networks with relu activation. *arXiv preprint arXiv:1705.09886*, 2017.
- [30] Shengfeng Liu, Yi Wang, Xin Yang, Shawn Li, Tianfu Wang, Baiying Lei, Dong Ni, and Li Liu. Deep learning in medical ultrasound analysis: A review. *Engineering*, 5:261–275, 03 2019.
- [31] Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.
- [32] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 06 2016.
- [33] Diganta Misra. Mish: A self regularized non-monotonic neural activation function. *arXiv preprint arXiv:1908.08681*, 4, 2019.
- [34] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.
- [35] Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. 04 2018.
- [36] Denis Prokopenko, Joël Valentin Stadelmann, Heinrich Schulz, Steffen Renisch, and Dmitry V Dylov. Unpaired synthetic image generation in radiology using gans. pages 94–101. Springer, Cham, 2019.
- [37] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

- [38] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352–2449, 2017. PMID: 28599112.
- [39] M. Rhu, N. Gimelshein, J. Clemons, A. Zulfiqar, and S. W. Keckler. vdn: Virtualized deep neural networks for scalable, memory-efficient neural network design. In *2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 1–13, 2016.
- [40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 05 2015.
- [41] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarczyk, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, 2017.
- [42] Viktor Shipitsin, Iaroslav Bespalov, and Dmitry V Dylov. Global adaptive filtering layer for computer vision. *arXiv preprint arXiv:2010.01177v3*, March 2021.
- [43] Connor Shorten and Taghi Khoshgohar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 12 2019.
- [44] P Sibi, S Allwyn Jones, and P Siddarth. Analysis of different activation functions using back propagation neural networks. *Journal of theoretical and applied information technology*, 47(3):1264–1268, 2013.
- [45] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. Springer, 2006.
- [46] Junho Song, Young Jun Chai, Hiroo Masuoka, Sun-Won Park, Su-jin Kim, June Choi, Hyoun-Joong Kong, Kyu Eun Lee, Joongseek Lee, Nojun Kwak, Ka Yi, and Akira Miyauchi. Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules. *Medicine*, 98:e15133, 04 2019.
- [47] Richard Szeliski. *Computer vision algorithms and applications*. Springer, London; New York, 2011.
- [48] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains, 06 2020.

- [49] Franklin Tessler, William Middleton, and Edward Grant. Thyroid imaging reporting and data system (ti-rads): A user’s guide. *Radiology*, 287:1082, 06 2018.
- [50] Nina Tuluptceva, Bart Bakker, Irina Fedulova, Heinrich Schulz, and Dmitry V Dylov. Anomaly detection with deep perceptual autoencoders. *arXiv preprint arXiv:2006.13265*, 2020.
- [51] Martin Vetterli, Jelena Kovačević, and Vivek K Goyal. *Foundations of signal processing*. Cambridge University Press, 2014.
- [52] Lei Wang, Shujian Yang, Shan Yang, Cheng Zhao, Guangye Tian, Yuxiu Gao, Yongjian Chen, and Yun Lu. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the yolov2 neural network. *World Journal of Surgical Oncology*, 17, 12 2019.
- [53] Z. Wang, Eero Simoncelli, and Alan Bovik. Multiscale structural similarity for image quality assessment. volume 2, pages 1398 – 1402 Vol.2, 12 2003.
- [54] Zhou Wang, Alan Bovik, Hamid Sheikh, and Eero Simoncelli. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13:600 – 612, 05 2004.
- [55] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation, 04 2020.
- [56] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.
- [57] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, PP:1–1, 12 2016.
- [58] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34, 08 2017.
- [59] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. 07 2018.
- [60] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 12 2019.

- [61] Wangmeng Zuo, Kai Zhang, and Lei Zhang. Convolutional neural networks for image denoising and restoration. *Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends*, pages 93–123, 2018.

## Appendix A

# Adjustment Algorithms

---

**Algorithm 2:** Spectrum Adjustment

---

**Input:**  $F$  – Real Fourier Transform of Image  
 $W$  – Weight Matrix

- 1:  $W = ReLU(W)$
- 2:  $F' = W * F$

**Output:**  $F'$  – Adjusted Image after global frequency filtering.

---

---

**Algorithm 3:** Spectrum Log Adjustment

---

**Input:**  $F$  – Real Fourier Transform of Image  
 $W$  – Weight Matrix

- 1:  $W = ReLU(W)$
- 2:  $S = \exp[W * \ln(1 + \|F\|_2)] - 1$
- 3:  $F' = S * F / \|F\|_2$

**Output:**  $F'$  – Adjusted Image after global frequency filtering.

---

---

**Algorithm 4:** Phase Adjustment

---

**Input:**  $F$  – Real Fourier Transform of Image  
 $W$  – Weight Matrix

- 1:  $W = clamp(W, 0, 2\pi)$
- 2:  $F' = \cos W * F_{\text{real}} - \sin W * F_{\text{imag}} + j(\sin W * F_{\text{real}} + \cos W * F_{\text{imag}})$

**Output:**  $F'$  – Adjusted Image after global frequency filtering.

---

## Appendix B

# Segmentation Results

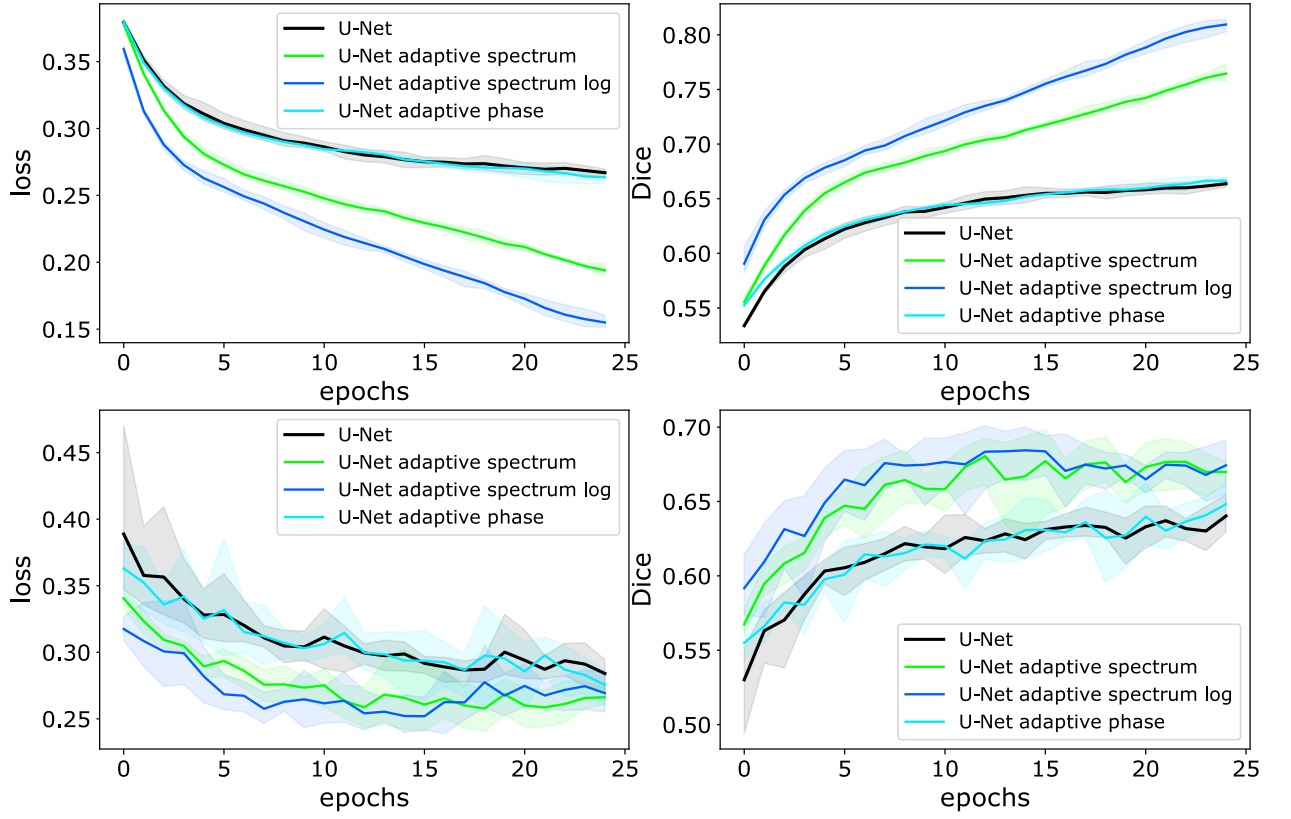


Figure B.1. Segmentation Results III. Top row: Loss function values and Dice metrics on BUSI train sets. Bottom row: the same on validation sets.



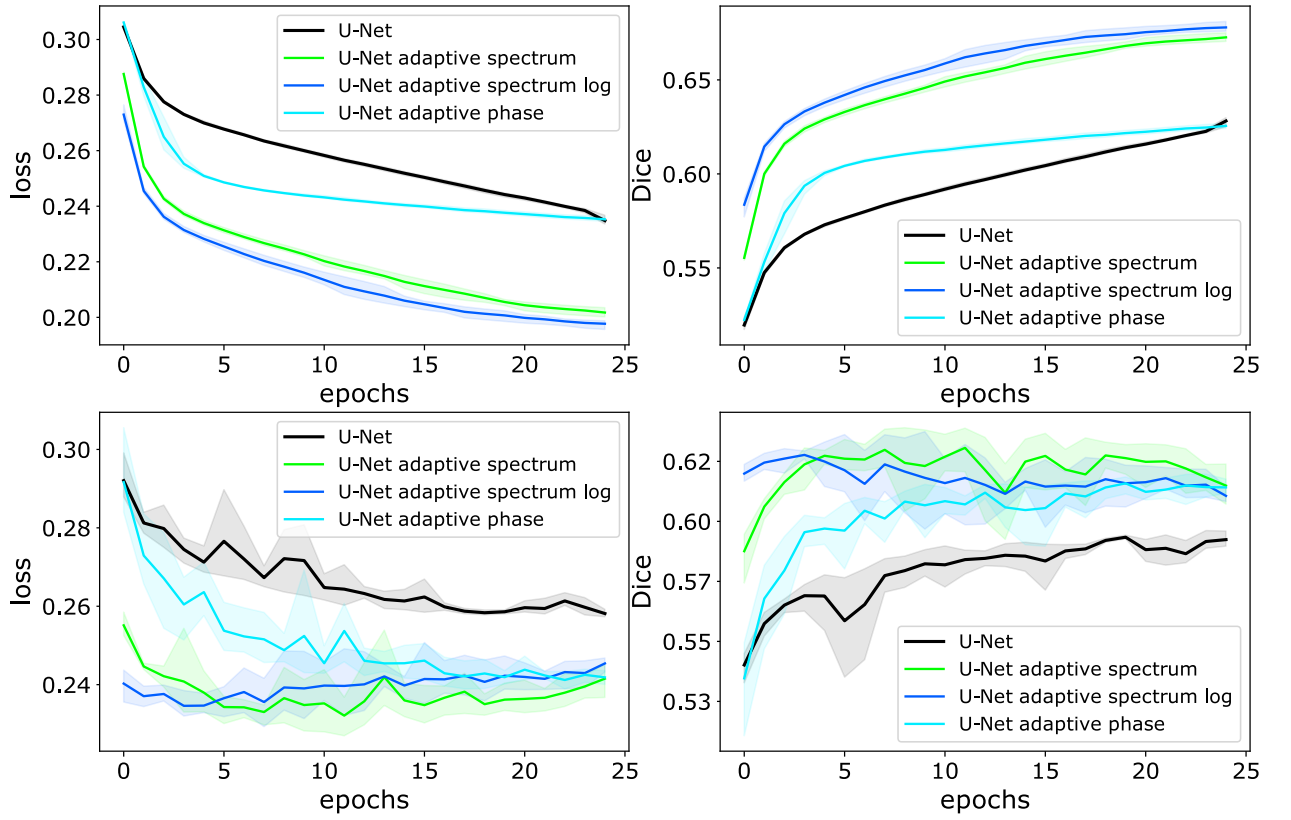


Figure B.2. Segmentation Results IV. Top row: Loss function values and Dice metrics on BPUI train sets. Bottom row: the same on validation sets.

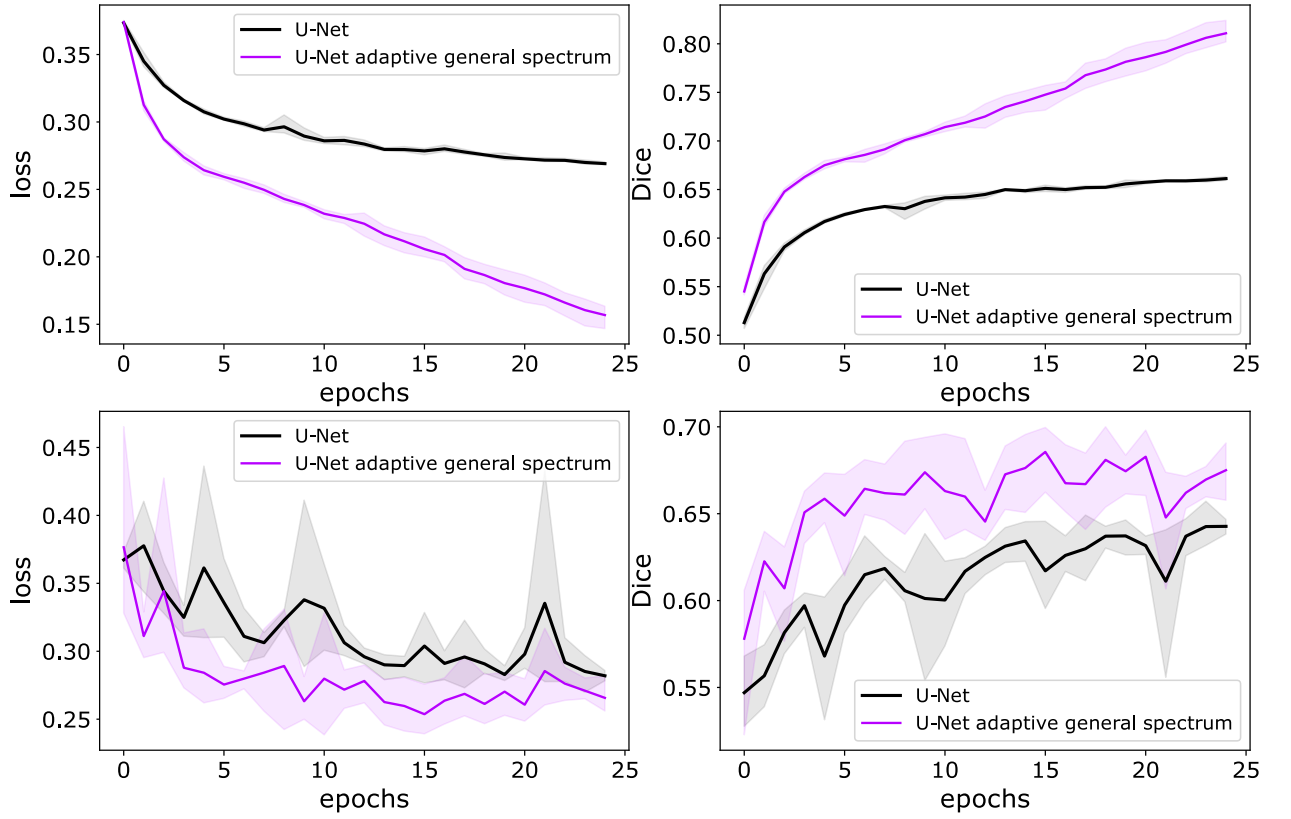


Figure B.3. Segmentation Results V. Top row: Loss function values and Dice metrics on BUSI train sets. Bottom row: the same on validation sets.

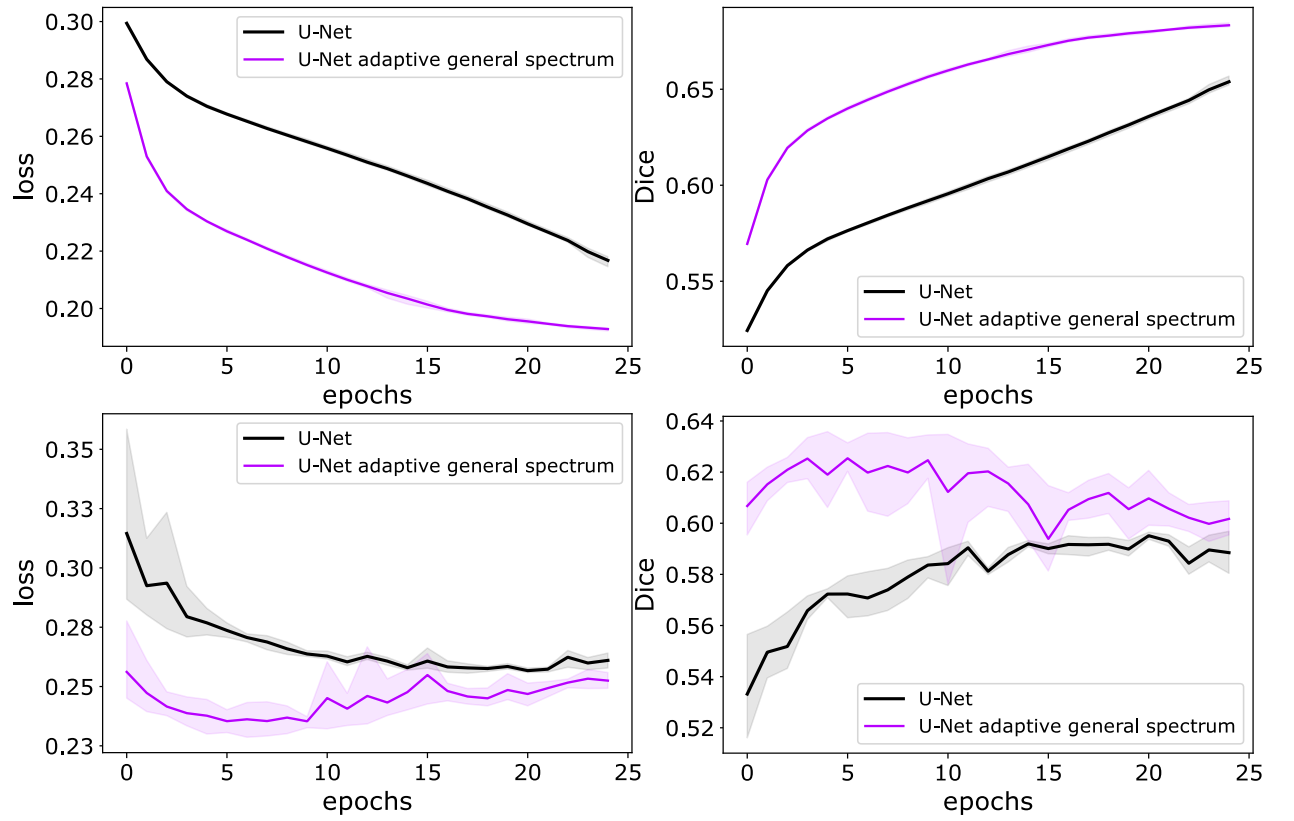


Figure B.4. Segmentation Results VI. Top row: Loss function values and Dice metrics on BPUI train sets. Bottom row: the same on validation sets.

## Appendix C

# Classification Results

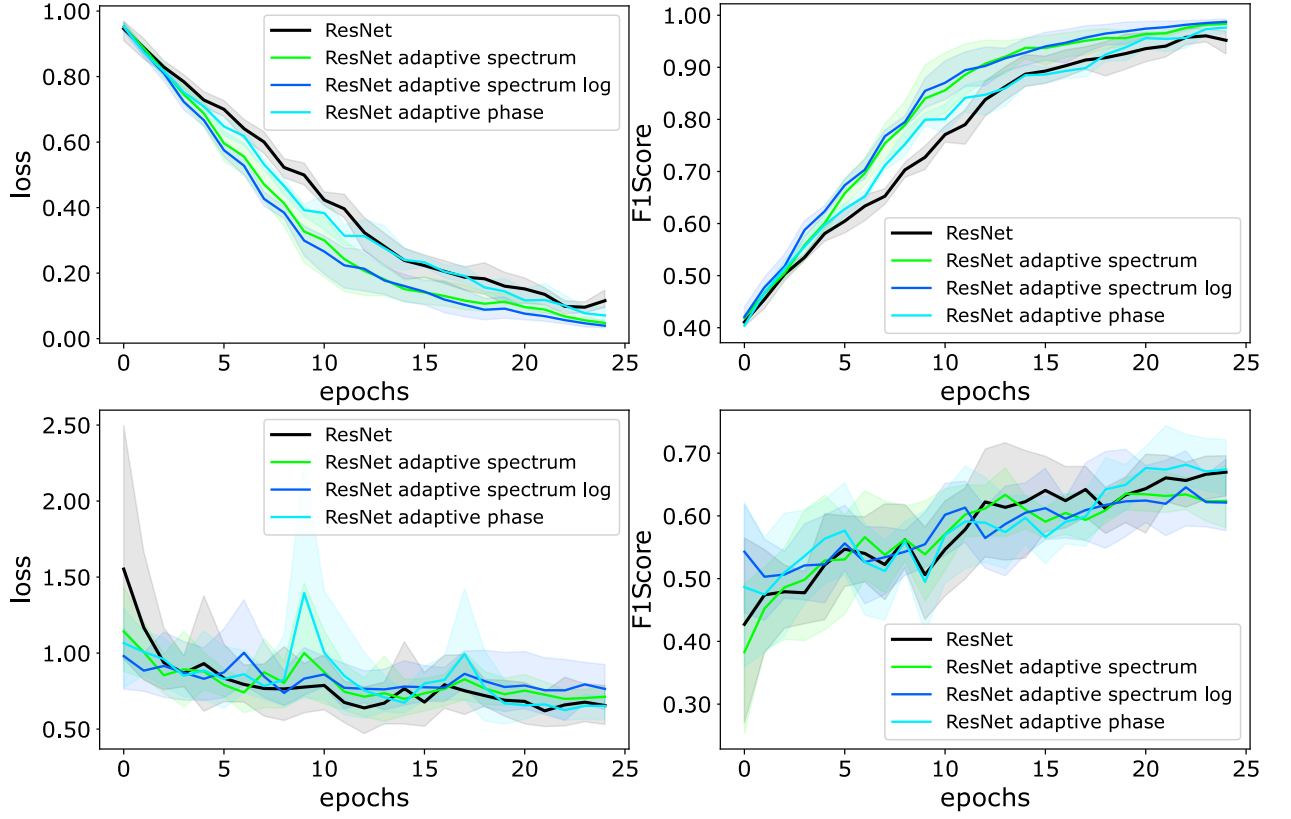


Figure C.1. Classification Results II. Top row: Loss function values and  $F_1$ -scores on BUSI train sets. Bottom row: the same on validation sets.

## Appendix D

# Denoising Results

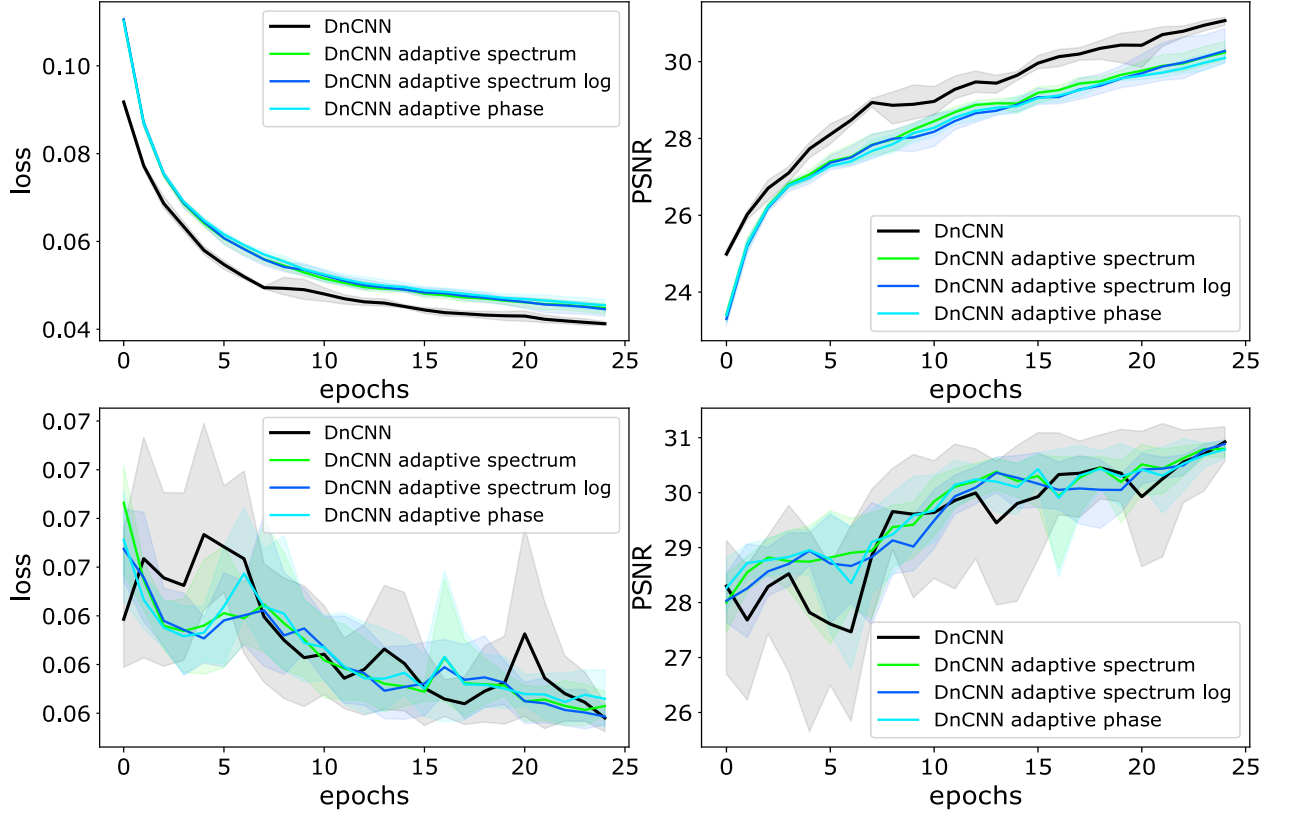


Figure D.1. Denoising Results. Top row: Loss function values and PSNR metrics on BUSI train sets. Bottom row: the same on validation sets.

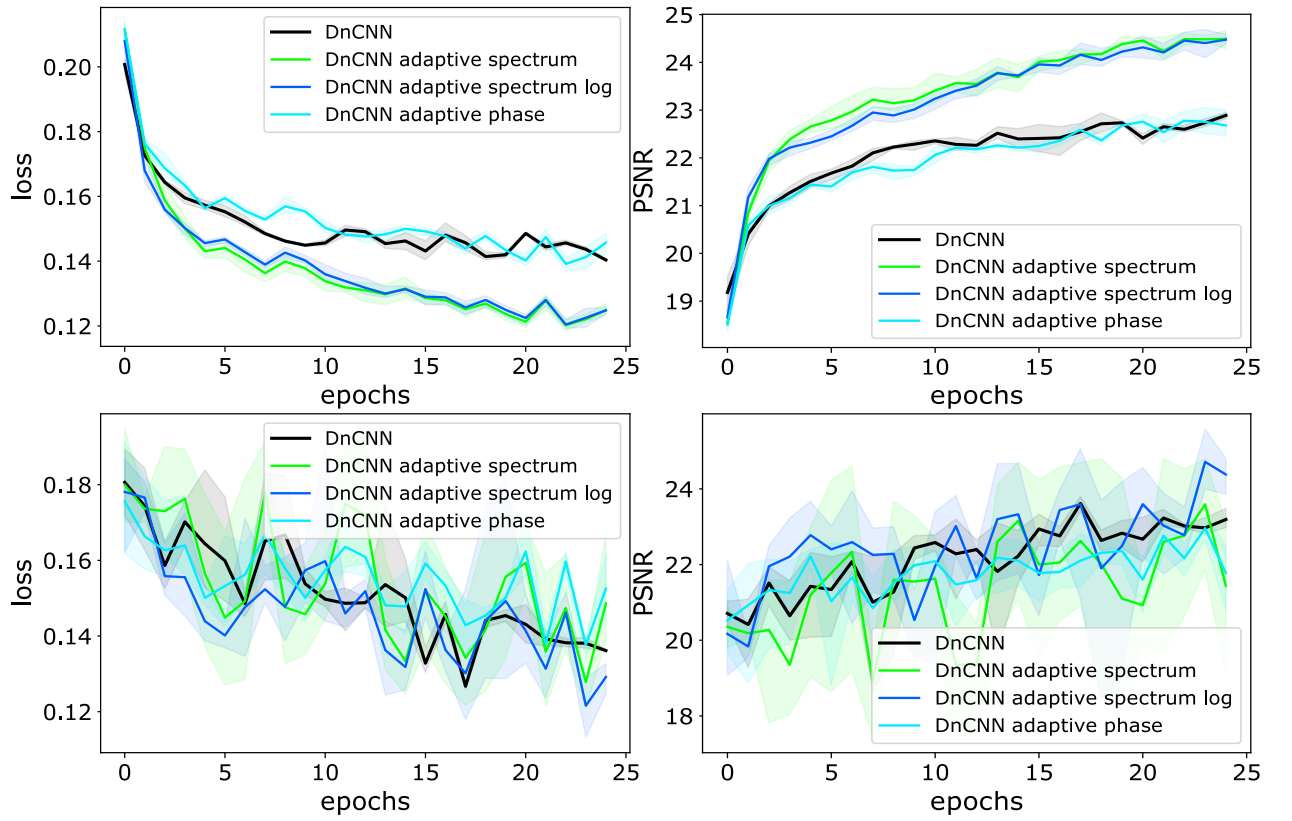


Figure D.2. Erasing Results. Top row: Loss function values and PSNR metrics on BUSI train sets. Bottom row: the same on validation sets.