FREQUENCY DOMAIN ANALYSIS OF U-NET SEGMENTED ULTRASOUND IMAGES

A Thesis

Submitted to the Faculty of Graduate Studies and Research

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science

In

Computer Science

University of Regina

By

Henrik Steenbock

Regina, Saskatchewan

August 2023

UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Henrik Reimer Steenbock, candidate for the degree of **Master of Science in Computer Science,** has presented a thesis titled, ***Frequency domain analysis of U-Net segmented ultrasound images,*** in an oral examination held on **August 3, 2023**. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:	Dr Aram Teymurazyan, Department of Physics
Supervisor(s):	Dr Sandra Zilles, Department of Computer Science
Committee Member:	Dr Xue Dong Yang, Department of Computer Science
Chair of Defense:	Dr Jaqui Gagnon, Faculty of Business Administration

Abstract

During prostate cancer brachytherapy, catheters are inserted into a patient's prostate for a highly localized radiation treatment. Accurately placed catheters are critical for successful treatment and ultrasound images are taken throughout the procedure to verify their exact positions. However, manually locating catheters on ultrasound images is extremely difficult, time consuming, and happens while the catheters are still in the patient. A fully automatic solution could significantly reduce procedure time and potentially even improve the precision. This thesis introduces a novel approach that segments 2D ultrasound images using the successful U-Net architecture to determine catheter candidates. These candidates are then extracted and Fourier Transformed into the frequency domain. De-convolution is performed directly in the frequency domain to reconstruct a number of frequency coefficients and remove noise. Additional features are calculated from the frequency coefficients to supplement the determined U-Net confidence and candidate location. Altogether, the features from each catheter candidate are classified by AdaBoost.

Acknowledgements

I am very thankful to my supervisor Dr. Sandra Zilles for her continued support and perspective not only during my research, but throughout my entire graduate studies. I have grown and learned many new things during this time. I am also very grateful for the generous funding she provided for my research.

I would also like to thank the Saskatchewan Cancer Agency for providing the medical ultrasound images and partial funding of my research. I would like to especially thank Dr. Derek Liu, who prepared the dataset in an easily usable format, provided a detailed background about brachytherapy, and answered all medical and technical questions about the data.

Further, I want to thank Shayantonee Tupor and Gopal Bala. Both of them worked with different approaches on related research before me and shared many valuable insights and data with me.

Lastly, am very grateful to my family and all their encouragement, support, and patience during my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Machine Learning and spectrum analysis in the frequency domain are fundamental tools in the domains of image and signal analysis. Both have been successfully utilized separately and together in an attempt to solve most common image analysis tasks [2, 9, 22, 40, 43, 47]. However, ultrasound images pose a big challenge to common image analysis techniques due to the amount of noise and a number of inconsistencies that are not easily removed [27], while especially the medical domain has a very strong need for accurate predictions [36]. This thesis proposes an approach combining Machine Learning and spectrum analysis to analyze and annotate medical brachytherapy ultrasound images.

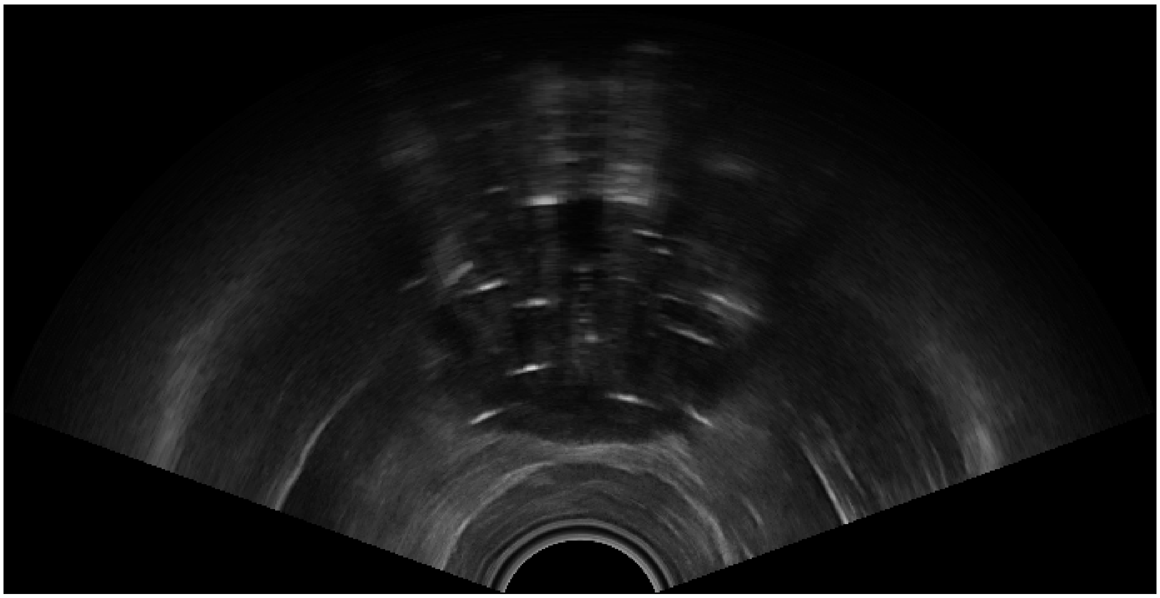


Figure 1.1: Sample ultrasound image of a prostate during brachytherapy [1].

1.1 Problem Statement and Motivation

Prostate cancer is among the most frequently occurring types of cancer in men [43, 47]. One of the most effective treatment options is brachytherapy, where needles (catheters) containing a high dose of radioactive material are inserted directly into the prostate for a highly localized treatment [23, 43, 47]. In an effort to not damage any healthy tissue and maximize precision, the exact position of these catheters is closely monitored by a specialist using ultrasound images taken during the procedure. However, this manual ultrasound image analysis is very time intensive, prone to human errors, and has to happen while the catheters are inserted into the patient, which means time is crucial to the patient's health [23, 43, 47].

The natural properties of ultrasound images are the reason why this task is so difficult to complete manually. As shown in Figure 1.1, the ultrasound images contain a substantial amount of speckle noise [27]. While some catheters are noticeable as bright regions of interest (ROI), there is no clear outline to exactly recognize and distinguish the bright regions. Further, the number of catheters causes shadows and artifacts that can appear like catheters, meaning the task at hand is not only a detection task, but also a classification task to distinguish catheters from non-catheters [23, 43, 47]. Partly due to both previously mentioned issues, the ultrasound images also suffer from brightness variance and distortion.

The main objective of this thesis is overcoming these issues by developing a Machine Learning model to automatically detect and classify catheters in 2D sequential brachytherapy ultrasound images in a much shorter time span. Layered on top of another, these annotated images can then be used to reconstruct a 3D model of the prostate, where true catheter locations and depths can be extrapolated using the 3D data.

1.2 Approach

Since the underlying task is twofold, the detection of regions of interest that could be catheters and the subsequent classification between catheters and non-catheters were done separately. An optimized U-Net [36] model as proposed by Tupor [40] was trained on ultrasound images to perform pixel-wise classification of ultrasound

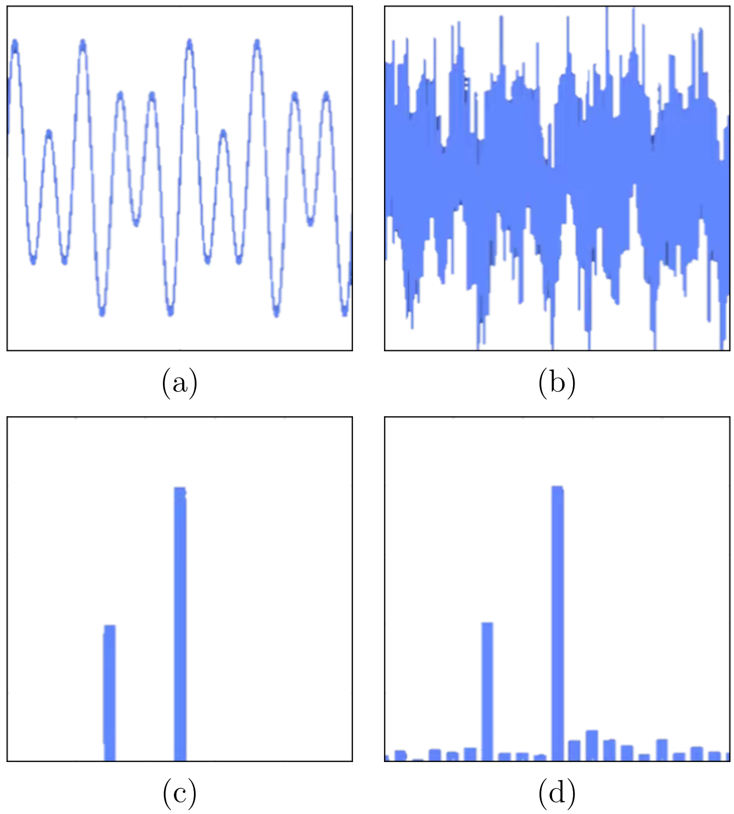


Figure 1.2: The same sum of two sinusoid functions (a) without noise and (b) with noise are shown as functions of time. (c) and (d) show the respective constituent sinusoid function magnitudes of (a) and (b) in the frequency domain as function of frequency. Adopted from [6].

images. Every connected group of catheter pixels was considered a region of interest. While U-Net already performs classification, the experiments from Bala [2] and this thesis indicate, that an alternative classification method based on ROI frequency domain features is promising with higher prediction confidence. Bala's results were promising because when a signal or image is decomposed into its constituent sinusoid functions in the frequency domain, the relevant components are much easier to recognize and extract from the unwanted noise as shown in Figure 1.2. Consequently, each detected ROI was extracted using a kernel and Fourier Transformed into the frequency domain. Convolution with the extraction kernel was then removed in the frequency domain based on the image spectrum reconstruction process introduced by Liu [22]. Multiple meaningful features were generated from the ROIs' reconstructed frequency coefficients and supplemented with additional features. This data was then classified by a simple AdaBoost predictor which appears to significantly outperform more complex classification methods in this instance, such as Neural Networks.

1.3 Contributions

This thesis proposes a novel twofold model for 2D brachytherapy ultrasound image segmentation and classification. The first model component uses the existing state-of-the-art U-Net architecture [36] as implemented and optimized by Tupor [40]. Unlike Tupor's approach, the U-Net component is not intended to only deliver ideal and accurate pixel-wise predictions, but it was meant to suggest as many catheter candidates as possible, regardless of them being true catheters. This goal was achieved by running two U-Net predictors each trained on differently pre-processed data on all data regardless of the pre-processing method.

During Bala's [2] experiments, their proposed model had the downside of being unable to perform ROI detection, limiting it exclusively to classifying manually annotated data with ROIs in the exactly correct locations. With a much larger

ROI pool than Tupor's, the set of automatically U-Net extracted ROIs was used for Bala's frequency domain analysis method as a second model component to evaluate their model's performance under more realistic conditions with fully automatic data. Throughout the frequency domain analysis process, a number of inaccuracies in Bala's research were found, thoroughly discussed, and analyzed in new experiments: The detected ROIs were extracted from the original ultrasound images using two different extraction kernels. The subsequent ROI frequency reconstruction in the frequency domain as proposed by Liu [22], was then modified to be performed with different combinations of the extraction kernels to determine their impact on the predictions.

Additional supplemental features, such as U-Net confidence of an ROI being a catheter and location information in the original ultrasound image were added to the data generated from the ROIs' frequency coefficients. The final numerical dataset was then classified by an AdaBoost ensemble learner for final predictions with an accuracy far exceeding the accuracy Tupor's [40] U-Net model and Bala's [2] Probing FCNN can achieve on their own.

1.4 Thesis Overview

The remaining chapters of this thesis are structured as follows:

- Chapter 2 provides an overview of recent research using U-Net based models in medical imaging and spectrum analysis of images.
- Chapter 3 introduces a deeper review of U-Net as a Convolutional Neural Network, the Fourier Transform, and spectrum analysis in the frequency domain.
- Chapter 4 explains the steps taken to generate a dataset of ROI candidates using U-Net. It then details the features extracted per ROI for frequency domain analysis used for the experiments in this thesis.
- Chapter 5 outlines all performed experiments and analyzes their results.
- Chapter 6 concludes this thesis with a discussion about its limitations and also potential for further research and improvements in the future.

Chapter 2

Related Research

Automatic image segmentation is one of the core disciplines in modern Machine Learning research due to the large number of possible applications. Self-driving cars [25], security systems [35], and face recognition [21] are just a few fields where image segmentation is actively researched for purposes of object detection and classification.

With an ever growing world population, efficiency in medical procedures is more important than ever, leading to the introduction of Machine Learning methods to automate and optimize manual analysis tasks [37]. Especially in medical imaging, applied Machine Learning methods not only save time, but show highly promising results in some domains, such as ultrasound image segmentation, where manual analysis is tedious, time intensive, and error prone [3, 12, 13, 47]. This chapter reviews multiple related approaches to address these issues which are also further outlined in

Section 1.1. Various recent research based on the renowned U-Net architecture [36] is discussed, followed by an alternative approach using image spectrum analysis.

2.1 U-Net Ultrasound Image Segmentation

U-Net [36] is among the most popular and successful convolutional neural network architectures for the medical imaging domain, where even in the case of brachytherapy ultrasound images only, multiple different research groups applied it independently since 2020 [40, 43, 47]. The reason for that is U-Net's unique architecture extracting local semantic features from repeatedly downsizing the image and putting them in the global context of the subsequently up-sampled whole image where noise was reduced due to the up-sampling, allowing it to accurately detect local and global features. Not only are objects detected accurately, their precise location is also determined and the model does not need a large sample amount to be trained. All of these factors make U-Net an excellent choice for medical purposes where both detection and localization are often a must and samples are scarce [36].

The simplest implementation of the U-Net architecture in this domain was done by Tupor [40], who focused solely on data pre-processing and hyperparameter-tuning their U-Net based model. Ultrasound images were cropped to the relevant areas to avoid false positives and they were contrast enhanced to make regions of interest, which also happen to be bright spots, more distinguishable from their surroundings.

Lastly, the dropout rate was increased to 50% for improved regularization. All these improvements led to a 2D accuracy of 79%. When the sequential image slices were stacked to reconstruct an annotated 3D prostate model and 2D catheter predictions were clustered to detect 3D catheters, an overall 3D accuracy of 89% was achieved by Liu et al. [23] based on Tupor's 2D data.

Zhang et al. [47] approached the task similarly with emphasis on pre-processing their data. They applied total variation regularization to reduce noise in their dataset. Additionally, they utilized the attention gate approach proposed by Oktay et al. [31] where attention gates are used in the U-Net skip connections to reduce false positives by assigning a lower weight to less relevant background features. While Zhang et al. did not quantify the accuracy of their 2D predictions, the accuracy in the resulting, reconstructed, 3D data was 96%, however, their dataset was not as clinically diverse and challenging as the one used by Tupor [40] and Liu et al. [23].

Wang et al. [43] implemented U-Net almost opposite to the modifications used by Zhang et al. [47]. Instead of trying to improve the data from U-Net skip connections, skip connections were omitted entirely to suppress noisy data from the encoder. Further, since the models of all three research groups [23, 43, 47] struggled detecting the tips of catheters, Wang et al. employed a secondary model with modified Vgg16 architecture [38] specifically designed to enhance catheter tip detection. Overall, their 3D model achieved an accuracy of 95% for catheter shafts.

The approach in this thesis builds on the ideas of Tupor [40] to utilize a simple but optimized U-net implementation with some data pre-processing. Similar to Wang et al. [43], a second model based on the work from Bala [2] as explained in Section 2.2 is applied in addition to U-Net, but it helps classify all catheters as a whole in 2D images, not just catheter tips as was the case for Wang et al.

2.2 Image Spectrum Analysis

Spectrum analysis in the frequency domain is an important technique traditionally used for signal analysis [9, 26]. However, it also has many application in image analysis for various image manipulation, filtering, and analysis purposes [4, 14, 22, 45].

In most cases, images are analyzed because they are not made up entirely of a repeating patterns, meaning they are discrete and non-stationary signals. As such, regions of interest have to be extracted and analyzed locally to avoid interference from other image regions [4, 14]. While extraction methods like the Gabor transform [34] already exist, they introduce a different kind of convolution to the extracted image regions. Liu [22] proposed a novel approach to remove this convolution. Bala [2] fully utilized Liu's solution on brachytherapy ultrasound images. They discovered a number of patterns in the frequency spectrum of catheter regions of interest and introduced features to capture these properties. Bala then used a Neural Network to classify regions of interest as catheters and non-catheters using these features.

They achieved an accuracy of 74% on 2D images which is lower than the results of Tupor [40] after the introduction of contrast enhanced images. However, the vast majority of predictions made by Bala's model had far more desirable confidence values than Tupor's U-Net model, indicating it could potentially enhance a U-Net model. The one downside of Bala's proposal is their model only being able to classify, but not detect regions of interest.

Chapter 3

Background

This chapter reviews background materials relevant to the experiments conducted during this thesis. It starts by explaining why image segmentation is important and how the Convolutional Neural Network architecture U-Net achieves this goal. Secondly, it provides a detailed description on the frequency domain, how different functions can be Fourier Transformed for frequency domain analysis, and what valuable insights can be obtained from doing so.

3.1 Image Segmentation

Thanks to deep learning technology, certain types of images can be automatically analyzed and labelled based on some object in them. In the context of some domains, such as medical imaging [36], however, it is highly important to not only know what is in an image, but where exactly the region of interest is located. Image segmentation

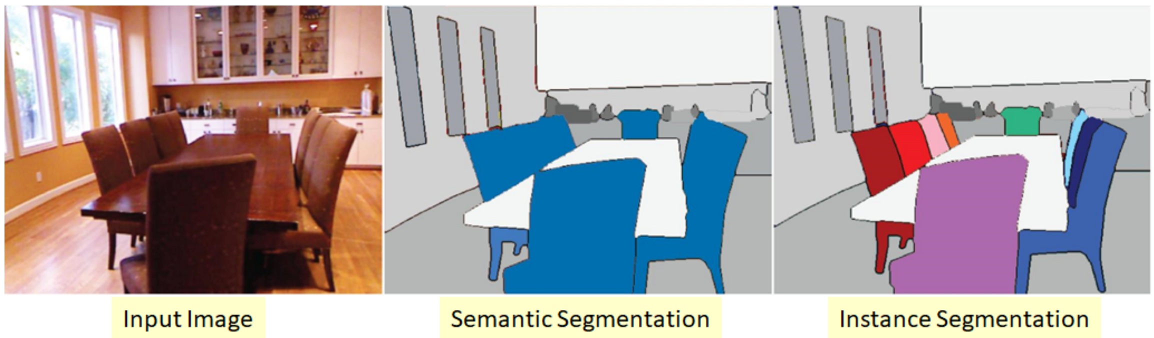


Figure 3.1: Visual difference between semantic image segmentation and instance segmentation of overlapping objects with the same class label. Adopted from [39].

achieves that goal by dividing an image into its various components based on some criteria [46].

Semantic segmentation is a kind of image segmentation that assigns a class label to each pixel of an image. When multiple instances of a class overlap without a clear border, pixel-wise classification does not infer the specific instance of that class to which the pixel belongs. The finer inference method instance segmentation is necessary to distinguish objects of the same class inside an image [16]. An example of the difference between semantic segmentation and instance segmentation is depicted in Figure 3.1.

3.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) [20] are a specialized group of neural networks that are modelled after the brain's reception of some types of inputs. Biological neurons in a brain each focus on a small, local subset of stimuli and collectively they can perceive the whole input. CNNs are widely used to apply that same technique in image classification and segmentation. As artificial neural networks, CNNs consist of multiple layers that perform some operations to extract information from the input. CNNs are centered around the linear convolution operation. Like neurons in a brain, the convolution operation extracts some useful information from a small region of an input to detect features that may not have been noticed in the global context of the entire input. Put together, these features result in a segmented version of the input. This makes CNNs ideal for data with grid-like topology, such as images [15].

3.2.1 CNN Operations

Convolution Operation

The most important operation CNNs are named after is the convolution operation. A kernel of pre-determined size containing real weight values is moved over an input to extract local features by using matrix multiplication. The results from moving the kernel in increments (called stride) across the entire input form a feature map quantifying the presence of that specific feature everywhere in the input [15]. An

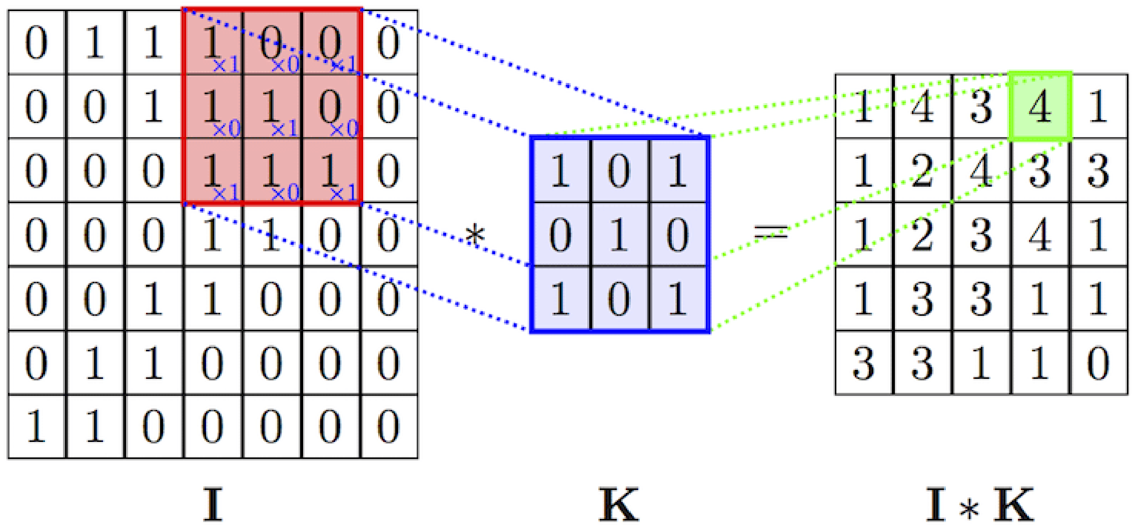


Figure 3.2: Visualization of a convolution operation on a 2D example input I using a 3×3 kernel K and stride 1. Adopted from [28].

example of this process can be observed in Figure 3.2.

Convolution operations often apply a number of kernels to an input to extract many feature maps at a time [15, 36]. The goal of training a CNN is optimizing the weights of these kernels to extract the most relevant feature information [15].

Activation Operation

After convolution operations, CNNs typically apply an activation function on the convolution output [15]. At their core, convolution operations are linear and cannot deal with non-linear inputs. Activation functions introduce non-linearity to make CNNs more robust to linear and non-linear inputs [44].

One of the most popular and to this thesis most relevant activation functions is rectified linear activation (ReLU). It maps negative values in feature maps to 0 and leaves non-negative values unchanged. Despite the simplicity of that change, it introduces enough non-linearity to feature maps to substantially improve a classifier's performance [44].

Another important activation function is the sigmoid function. Some CNN architectures use it in the output layer after a 1×1 convolution step to map class probabilities to that part of the output [36].

Pooling Operation

Pooling operations usually down-sample an input to a smaller size by extracting some summary of each region of an activated feature map [15, 36]. This step helps extract semantic context and it introduces invariance to small input translations [15]. For the purposes of this thesis, the most pertinent pooling operation is max pooling, which simply extracts the highest value in that region of activated features.

De-Convolution Operation

As implied by the name, de-convolution or transposed convolution can be seen as the counterpart to a convolution operation. Each value in a feature map is multiplied with some kernel of pre-determined size to up-sample and reconstruct an input [36, 40].

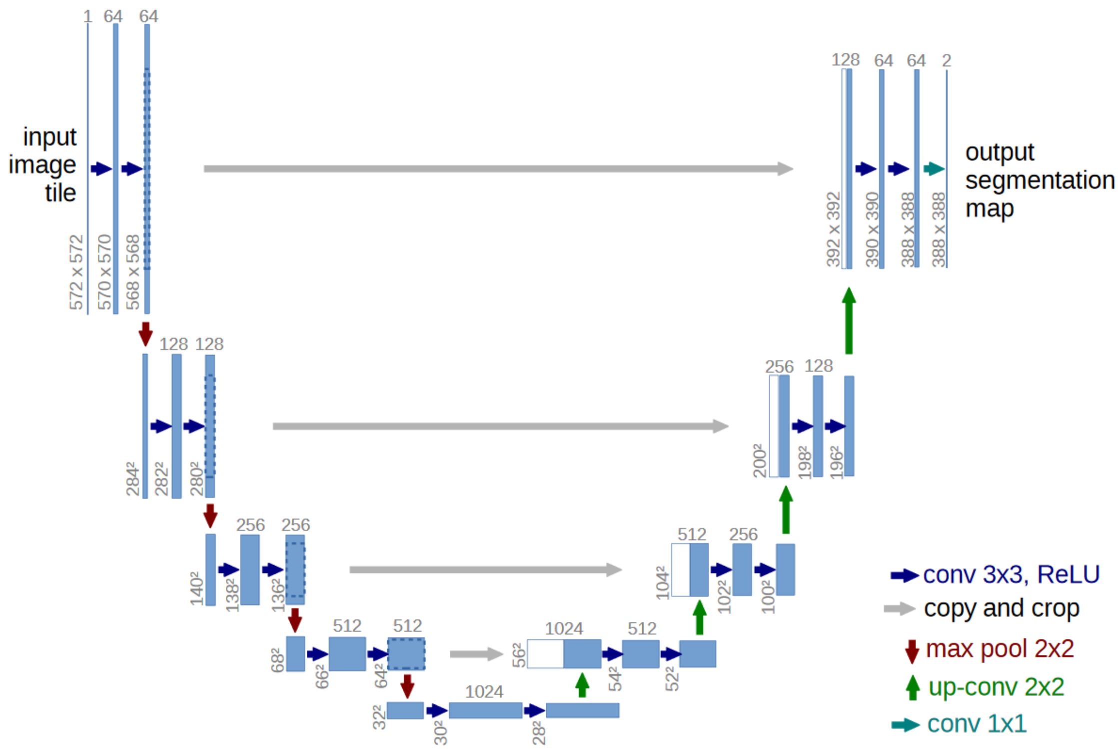


Figure 3.3: Schematic presentation of the U-Net architecture outlining the components, the data's path, and used data dimensions. Adopted from [36].

3.2.2 U-Net

Although first proposed 7 years ago, the CNN architecture U-Net designed by Ronneberger et al. [36] is still a state-of-the-art framework for CNN based image segmentation, with an ever growing number of derivative models [3, 43, 47]. Originally developed for the biomedical imaging domain, it meets the high demand not only for accurate classifications, but also precise localization of regions of interest. As a fully convolutional network [24], U-Net performs pixel-wise predictions, outperforming conventional CNN models in locating regions of interest.

U-Net was named after its U-shaped encoder-decoder architecture which can be seen in Figure 3.3. The encoder contains four steps that contract the data to lower resolutions, but with an increasing number of feature maps, to encode semantic information about low-level features in the input image. Each step performs two convolution operations with ReLU activation and stride 1, before the results are down-sampled by a max pooling operation with stride 2. In the initial encoder step, the first convolution operation generates the desired number of feature maps and the first convolution operation in subsequent steps doubles that number. The so-called bottleneck takes the encoded data and performs a pair of ReLU activated convolution operations as an encoder step would and passes it to the decoder without performing another pooling operation.

The decoder is more or less symmetric to the encoder and up-samples data to higher resolutions using information propagated by the feature maps and data received directly from the encoder through skip connections to decode high-level features and their precise locations. The four steps in the decoder start with a deconvolution operation to up-sample the image to twice its resolution with only half the number of feature maps. The up-sampled data is then concatenated with the appropriately cropped output of the second activated convolution operation in the corresponding step of the encoder. Each decoder also performs a pair of ReLU activated convolution operations of stride 1. The first convolution in each decoder step halves the number of feature maps. Lastly, a final convolution with kernel size 1×1 is used to map each pixel of the reconstructed image to a class label, resulting in a segmentation map of the input [36]. In the case of binary classes, a sigmoid activation function is used after the final convolution operation to receive class probabilities instead of class labels [40].

3.3 Frequency Domain

Images are traditionally represented by a 2D matrix where the matrix coordinates match the image coordinates and matrix entries represent the corresponding pixel intensity and color. While it is visually more intuitive, this image representation in the spatial domain is not always ideal for image processing. In addition to better efficiency, some operations and concepts are far easier to achieve in the frequency domain [4, 14].

A visual representation of the same one-dimensional signal in spatial and frequency domain are shown in Figure 3.4 [5]. The three wave functions in blue, green, and purple are components that result in the red function f if added together. In the spatial domain, f is a sinusoidal function of time. To get to the frequency domain, f is transformed to the complex-valued function F where most insights are usually gained from the magnitude. Consequently, frequency domain analysis looks at F as a function of frequency, showing the magnitude of the component signals as frequency

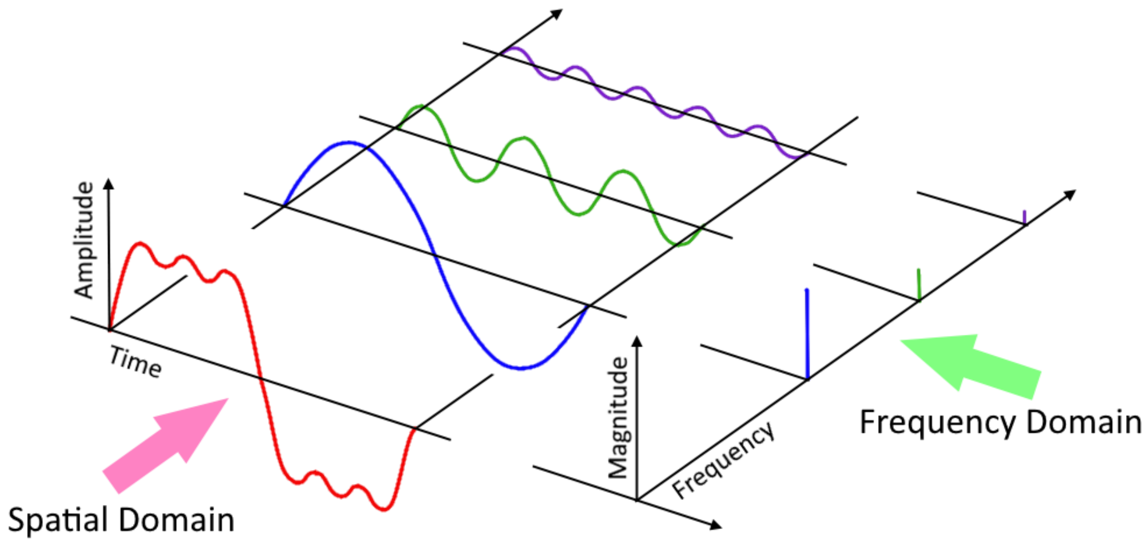


Figure 3.4: Representation of the different viewpoints from spatial/time domain and frequency domain on a one-dimensional function (red) composed of three other functions. Adopted from [5].

spectrum [7]. Nonetheless, the complex phase may also provide valuable insights [10].

3.3.1 Fourier Transform

The Fourier Series is a means of approximating any periodic function into a sum of sinusoid functions. Since trigonometric functions are very well understood and researched fundamental tools, their properties can be leveraged to gain invaluable insights that would otherwise be difficult to obtain. Most importantly, a function represented by a Fourier Series can be completely reconstructed to return from analysis in the frequency domain to the spatial domain without any information loss [4, 14]. Figure 3.5 shows examples of a periodic function encoded by varying numbers of

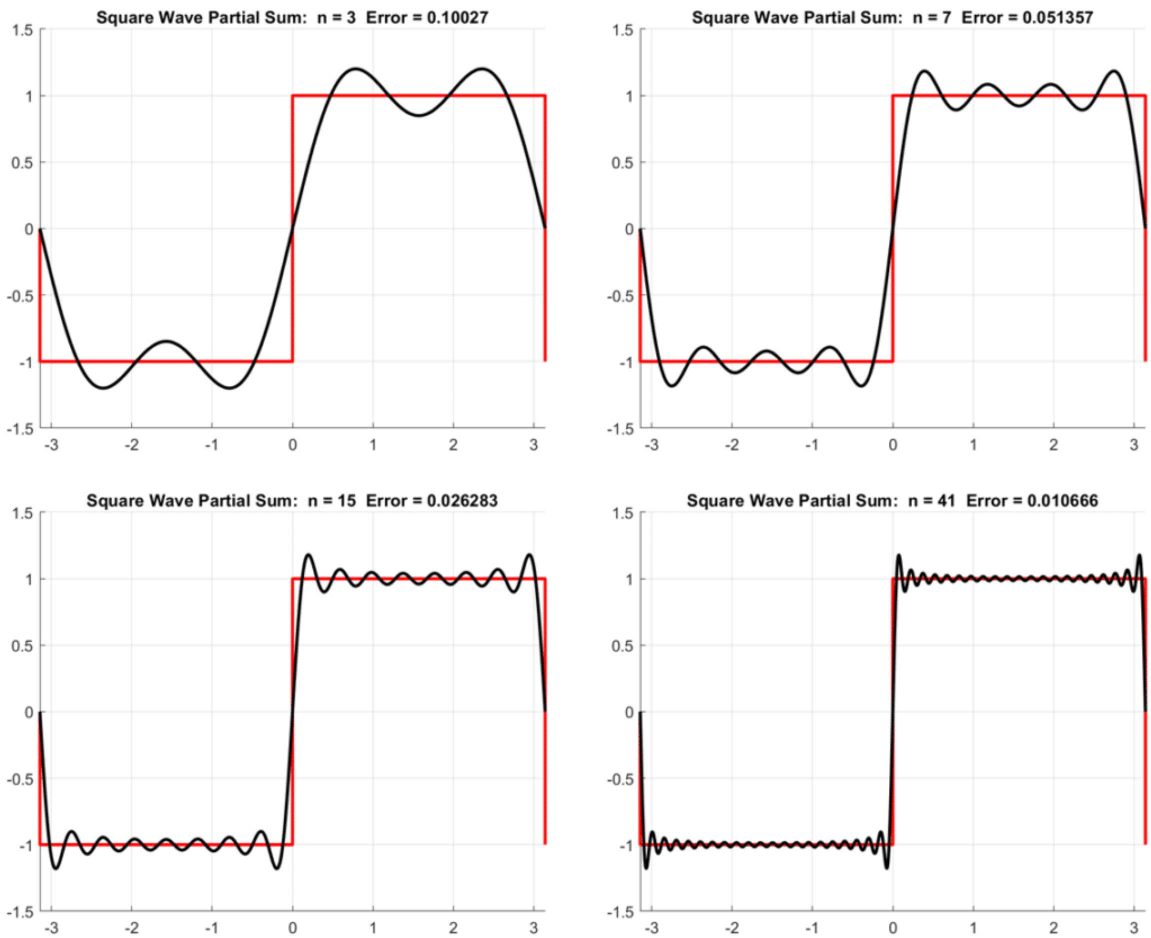


Figure 3.5: Example of Fourier Series approximating a periodic function (red) with the sum of n sinusoid functions. Adopted from [42].

sinusoid functions.

The Fourier Transform is a logical extension of the Fourier Series for non-periodic functions that are continuous and integrable with the goal of enabling frequency domain analysis of those functions. Similar to the Fourier Series, a function $f(x)$ is encoded as a sum of sinusoid functions. Unlike the Fourier Series, that sum is multiplied with a weighting function and then integrated, resulting in the transformed function $F(u)$ in the frequency domain, as shown in Equation 3.1 [14].

$$F(u) = \int_{-\infty}^{\infty} f(x) \cdot e^{-i2\pi ux} \, dx \tag{3.1}$$

The inverse Fourier Transform takes a signal function $F(u)$ in the frequency domain and restores the original signal function $f(x)$ in the spatial domain as shown in Equation 3.2 [4, 14].

$$f(x) = \int_{-\infty}^{\infty} F(u) \cdot e^{i2\pi ux} \, du \tag{3.2}$$

For the purpose of this thesis, we assume $f(x)$ to be in the real number space and $F(u)$ to be in the complex number space. With that in mind, Euler's notation [4, 14] for a complex number c denoted in Equation 3.3 shows how the transformation $F(u)$ from $f(x)$ can be written in the characteristic sum of sinusoid wave functions in the frequency domain.

$$c = e^{i\theta} = cos(\theta) + i \cdot sin(\theta) \tag{3.3}$$

In the case of 2D signal functions $f(x, y)$ from the spatial domain that are continuous and integrable, the forward and inverse Fourier Transform Equations 3.1 and 3.2 can be extended as shown in Equations 3.4 and 3.5 [4, 14] where u and v are frequencies along the x-axis and y-axis respectively.

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \cdot e^{-i2\pi(ux+vy)}\, dx\ dy \tag{3.4}$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v) \cdot e^{i2\pi(ux+vy)}\, du\ dv \tag{3.5}$$

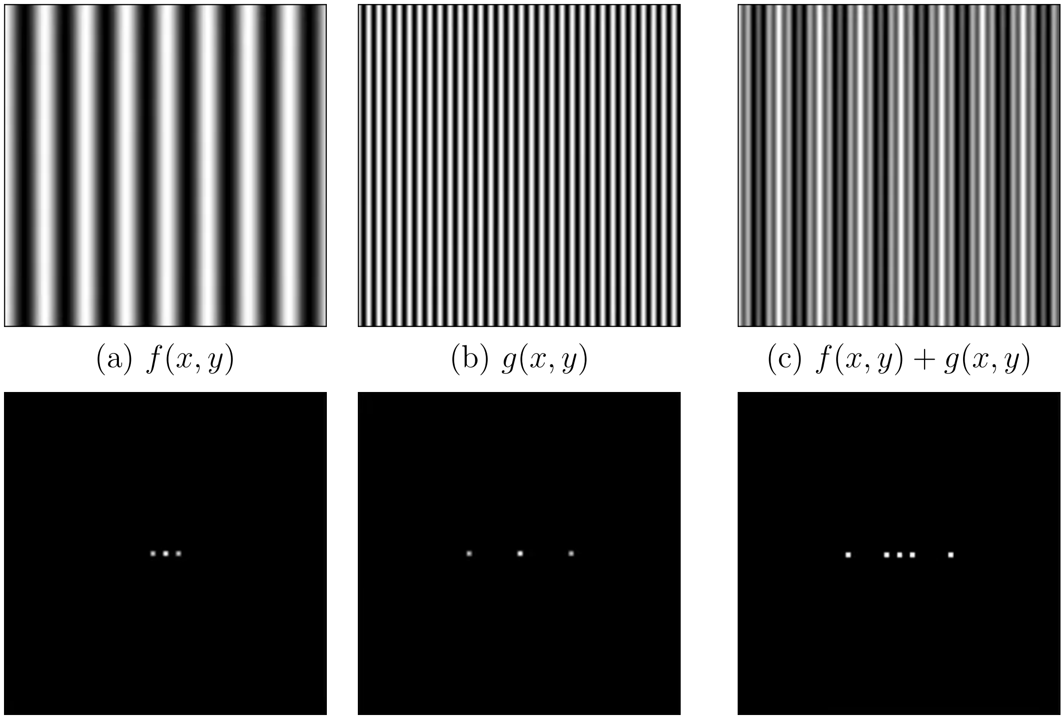
Since the signal $f(x, y)$ is a discrete image matrix with M rows and N columns in the case of this thesis, the Discrete Fourier Transform is necessary to handle discretely sampled input data. Equations 3.6 and 3.7 show, how frequencies are summed across columns and rows. The resulting Fourier Transform is another matrix of size $M \times N$ containing complex numbers [4, 14].

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \cdot e^{-i2\pi(ux/M+vy/N)} \tag{3.6}$$

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v) \cdot e^{i2\pi(ux/M+vy/N)} \tag{3.7}$$

3.3.2 Frequency Domain Analysis

Figure 3.6 visualizes what information the magnitudes of 2D images encode. Two simple *cos* functions f and g of different frequencies are shown in (a) and (b), whilst (c) shows the sum $f + g$. (d) through (f) show the respective absolute magnitude values of which there are two per sinusoid function plus an additional one represented by a dot in the image center. The latter magnitude represents the constant



(a) $f(x, y)$ (b) $g(x, y)$ (c) $f(x, y) + g(x, y)$

(d) $Log(|F(x, y)|)$ (e) $Log(|G(x, y)|)$ (f) $Log(|F(x, y) + G(x, y)|)$

Figure 3.6: (a) and (b) show simple sinusoid functions in a 2D image and (c) shows the combination of those two functions f and g. (d) through (f) show the respective magnitudes. For visualization purposes, magnitudes are shown as *log* of the absolute magnitude values. Adopted from [29].

value added to the sinusoid functions to avoid negative pixel values which cannot be displayed. (d) through (f) clearly indicate how easy it can be to combine, but also distinguish component functions in the frequency domain as opposed to the spatial domain. However, not all images are simple 2D representations of a few sinusoid functions. More realistic and complex images require far more sinusoid components to be represented, as shown in Figure 3.7. A notable observation is the many strong magnitudes in (d) along the diagonal from the bottom left to the top right which are

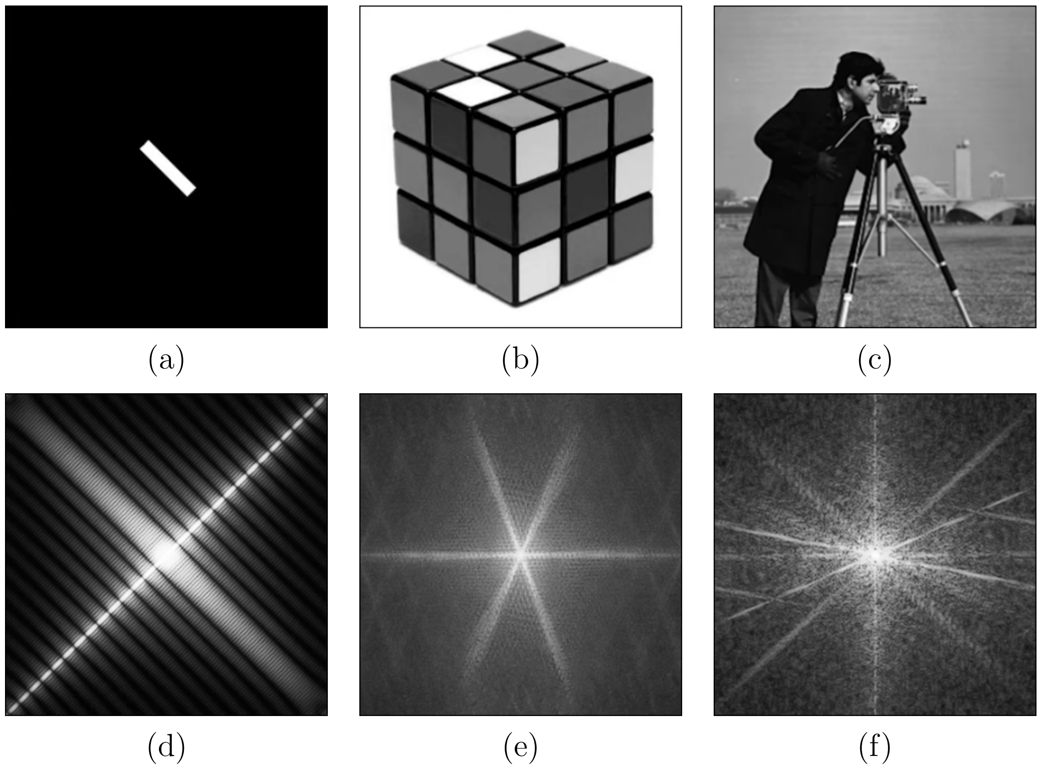


Figure 3.7: Examples of more complex 2D images (a) through (c) and the corresponding signal magnitudes in the frequency domain shown in (d) through (f). For visualization purposes, magnitudes are shown as *log* of the absolute magnitude values. Adopted from [29].

necessary to encode the sharp long edges in (a) of the rectangle which are orthogonal to that same diagonal line. Upon closer inspection, similar patterns can be seen in the other images and their respective magnitudes [29].

3.3.3 Signal Convolution and De-Convolution

As explained in Section 3.2.1, Convolution is a crucial mathematical operation in Neural Networks, but it is also a core tool in signal analysis. Let $f(x, y)$ and $g(x, y)$

be a continuous function and kernel with respective Fourier Transforms $F(u,v)$ and $G(u,v)$. The Convolution Theorem states, that the result $h(x,y)$ of the convolution $f(x,y) \otimes g(x,y)$ in the spatial domain can also be obtained from the multiplication $F(u,v) \cdot G(u,v) = H(u,v)$ in the frequency domain (Equation 3.8), where $H(u,v)$ is the Fourier Transform of $h(x,y)$, and vice versa, and where $\otimes$ denotes the convolution operation (Equation 3.9) [14].

Convolution Theorem

$$f(x,y) \otimes g(x,y) = h(x,y) \Longleftrightarrow F(u,v) \cdot G(u,v) = H(u,v) \tag{3.8}$$

$$f(x,y) \cdot g(x,y) = h(x,y) \Longleftrightarrow F(u,v) \otimes G(u,v) = H(u,v) \tag{3.9}$$

In the case of discrete functions, $F(u,v) \otimes G(u,v)$ in Equation 3.9 has to be divided by MN, but the following implications remain the same [14]. Practically applied, simple multiplication in one domain can be used to avoid computationally more complex convolution in the other domain utilizing the forward and inverse Fourier Transform. In theory, this also means convolution in one domain can be reversed through division in the other domain, but due to noise in observed functions this de-convolution process is not trivial [22, 33].

Most research focuses on de-convolution through division in the frequency domain of a convolution in the spatial domain [18, 19], but Liu [22] developed a novel approach to inverse convolution entirely in the frequency domain. Similar to later experiments in this thesis, Liu extracted a region of interest from an image $f(x, y)$ through multiplication with a Gaussian kernel [7] $g(x, y)$. The impact of that kernel could easily be removed through division $h(x, y)/g(x, y) = f(x, y)$ in the spatial domain, but then the image $f(x, y)$ would no longer be restricted to the desired region of interest. Instead, they had to perform de-convolution in the frequency domain to restore the image $F(u, v) = H(u, v) \oslash G(u, v)$, where $\oslash$ denotes the de-convolution operation as inverse to convolution for this thesis. Section 3.3.4 outlines Liu's approach to solving this de-convolution problem using an iterative reconstruction process.

3.3.4 Image Spectrum Reconstruction

Liu's Probing Detector [22] assumes a local region of a function $f(x)$ was extracted using multiplication with a Gaussian kernel $g(x)$ in the spatial domain resulting in image $h(x)$. Given are the Fourier Transforms $G(u)$ and $H(u)$ of $g(x)$ and $h(x)$ respectively. With the Convolution Theorem, we know $H(u)$ is the result of the convolution $F(u) \otimes G(u)$ and the goal is to iteratively reconstruct each frequency coefficient u composing the unknown Fourier Transform $F(u)$ of the extracted region $f(x)$ in the original image. Figure 3.8 visualizes this process with a one-dimensional example. (a) shows the frequency coefficient spectrum that is to be reconstructed, in (b) is

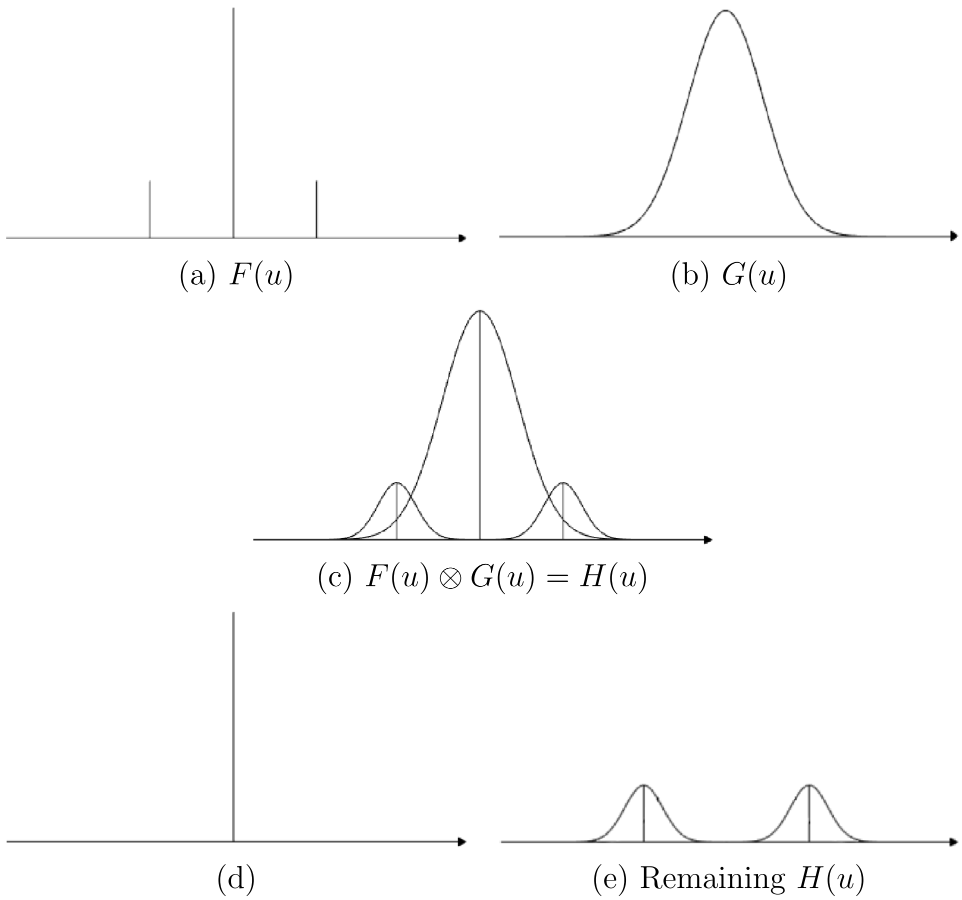


Figure 3.8: Visualization of the one-dimensional convolution between the Fourier Transform of a function in (a) and a Gaussian kernel in (b) resulting in (c). The frequency coefficient with largest magnitude in (d) is identified and its influence on (c) removed, resulting in (e). Adopted from [22].

the Fourier Transform of the Gaussian kernel, and (c) shows the convolution of both. Despite some peaks overlapping due to the convolution with the Gaussian kernel, the largest peak in (c) is assumed to be sufficiently distinguishable and a close estimate of some frequency u and is saved in the originally empty array $F(u)$ in (d) at that same location as approximation. $G(u)$ is then scaled by the magnitude of the extracted frequency coefficient and subtracted from $H(u)$, to remove that coefficient's peak and reveal peaks correlating to smaller frequency coefficient magnitudes as shown in (e). By introducing a maximum number of iterations N and extracting the N largest peaks only, noise frequency coefficients of smaller magnitude are filtered from the image.

The 2D version of the described algorithm as it was used in this thesis is formalized as follows:

Input: Image $f(x, y)$, probing location (x_0, y_0), window size σ.

Output: Reconstructed magnitudes of frequency coefficients $F(u, v)$ from the region of interest.

Step 1: Initialize a 2D array $F(u, v)$ and define the number N of maximum iterations.

Step 2: Compute a Gaussian kernel $g(x, y)$ given the probing location (x_0, y_0) and window size σ.

Step 3: Compute the product of $f(x, y)$ and $g(x, y)$, denoted as $h(x, y)$.

Step 4: Compute the Fourier Transforms $H(u, v)$ and $G(u, v)$ of $h(x, y)$ and $g(x, y)$ respectively.

Step 5: For each iteration:

(a) Find the largest magnitude peak in the frequency spectrum of $H(u, v)$ and store it in $F(u, v)$.

(b) Scale every element of $G(u, v)$ using the largest frequency coefficient magnitude and store the result in $G'(u, v)$.

(c) Subtract $G'(u, v)$ from $H(u, v)$.

Return: $F(u, v)$.

Chapter 4

Dataset Generation

This chapter gives a detailed explanation of the provided dataset and how informative features were extracted before evaluation during the experiments.

4.1 Ultrasound Image Dataset

Prostate ultrasound images of brachytherapy patients were kindly provided by the Saskatchewan Cancer Agency (SCA) [1]. Overall, their dataset consists of records from 303 different anonymous patients with each record containing 134 numbered image slices. Higher numbered slices are closer to the catheter insertion point. Most images have a size of 692×354 pixels. However, 13 records use different image sizes or have missing data, and thus were discarded.

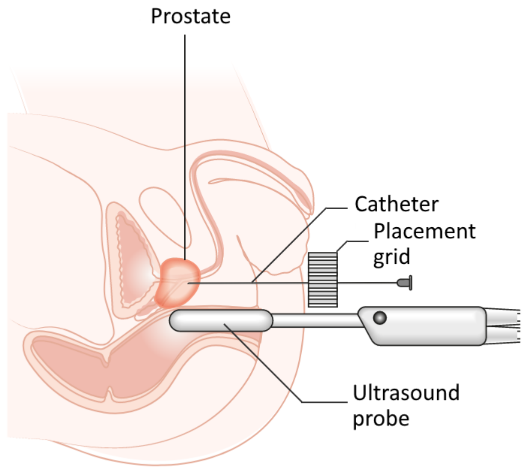


Figure 4.1: Visualization of the recording process for the ultrasound images provided by the Saskatchewan Cancer Agency [1] during prostate cancer brachytherapy. Adopted from [41].

4.1.1 Ultrasound Image Acquisition

As shown by Figure 4.1, catheters were inserted into a prostate using a placement grid. Afterwards, an ultrasound probe inside the rectum took 2D image slices of the prostate. Stacked on top of one another, the 2D ultrasound image slices comprise a 3D model of the prostate and the catheters inside it. While the same placement grid was used for all patients, the number and pattern of catheters vary between patients. Each slice contains between 11 and 18 true catheter labels for a mean number of 16 catheters.

Each image comes with four different masks describing the location of the prostate, catheters, urethra, and rectum. Of those masks, only the catheter labels are used in

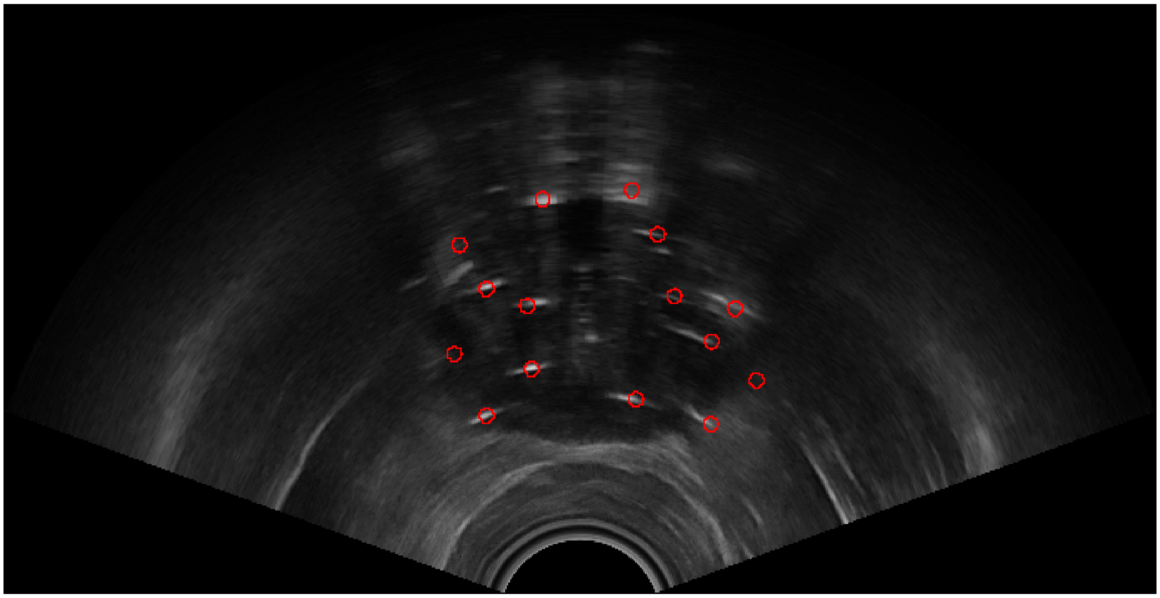


Figure 4.2: Sample ultrasound image from the SCA dataset with catheter ground truth highlighted in red [1].

the following experiments. These masks were created manually by medical professionals.

4.1.2 Dataset Challenges

Current brachytherapy ultrasound images pose a number of challenges that make it difficult to analyze them. Due do the nature of ultrasound images, they contain speckle noise that blurs regions of interest and distorts them the further away from the probe they are, as shown by Figure 4.2. Both of those issues combined lead to a variance in brightness and size of catheters. Not only does this make it harder to segment images, but it also causes catheters to not be visible at all in some of the 2D image slices. The catheters can also reflect ultrasound waves causing *echoes* that are often indistinguishable from a catheter in the ultrasound image to people without medical background.

Another issue is the physical shape of the catheters. They are long and thin needles that may bend inside the tissue, despite the placement grid being used during insertion. The further inside a body they are, the more their location may deviate from their original grid position.

Since the dataset was labelled by medical professionals for medical purposes, the labels are presumably as accurate as they can get. However, the data requirements for medical and machine learning processes are different. Even if a catheter is just barely or not at all visible in an ultrasound image, the catheter position was determined using the neighbouring 2D image slices to add a catheter label at that position anyway. While that is important for medical purposes, it leads to some ultrasound image sections labelled as catheter based on 3D information outside of that 2D image slice. Additionally, the brightness and contrast of entire ultrasound images, but also individual sections of them, were adjusted per image to ideally facilitate the manual segmentation process. This leads to further variance in contrast and brightness throughout the dataset and even within individual ultrasound images.

4.2 FDA Dataset

A number of steps have to be taken to generate a dataset suitable for Frequency Domain Analysis (FDA) of the provided ultrasound images. First, the ultra-sound images have to be pre-processed to have a U-Net classifier suggest candidate regions of interest that could be a catheter. The suggested image coordinates then have the steps outlined in Section 4.2.3 applied to them to generate meaningful features to classify that region of interest. Lastly, further supplemental features are added to the dataset for experiments to be performed on it.

4.2.1 ROI Candidate Generation with U-Net

Even seven years after it was first proposed, U-Net is still the state-of-the-art image segmentation architecture [36]. Tupor [40] has accordingly chosen a U-Net based model in their thesis to find catheters in ultrasound images from this very dataset and their model achieved an accuracy of 79%. For that reason, two different U-Net classifiers were trained as part of this thesis according to the optimized hyper-parameter specifications determined by Tupor. The first classifier was trained and evaluated on a cropped version of the original ultrasound images (*original classifier*), while the other one is using the same cropped ultrasound images after their contrast was enhanced through histogram equalization (*contrast classifier*) [30].

Tupor only had access to the first 185 records of the ultrasound image dataset and they trained as well as validated their model on 2D slices from within the range of 10 to 130 of those records. Other slices were discarded because they contained faulty images without proper ultrasound readings, or did not show the prostate. The two classifiers were also trained and evaluated on the image slices 10 to 130 from the records 1 to 185 as Tupor did. The core difference to Tupor's approach is that the generated list of potential catheter centroids from the pixel-wise U-Net prediction was not trimmed to the most likely X candidates, where X is the known number of catheters in that image. Instead, every single potential candidate was used to generate entries in the dataset for frequency domain analysis, assuming FDA is better at detecting catheters as indicated by Bala's findings [2].

Each predictor classified both original and contrast enhanced image slices 10 to 130 of the records 186 to 303, to avoid any bias the classifiers might have to the images on which they were trained. Pixel-wise classification used a confidence threshold of 20%. This process resulted in four lists of catheter centroid candidates from ROIs as those shown in Figure 4.3. Especially in the case of the *original classifier* predicting contrast enhanced images and the *contrast classifier* predicting unaltered images, the candidate lists contained false positive ROIs. These intentionally poor results add additional candidates which the more optimized pairings of classifiers, predicting the types of images they were trained on, did not find. While the odds of these extra candidates actually being catheters were presumed to be low, this step added a large

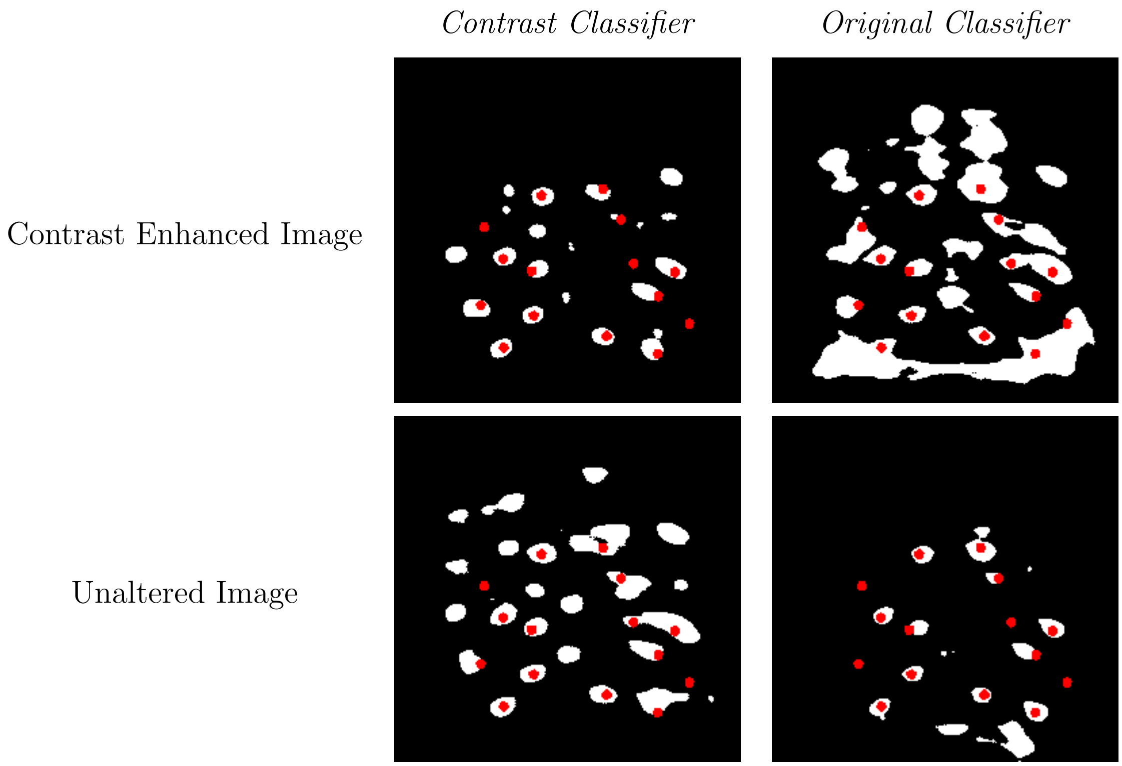


Figure 4.3: Sample predictions of the two U-Net models trained on contrast enhanced data and unaltered data respectively on both contrast enhanced and unaltered images. Ground truth labels are highlighted in red.

number of negative examples needed to train the subsequent FDA classifier. The four separate candidate lists were first filtered to remove small regions of interest with bounding box $pixelwidth + pixelheight < 12$. The threshold 12 was determined to remove the smallest of ROIs which in the vast majority of cases are non-catheters of which a sufficiently large number was already being detected. The few actual catheter ROIs of that small size are in most cases also detected by a different classifier and image type pairing. After filtering, the candidate lists were concatenated in the following order:

1. Candidates from *contrast classifier* predicting contrast enhanced images

2. Candidates from *original classifier* predicting unaltered images

3. Candidates from *contrast classifier* predicting unaltered images

4. Candidates from *original classifier* predicting contrast enhanced images

Each element of the resulting ordered list was then called in order, removing all subsequent list entries with a position within euclidean distance of 3 mm (11.33 pixels) to remove duplicates. Overall, the mean number of extracted unique candidates from each image slice was 36, more than twice the mean number of 16 true catheters. Similar to Tupor's method [40], each candidate was assigned two different class labels. The *coarse class label* states whether the centroid is within 3 mm (11.33 pixels) of a true catheter centroid and the *fine label* is true if the centroid is within 2 mm (7.5 pixels) of a true catheter centroid.

4.2.2 ROI Detection Accuracy

Since all candidates were solely determined by the explained U-Net predictors without being manually supplemented with all true catheters, some true catheters were not found by the classifiers and did not make it into the candidate list. The larger number of candidates per 2D image slice are introduced to reduce the risk of false negatives while increasing the number of false positives. However, the evaluation by a more accurate second model and a potential later 3D reconstruction address the false positives. 92.68% of catheters were detected according to the coarse class labels and 81.03% of them were detected using the finer class label. Even if the classification on FDA features had perfect accuracy, it would have to be weighted by these percentages to reflect the true percentage of correctly identified catheters as product of detection and classification accuracy.

4.2.3 Frequency Domain Feature Generation

Once coordinates of a region of interest candidate had been determined, a 33×33 pixel window was placed around the centroid of that area in the original image extracting an image excerpt using multiplication with a kernel. In Bala's [2] experiments this window was a Gaussian kernel $g(x, y)$ resulting in the extracted image region $h(x, y)$, but for reasons further outlined in Section 5.2, this step was also repeated with a rectangular kernel $d(x, y)$ resulting in image excerpt $h'(x, y)$. The contents of the ROI were then placed in the center of an otherwise black 256×256 pixel image. Afterwards, the brightness of each pixel in the ROI was normalized to the range $[0, 255]$ to address brightness variance, with 0 denoting a black, and 255 denoting a white pixel. The kernels and extracted and transformed example ROIs are shown in Figure 4.4. The resulting images $h(x, y)$ and $h'(x, y)$, the Gaussian kernel $g(x, y)$, and the rectangular kernel $d(x, y)$ were then transformed into the frequency domain

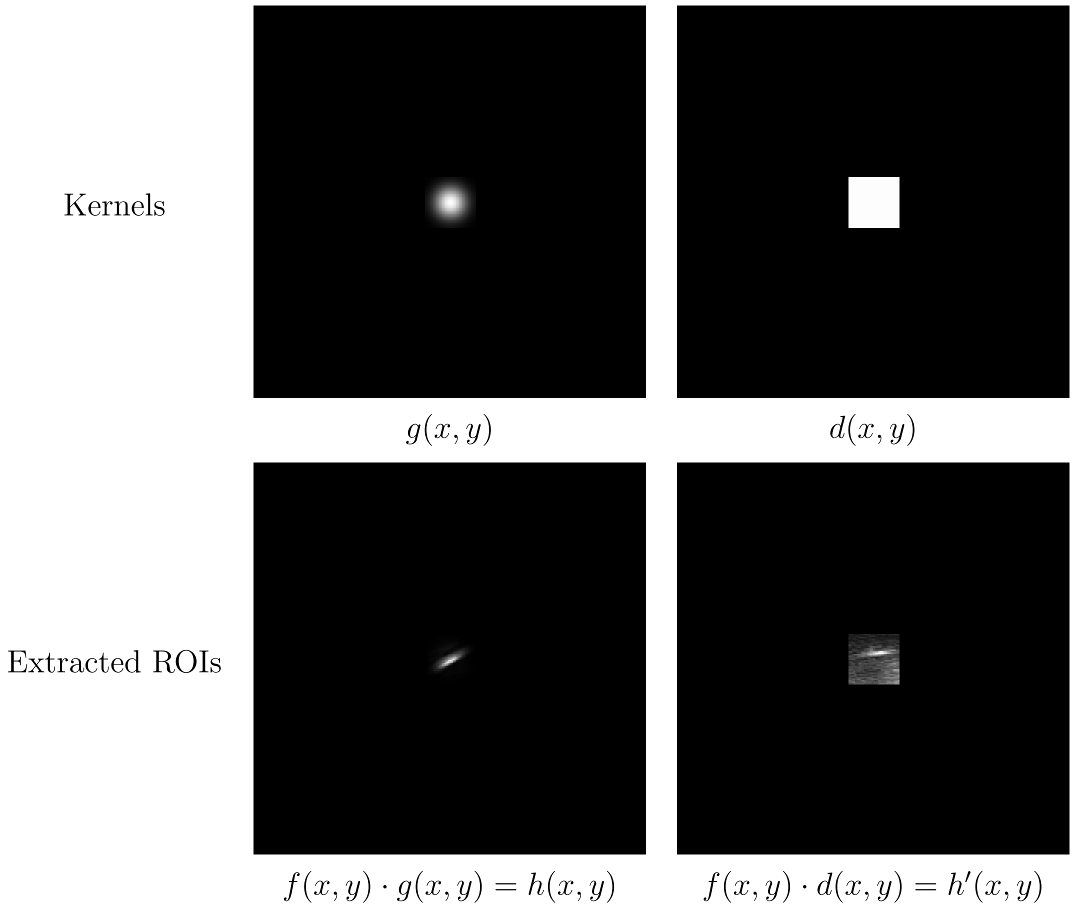


Figure 4.4: Gaussian kernel $g(x, y)$ and rectangular kernel $d(x, y)$ used to extract a region of interest $h(x, y)$ and $h'(x, y)$ respectively from an image $f(x, y)$. Each kernel extracted a different ROI.

using the Fast Fourier Transform. The transformed images are $H(u, v)$, $H'(u, v)$, $G(u, v)$, and $D(u, v)$ respectively. With the ROI in the frequency domain, 200 frequency coefficients were iteratively calculated to reconstruct the de-convoluted image frequency spectrums $F(u, v) = H(u, v) \oslash G(u, v)$, $F'(u, v) = H'(u, v) \oslash D(u, v)$, and $F''(u, v) = H(u, v) \oslash D(u, v)$. The following is a full overview of Liu's [2, 22] modified probing detector algorithm to reflect those modifications:

Input: Image $f(x, y)$, probing location (x_0, y_0), window size σ.

Output: Reconstructed magnitudes of frequency coefficients $F(u, v)$, $F'(u, v)$, and $F''(u, v)$ from the region of interest.

Step 1: Initialize 2D arrays $F(u, v)$, $F'(u, v)$, $F''(u, v)$, and $G'(u, v)$ and define the number N of maximum iterations.

Step 2: Compute a Gaussian kernel $g(x, y)$ and a rectangular kernel $d(x, y)$ given the probing location (x_0, y_0) and window size σ.

Step 3: Compute the products $f(x, y) \cdot g(x, y) = h(x, y)$ and $f(x, y) \cdot d(x, y) = h'(x, y)$.

Step 4: Normalize $h(x, y)$ and $h'(x, y)$ independently to the range $[0, 255]$.

Step 5: Move the regions extracted with kernels to the center of the respective image.

Step 6: Compute the Fourier Transforms $H(u, v)$, $H'(u, v)$, $G(u, v)$, and $D(u, v)$ of $h(x, y)$, $h'(x, y)$, $g(x, y)$, and $d(x, y)$ respectively.

Step 7: For each iteration:

- **(a)** Find the largest magnitude peak in the frequency spectrum of $H(u, v)$ and store it in $F(u, v)$.
- **(b)** Copy $G(u, v)$ to $G'(u, v)$ and scale every element of $G'(u, v)$ using the largest frequency coefficient magnitude from (a).
- **(c)** Subtract $G'(u, v)$ from $H(u, v)$.

Step 8: For each iteration:

(a) Find the largest magnitude peak in the frequency spectrum of $H'(u, v)$ and store it in $F'(u, v)$.

(b) Copy $D(u, v)$ to $G'(u, v)$ and scale every element of $G'(u, v)$ using the largest frequency coefficient magnitude from (a).

(c) Subtract $G'(u, v)$ from $H'(u, v)$.

Step 9: Restore $H(u, v)$ as the Fourier Transform of $h(x, y)$.

Step 10: For each iteration:

(a) Find the largest magnitude peak in the frequency spectrum of $H'(u, v)$ and store it in $F''(u, v)$.

(b) Copy $G(u, v)$ to $G'(u, v)$ and scale every element of $G'(u, v)$ using the largest frequency coefficient magnitude from (a).

(c) Subtract $G'(u, v)$ from $H'(u, v)$.

Return: $F(u, v)$ and $F'(u, v)$.

Despite calculating frequency coefficients for $N = 200$ iterations, Bala only used the frequency coefficients from the first $N = 40$ iterations [2]. However, the results shown in Section 5.1 indicate a higher accuracy using a larger number of coefficients. Consequently, $N = 200$ frequency coefficients were extracted for each region of interest. Coefficients and subsequently generated data from $F(u, v)$, $F'(u, v)$, and $F''(u, v)$ were kept strictly separate, resulting in three different datasets.

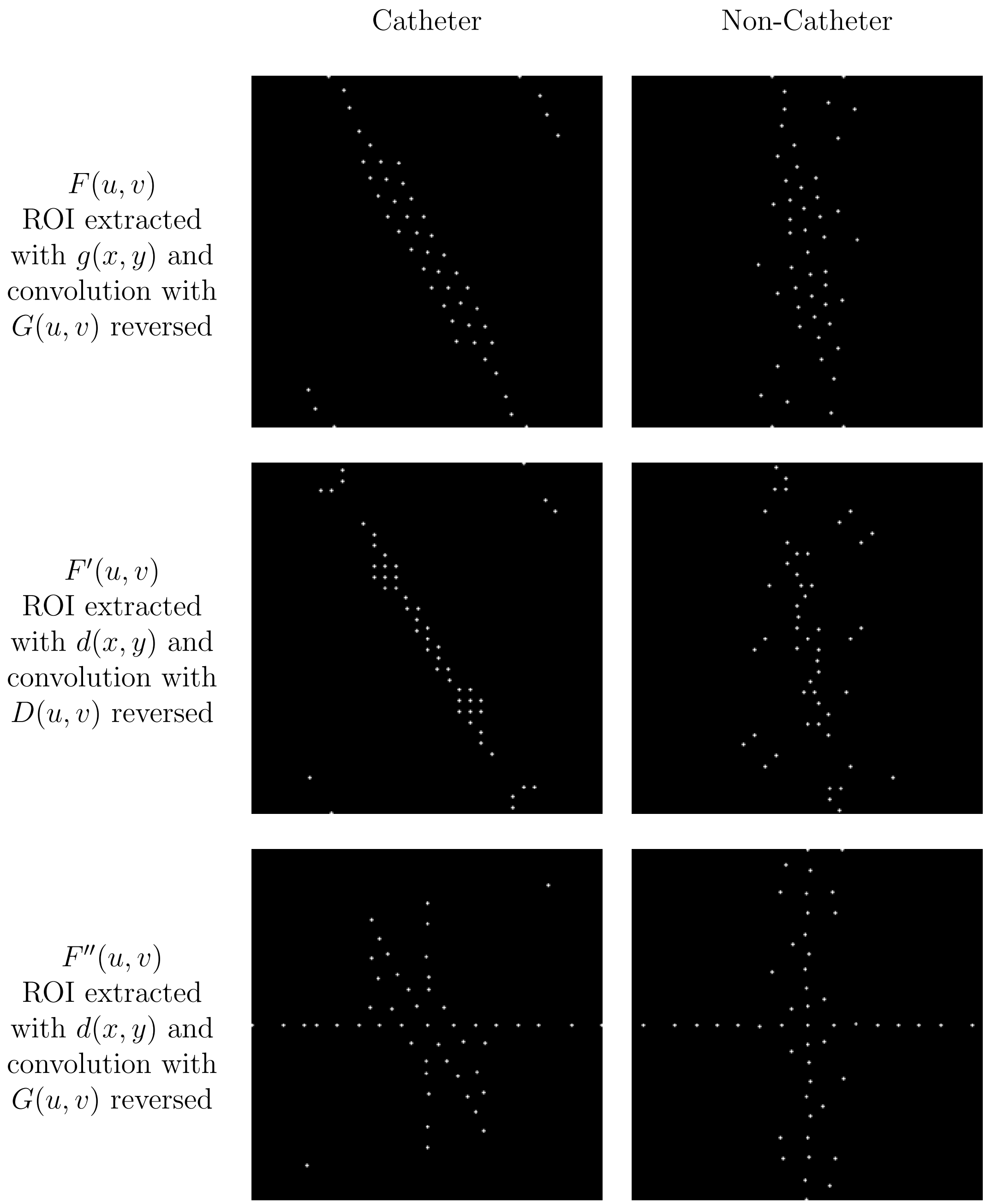


Figure 4.5: Plotted locations of the first 50 frequency coefficients of the same catheter and non-catheter in each of the three datasets. For visualization purposes, all coefficients are shown as white pixels regardless of the corresponding magnitude values.

Figure 4.5 visualizes the differences in frequency spectrums of the three different datasets. A noticeable outlier are both images from $F''(u, v)$ with the many frequency coefficients along or near the x-axis and y-axis. Based on the findings in Section 3.3.2, this observation is expected, as the corresponding sinusoid waves along those axes represent the sharp edges of the rectangular kernel $d(x, y)$. Those frequencies are not being removed, as de-convolution in that case was performed with $G(u, v)$, which does not have those sharp kernel edges. With these near-axes coefficients out of the way, the images in each column are looking much more alike. In the case of both catheters and non-catheters, most frequency coefficients are located on or around a line going through the image center, rotated to be orthogonal to the long edge of bright catheter candidates (Figure 4.4), which is also expected in accordance with the findings of Section 3.3.2. The core difference between catheter and non-catheter frequency spectrums appears to be the more structured coefficient scattering for catheters, while non-catheter frequency coefficients are scattered more chaotically. The exact scattering patterns vary from ROI to ROI, but they generally follow the observations made about the shown examples.

To further analyze the calculated frequency coefficients, the four additional features distance to center, spectral rotation, spectral deviation, and sector index were calculated for FDA classification in addition to the magnitude value, as proposed by Bala [2]. The intent is highlighting and leveraging the differences between catheters and non-catheters in their scatter behaviour of their frequency coefficients.

Magnitude

Magnitude is simply the value of a frequency coefficient. Due to the way frequency coefficients are extracted, the magnitudes of coefficients are an ordered set with the magnitude from the first iteration being largest. Magnitude values are positive real numbers.

Distance to Center

This feature denotes the euclidean distance of a frequency coefficient to the center of the reconstructed image. In Figure 4.5 it is apparent, that all coefficients are in some way scattered around the center, and the first coefficient is also always exactly in the center.

Spectral Rotation

Spectral rotation describes the counter-clockwise rotation of a frequency coefficient from the positive x-axis. Bala measures this feature as angle θ [2], but for this thesis, the tuple $(sin(\theta), cos(\theta))$ was added as alternative feature to describe an angle. In theory, that tuple represents the cyclic properties of an angle better than degrees because angles like $\theta_1 = 1°$ and $\theta_2 = 359°$ would be numerically closer in the tuple than they are as angles in degrees.

Spectral Deviation

If the coefficients C_1, C_2, and C_3 are from three successive iterations, the spectral deviation angle θ_1 for C_2 is defined as $180° - \theta_2$. θ_2 is the clockwise rotation angle at C_2 from line C_1C_2 to line C_2C_3. As with spectral rotation, this feature was added as angle θ and also as tuple $(sin(\theta), cos(\theta))$.

Sector Index

The sector index is an aggregated version of spectral rotation. A circle was arbitrarily divided in 16 evenly sized sectors and frequency coefficients are assigned one of these indices based on their spectral rotation.

4.2.4 Final Dataset

The final comprehensive tabular dataset consists of an entry for each region of interest. With 3 non-rotation based FDA features and both rotations represented by 3 features each, ROIs are described with a total of 9 FDA features per iteration. This results in $200 \times 9 = 1800$ frequency domain features, the fine and coarse class labels for whether or not it is a catheter, and the following 6 supplemental features that Bala did not use [2].

Slice Index

The index of the 2D image slice is included because slices closer to the insertion point more closely resemble the pattern on the placement grid. Machine learning methods could use it to weigh the importance of knowledge about commonly used catheter patterns.

Position in Original Image

Similar to the slice index, the position of the region of interest in the original ultrasound image is important to learn frequently used catheter patterns on the placement grid. Additionally, it provides information about the amount of distortion and speckle noise at that location in the image because both get worse further away from the ultrasound probe. Due to the radial property of the given ultrasound images, the centroid location is encoded in polar coordinates as (d, ϕ). Here, d denotes the distance to the center of the probe and ϕ denotes the centroid's angular offset from the positive x-axis, where the origin is the probe location. For reference, euclidean centroid coordinates are also included in the dataset for a total of four location features.

U-Net Confidence

Since the U-Net predictor already performs a full prediction based on Tupor's [40] model, U-Net confidence values of a ROI being a catheter can provide useful insights.

Chapter 5

Experimental Results and Analysis

This chapter describes all experiments performed throughout this thesis research and analyzes their respective results.

Since the datasets detailed in the previous chapter contain a large number of data entries, the experiments only used a randomly sampled subset of slices and all their corresponding regions of interest across all patients to represent the performance of the model in a more realistic setting. Entire slices were used to have rotation invariance to the line along which frequency coefficients are scattered as explained in Figure 4.5. Experiments based on the datasets generated for this thesis are run on 2,000 random slices across all patients which is around 76,000 regions of interest.

All experiments were validated using 5-fold cross validation. Unless otherwise specified, all listed accuracy values show the mean accuracy of this cross validation. Datasets were always split into training and testing data using stratified sampling.

Dataset splits were based on a shared random seed within the scope of each experiment to avoid variance in results caused by (dis-)advantageous dataset splits for the compared predictors by assuring identical dataset splits for each predictor.

The 5x2 paired t-test method proposed by Dietterich [8] was used for all significance tests. Significance tests were always performed on the more refined classifiers trained on fine class labels. The used α level of significance was $\alpha = 0.05$.

5.1 Probing FCNN Comparison

Bala's [2] experiments have shown, that their frequency domain analysis (FDA) features provide meaningful information that facilitate predictions with a confidence far higher than that of the U-Net based model by Tupor [40]. Inspired by this success, their experiments were repeated within the scope of this thesis with the exact same dataset, but using simpler off-the-shelf classifiers from the scikit-learn library[1] to compare their performance to Bala's neural network [32].

First of all, a Decision Tree as simplest possible predictor and AdaBoost [11] as ensemble learner were tested on Bala's [2] dataset of FDA features from various numbers of iterations N to evaluate their Probing FCNN under the same circumstances. This dataset used by Bala themselves is comprised of 38,310 ROI entries. Both classifiers used default scikit-learn [32] hyper-parameters. From each iteration, this experiment

[1] https://scikit-learn.org/stable/index.html

used all 5 features because preliminary experiments with smaller feature maps as applied by Bala yielded poorer results. The results shown in Table 5.1 indicate that a simple Decision Tree performs just slightly worse than Bala's Probing FCNN, with accuracy slightly increasing from 73.87% by about 2% as iterations are increased from the $N = 40$ used by Bala to the $N = 200$ for which they generated features. AdaBoost's accuracy increases at a similar rate with features from more iterations, indicating that features from iterations beyond 40 do in fact improve predictions, contrary to Bala's claim [2]. Moreover, even the lowest AdaBoost accuracy of 80.38% with data from 40 iterations is far higher than the 74% of Probing FCNN, implying a neural network may be unnecessarily complicated to classify data represented with these features.

Applying the insights gained from this first experiment, all further experiments used FDA features from $N = 200$ generations. Since a single Decision Tree did not make predictions as well as the ensemble learner AdaBoost, that predictor was replaced with Random Forest [17] as another ensemble method. Once hyperparameter tuned, AdaBoost used 100 estimators and reached a mean accuracy of 82.9% with Random Forest using already optimal default scikit-learn [32] hyper-parameters close behind at 81.82% on Bala's dataset, exceeding Tupor's highest achieved accuracy of 79% [40]. Due to the higher accuracy and faster computation times, AdaBoost was used for all further experiments.

Iterations	Probing FCNN [2]	Decision Tree	AdaBoost [11]
40	0.74	0.7387	0.8038
50		0.7339	0.8037
100		0.7501	0.8197
150		0.7589	0.8206
200		0.7585	0.8244

Table 5.1: Comparison on Bala's dataset of Probing FCNN [2] accuracy to mean accuracy from 5-fold cross validated tests on the same dataset with a Decision Tree and AdaBoost [11] using features from various numbers of iterations.

5.2 Impact of ROI Extraction Kernel

In their thesis, Bala describes how regions of interest of an ultrasound image are extracted with a Gaussian kernel $g(x, y)$ [2]. Once converted to the frequency domain, the convolution with the Fourier Transform of a Gaussian kernel $G(u, v)$ is then removed during the extraction of frequency coefficients. However, a careful inspection of their source code revealed that they applied a rectangular kernel $d(x, y)$ for ROI extraction instead of a Gaussian kernel, while still removing the hypothetical convolution with the Fourier Transform of a Gaussian kernel $G(u, v)$ during frequency coefficient reconstruction. Nonetheless their method generates data that allows for very precise classification of regions of interest as shown by the previous section and their thesis. This observation raises the questions why the predictions are that good and how the outcome would change if a Gaussian kernel had been applied to extract ROIs. Additionally, how would the predictor perform if the rectangular kernel was used for ROI extraction and again later to remove convolution with it in the frequency

	Coarse Label	Fine Label
$F(u,v) = H(u,v) \oslash G(u,v)$	0.8684	0.8853
$F'(u,v) = H'(u,v) \oslash D(u,v)$	0.8622	0.8837
$F''(u,v) = H'(u,v) \oslash G(u,v)$	0.8641	0.8902

Table 5.2: Mean accuracy from 5-fold cross validated tests comparing AdaBoost predictions based on reconstructed frequency coefficients $F(u,v)$, $F'(u,v)$, and $F''(u,v)$ as defined in Section 4.2.3.

domain.

To answer these questions, AdaBoost was trained to compare the predictions based on reconstructed frequency coefficients $F(u,v)$, $F'(u,v)$, and $F''(u,v)$ where $H(u,v)$ denotes the Fourier Transform of an ROI extracted with a Gaussian kernel $g(x,y)$ and $H'(u,v)$ denotes the Fourier Transform of an ROI extracted with a rectangular kernel $d(x,y)$ and $G(u,v)$ and $D(u,v)$ are the Fourier Transforms of these kernels. $\oslash$ denotes the de-convolution operation as defined in Section 3.3.4. The results shown in Table 5.2 indicate that mean accuracy is very similar with both the coarse and fine class labels. Using Dietterich's [8] 5x2 paired t-test method to compare these results leads to the p-values 0.4005 when comparing $F(u,v)$ and $F'(u,v)$, 0.6364 when comparing $F'(u,v)$ and $F''(u,v)$, and 0.7339 when comparing $F(u,v)$ and $F''(u,v)$, far exceeding the commonly applied level of significance $\alpha = 0.05$. That means the null hypothesis - no significant difference in predictions from the tested predictor pair - cannot be rejected for any pair and there is most likely no significant difference between AdaBoost predictions on the three datasets. $F(u,v)$ and $F'(u,v)$ are similar as was expected, because the concept of the used de-convolution process from Liu [22] is removing the convolution with the extraction kernel, resulting in the reconstructing of the same image in two different ways. A possible explanation for why $F''(u,v)$ yields just as good or better results is the fact that the Fourier Transformed kernels $G(u,v)$ and $D(u,v)$ are actually very similar when compared in the frequency domain. Another potential cause could be that the noisy and blurred nature of ultrasound images is similar to the effect applying a Gaussian kernel to an ROI would have even when extracting it with a rectangular kernel. By performing de-convolution using the Fourier transform of a Gaussian kernel, some of the Gaussian-like noise already present in ultrasound images [27] that otherwise would not be removed by the de-convolution method may also be removed producing meaningful data for predictions.

Regardless of the reason, all further experiments used data generated without applying a Gaussian kernel during the window extraction step for consistency with Bala's method.

5.3 Feature Map Selection

Once the choice of predictor and method for dataset generation were finalized, experiments were conducted to evaluate the additional features that were added beyond what Bala [2] had used.

5.3.1 Angle Representation

The calculated angles from frequency coefficients for spectral rotation and deviation were expressed both in degrees and as tuple of $sin(\theta)$ and $cos(\theta)$ of the respective angle θ. While degrees are easier for humans to interpret, they do not represent the periodic properties of angles very well, whereas a tuple of sin and cos of an angle does because both functions are also periodic and describe the unit circle. AdaBoost was trained using these representations separately and together. The experiment results in Table 5.3 indicate only marginal mean accuracy differences when angles were represented by the tuple $sin(\theta)$ and $cos(\theta)$ of an angle θ as opposed to degrees, or both at once. This suspicion was confirmed after performing another significance test to compare the fine label classifiers using different angle representations and to compare using the tuples to using both angle representation. The calculated p-values were 0.5576 and 0.5247 respectively, with both values well above $\alpha = 0.05$ meaning it is quite likely that all three classifiers have no significant differences [8]. Considering each dataset entry contains 400 angles, spectral rotation and deviation angles for each frequency coefficient iteration, the 400 features representing those angles in degrees were omitted in the following experiments to reduce computation time and the large feature map.

Angle Representation	Coarse Label	Fine Label
$sin(\theta)$ And $cos(\theta)$	0.8690	0.8909
Degrees	0.8684	0.8901
Both	0.8688	0.8911

Table 5.3: Mean accuracy from 5-fold cross validated tests evaluating AdaBoost predictions using angles represented in number of degrees, as tuple of $sin(\theta)$ and $cos(\theta)$ of the respective angle, and with both representations.

5.3.2 Location Information

In regards to the location of each ROI, two different ideas were evaluated. The first experiment with results shown in Table 5.4 tests that the predictor is not just learning coordinate patterns from the location features, but is leveraging both location and FDA features for better predictions. A predictor was trained on just the FDA data and scored against another predictor trained solely on slice index, coordinates, and U-Net accuracy.

The results for the FDA data only predictor align with the findings from Section 5.1: With a more coarse class label, accuracy is lower than 82.9%, which was expected because Bala's dataset used the exact center of ROI ground truths as center of the extracted ROI [2]. The center of the U-Net determined ROIs can be offset by up to 3mm (11.33 pixels) from the ground truth center in the coarsely labelled ROIs. On the other hand, the predictor trained on finely labelled ROIs has a higher accuracy of around 84.11%, even though the ROIs are still more coarse than for Bala's data. This improvement is likely caused by the larger training set.

	Coarse Label	Fine Label
FDA Features Only	0.8142	0.8411
No FDA Features	0.8582	0.8763
Both	0.8692	0.8913

Table 5.4: Mean accuracy from 5-fold cross validated tests verifying AdaBoost predictions with FDA features only, excluding FDA features, and with all features.

Ignoring all FDA features, the predictor achieved a higher mean accuracy than with only FDA features. Simply learning coordinate patterns of ROIs combined with the U-Net confidence for the center of that ROI candidate appears to be a viable strategy. A 5x2 paired t-test further confirmed this observation with the p-value of 0.0004 [8]. It is so far below $\alpha = 0.05$ that a significant difference can safely be assumed. Combining all data, however, improves the accuracy substantially using both the coarse and the fine class labels, as supported by another 5x2 paired t-test with a p-value of 0.0005 well below $\alpha = 0.05$.

The next step was about comparing the model on ROI locations in cartesian and polar coordinates to see which representation is more beneficial in this domain. The differences in accuracy shown in Table 5.5 using only features from one coordinate system are marginal with polar coordinates appearing to facilitate slightly better predictions, but a significance test was necessary. As expected, the 5x2 paired t-test resulted in a p-value of 0.4207 above the $\alpha = 0.05$ threshold and a significant difference is very unlikely. Using both coordinate systems together seems to be another marginal improvement, but a 5x2 paired t-test comparison of the classifier using both location

Location Representation	Coarse Label	Fine Label
Cartesian Coordinates	0.8655	0.8894
Polar Coordinates	0.8673	0.8902
Both	0.8687	0.8906

Table 5.5: Mean accuracy from 5-fold cross validated tests evaluating AdaBoost predictions using ROI locations represented in cartesian coordinates, polar coordinates, and with both.

representations to the one using polar coordinates surprisingly leads to a p-value of 0.0466, meaning the classifiers are not substantially identical. Consequently, neither coordinate features were omitted to fully utilize the higher value of both coordinate representations.

5.4 Final Test Results

ROI candidates are generated by multiple U-Net models and there is no guarantee that all true catheters are actually included in those candidates. As mentioned in Section 4.2.2, it is important to consider the percentage of true catheters detected among those ROIs and to weigh classification accuracies accordingly. Using the coarse ROI classification, 92.68% of true catheters are among the candidates. Weighting the best cross-validated mean accuracy of 86.92% determined in the previous experiments with this percentage leads to an overall detection rate of 80.56% for this combined U-Net and FDA model classifier. In the case of fine ROI classification, only 81.03% of true catheters are detected and with an accuracy of 89.13% for fine predictions this results in an overall detection rate of 72.77% for the combined model classifier.

Since ROI predictors trained on fine labels have consistently far higher accuracy than those trained on coarse labels, a last experiment was performed training a predictor on fine class labels, but testing it on coarse labels. This improves the accuracy of predictions on the coarse labels by over 1% to 87.94%. Leveraging the high accuracy of fine labels combined with the higher percentage of true catheters among the coarse ROI candidates leads to a final 2D combined U-Net and FDA model classifier accuracy of 81.5%.

Chapter 6

Conclusion

Prostate cancer is one of the most common types of cancer and yet, its brachytherapy treatment relies heavily on difficult and error prone manual segmentation of ultrasound images [43, 47]. This thesis proposes a fully automatic novel approach that combines two individually successful models in this domain, leveraging the benefits of both to help overcome their respective disadvantages. An optimized U-Net [36] implementation performs pixel-wise predictions on unaltered and contrast enhanced 2D prostate ultrasound images to generate a pool of regions of interest that could be catheters. This pool is purposely large to minimize the number of false negatives with the trade-off of increasing the number of false positives as false positives can later be removed when evaluated by the second predictor or during 3D reconstruction [23]. Every ROI is then extracted using a Gaussian and rectangular kernel and Fourier Transformed into the frequency domain. A modified version of Liu's [22] spectrum reconstruction method is applied to remove convolution and noise from the transformed ROIs. A number of relevant features are calculated from the extracted frequency coefficients as proposed by Bala [2]. These features are combined with the previously determined ROI centroid's U-Net confidence, location, and slice index and finally predicted by an AdaBoost predictor.

Optimized AdaBoost predictions have a 2D accuracy of 89.13% and 86.92% for ROI centroids within 2 mm and 3 mm of ground truth catheter centroids respectively. Due to false negatives in U-Net predictions, the overall accuracy of correctly identified and classified catheters is 72.77% and 80.56% for 2 mm and 3 mm tolerance respectively. When trained on fine 2 mm tolerance data and tested on coarse 3 mm tolerance data, the overall 2D model accuracy improves to 81.5%, exceeding the 2D accuracies 79% and 74% previously achieved by Tupor [40] and Bala [2].

6.1 Limitations and Future Work

Despite highly accurate AdaBoost predictions, the overall 2D model accuracy is reduced by the still substantial percentages of 18.97% and 7.32% false negatives during U-Net [36] ROI candidate generation for fine 2 mm and coarse 3 mm tolerance respectively. The reasons are most likely twofold. As explained in Section 4.1.2, some catheters are obscured while still being labelled catheters. Beyond these hard to detect catheters, U-Net is generally expected to make some errors.

Since the original catheter grid insertion positions are known, this knowledge could be passed to the model once at the beginning or during a brachytherapy session. An increasing number of additional ROI candidates could be added to the U-Net generated pool in the area around each expected catheter grid position depending on the 2D image slice index (catheter depth) where deeper slices add more candidates as the catheter location could deviate more from the original grid position. This would essentially eliminate false negatives during that part of the model and also preserve the potentially valuable U-Net confidence as AdaBoost prediction feature. While it is not given that AdaBoost will correctly classify these hard to detect catheters with the same accuracy it has for the U-Net detected catheters, even a single correctly classified ROI that U-Net could not detect would improve the accuracy of the proposed model.

A more drastic approach would be replacing the U-Net model component entirely with an alternative method to generate ROI candidates more reliably, such as the one explained above. However, losing the U-Net confidence as feature for the proposed AdaBoost predictor may negate some or all of the benefit of finding more false negatives.

The proposed AdaBoost predictor was applied to three different datasets based on different combinations of ROI extraction kernels and subsequently used kernels during spectrum reconstruction, however, it was optimized on only one of these datasets. Future research should optimize AdaBoost on each dataset individually, especially in regards to the number of iterations N during the spectrum reconstruction process. The underlying case where the used $N = 200$ is optimal possibly requires that many frequency coefficients, because some coefficients are not from the original image, but convolution from the extraction kernel due to the mismatched kernel combination. The other two datasets using matching kernel combinations may not need as many iterations and might even have noise introduced from too many iterations. In that context, it may be worth exploring entirely new combinations of kernels for ROI extraction and spectrum reconstruction.

The most logical next step for the proposed model is utilizing the 2D predictions to reconstruct a 3D prostate model and evaluate catheter detection in 3D space. The 3D reconstruction method by Liu et al. [23] could natively input 2D data from the proposed model and they indicated an expected accuracy improvement of up to 95% in their 3D predictions due to higher 2D prediction accuracy and the lower number of false negatives compared to the currently underlying 2D predictions from Tupor [40]. With clinically diverse underlying data provided by the Saskatchewan Cancer Agency [1, 23] leading to more robust predictions, the theoretical 3D accuracy of that magnitude would put the model on par with or ahead of the results from Wang et al. [43] and Zhang et al. [47].

Bibliography

[1] Saskatchewan Cancer Agency. `http://www.saskcancer.ca/`.

[2] Gopal Bala. *Distinguishing Catheters from their Echoes in Ultrasound Images.* Master's thesis, University of Regina, 2020.

[3] Pablo J. Blanco, Paulo G.P. Ziemer, Carlos A. Bulant, Yasushi Ueki, Ronald Bass, Lorenz Räber, Pedro A. Lemos, and Héctor M. García-García. Fully automated lumen and vessel contour segmentation in intravascular ultrasound datasets. *Medical Image Analysis*, 75:102262, 2022.

[4] Wilhelm Burger and Mark J. Burge. *Digital Image Processing: An Algorithmic Introduction Using Java.* Springer Publishing Company, Incorporated, 2nd edition, 2016.

[5] Kinder Chen. Denoising data with fast fourier transform. `https://kinder-chen.medium.com/denoising-data-with-fast-fourier-transform-a81d9f38cc4c`, 2020. Accessed on April 18, 2023.

[6] Mike Cohen. Time and frequency domains. `https://www.youtube.com/watch?v=fYtVHhk3xJ0`, 1019. Accessed on April 18, 2023.

[7] Robert I. Cowles. *An Adjustable Gaussian Filter for Fourier Transformation of Arbitrary Image Regions*. Master's thesis, University of Regina, 2002.

[8] Thomas G. Dietterich. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

[9] Daniel P.W. Ellis. An introduction to signal processing for speech. Technical report, LabROSA, Columbia University, New York, 2008.

[10] David Fleet and Allan Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[11] Yoav Freund and Robert E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In Paul Vitányi, editor, *Computational Learning Theory. Proceedings of the Second European Conference*, pages 23–37, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

[12] Mino George and H. B. Anita. Analysis of kidney ultrasound images using deep learning and machine learning techniques: A review. In G. Ranganathan, Robert Bestak, Ram Palanisamy, and Álvaro Rocha, editors, *Proceedings of the*

International Conference on Pervasive Computing and Social Networking, pages 183–199, Singapore, 2022. Springer Nature Singapore.

[13] Stav Gofer, Oren Haik, Ron Bardin, Yinon Gilboa, and Sharon Perlman. Machine learning algorithms for classification of first-trimester fetal brain ultrasound images. *Journal of Ultrasound in Medicine*, 41(7):1773–1779, 2022.

[14] Rafael Gonzales and Richard Woods. *Digital Image Processing*. Pearson Education, 4th edition, 2018.

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`. Accessed on April 18, 2023.

[16] Abdul M. Hafiz and Ghulam M. Bhat. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval*, 9(3):171–189, 2020.

[17] Tin K. Ho. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282. IEEE, 1995.

[18] Ming Jiang and Ge Wang. Development of blind image deconvolution and its applications. *Journal of X-Ray Science and Technology*, 11:13–19, 2003.

[19] Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.

[20] Yann LeCun and Yoshua Bengio. *Convolutional Networks for Images, Speech, and Time Series*, page 255–258. MIT Press, Cambridge, MA, USA, 1998.

[21] Menghan Li, Bin Huang, and Guohui Tian. A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, 110:104669, 2022.

[22] Bingyang Liu. *Probing Detector for Image Local Frequency Analysis*. Master's thesis, University of Regina, 2017.

[23] Derek Liu, Shayantonee Tupor, Jaskaran Singh, Trey Chernoff, Nelson Leong, Evgeny Sadikov, Asim Amjad, and Sandra Zilles. The challenges facing deep learning–based catheter localization for ultrasound guided high-dose-rate prostate brachytherapy. *Medical Physics*, 49(4):2442–2451, 2022.

[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440. IEEE, 2015.

[25] Zhihan Lv, Liang Qiao, Shuo Yang, Jinhua Li, Haibin Lv, and Francesco Piccialli. Memory-augmented neural networks based dynamic complex image segmentation in digital twins for self-driving vehicle. *Pattern Recognition*, 132:108956, 2022.

[26] Jinming Ma, Ran Tao, and Yongzhe Li. Spectrum analysis for multiband signals with nonuniform sub-nyquist sampling in the fractional fourier domain. *IEEE Transactions on Signal Processing*, 70:3632–3646, 2022.

[27] Oleg V. Michailovich and Allen Tannenbaum. Despeckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 53(1):64–78, 2006.

[28] Ihab S. Mohamed. *Detection and Tracking of Pallets using a Laser Rangefinder and Machine Learning Techniques*. Master's thesis, University of Genova, 2017.

[29] Shree Nayar. Image filtering in frequency domain. `https://www.youtube.com/watch?v=OOu5KP3Gvx0`, 2021. Accessed on April 18, 2023.

[30] Wayne Niblack. *An Introduction to Digital Image Processing*. Strandberg Publishing Company, 1985.

[31] Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Lee Matthew C. H., Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla,

Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. *arXiv e-prints*, 1804.03999, 2018.

[32] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[33] Robi Polikar. *The Wavelet Tutorial*. Academic Press, 2nd edition, 2001.

[34] S. Qian and D. Chen. Discrete gabor transform. *IEEE Transactions on Signal Processing*, 41(7):2429–2438, 1993.

[35] G. Rajakumar and T. Ananth Kumar. *Design of Advanced Security System Using Vein Pattern Recognition and Image Segmentation Techniques*, pages 213–225. Springer Singapore, Singapore, 2022.

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, Cham, 2015. Springer International Publishing.

[37] Mohammad Shehab, Laith Abualigah, Qusai Shambour, Muhannad A. Abu-Hashem, Mohd K. Y. Shambour, Ahmed I. Alsalibi, and Amir H. Gandomi.

Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, 145:105458, 2022.

[38] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, 1409.1556, 2014.

[39] Sik-Ho Tsang. Review: Deepmask (instance segmentation). https://towardsdatascience.com/review-deepmask-instance-segmentation-30327a072339, 2018. Accessed on April 18, 2023.

[40] Shayantonee Tupor. *Segmentation of Ultrasound Images Using Deep Neural Networks*. Master's thesis, University of Regina, 2021.

[41] Cancer Research UK. Review: Deepmask (instance segmentation). https://commons.wikimedia.org/w/index.php?curid=46505428, 2016. Accessed on April 18, 2023.

[42] Mathuranathan Viswanathan. Understanding fourier series. https://www.gaussianwaves.com/2013/05/a-tutorial-on-fourier-analysis-fourier-series/, 2012. Accessed on April 18, 2023.

[43] Fuyue Wang, Lei Xing, Hilary Bagshaw, Mark Buyyounouski, and Bin Han. Deep learning applications in automatic needle segmentation in ultrasound-guided prostate brachytherapy. *Medical Physics*, 47(9):3797–3805, 2020.

[44] Yingying Wang, Yibin Li, Yong Song, and Xuewen Rong. The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, 10(5), 2020.

[45] Qi Xiao, Rui Wang, Hongyu Sun, and Limin Wang. Objective evaluation of fabric pilling based on image analysis and deep learning algorithm. *International Journal of Clothing Science and Technology*, 33(4):495–512, 2021.

[46] Yu-Jin Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

[47] Yupei Zhang, Yang Lei, Richard L.J. Qiu, Tonghe Wang, Hesheng Wang, Ashesh B. Jani, Walter J. Curran, Pretesh Patel, Tian Liu, and Xiaofeng Yang. Multi-needle localization with attention U-Net in us-guided hdr prostate brachytherapy. *Medical Physics*, 47(7):2735–2745, 2020.