Master Thesis

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Computer Science

Improving atherosclerotic plaque segmentation and estimating their clinically relevant parameters

Bc. Ondřej Stejskal

Supervisor: prof. Dr. Ing. Jan Kybic
Field of study: Open Informatics
Subfield: Artificial Intelligence
January 2024

MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Stejskal Ondřej** Personal ID number: **483739**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Computer Science**

Study program: **Open Informatics**

Specialisation: **Artificial Intelligence**

II. Master’s thesis details

Master’s thesis title in English:

Improving atherosclerotic plaque segmentation and estimating their clinically relevant parameters

Master’s thesis title in Czech:

Zlepšení segmentace aterosklerotických plátů a odhad jejich klinicky relevantních parametrů

Guidelines:

This project is about automatic processing of ultrasound images of the carotid artery. The main goal is automatic evaluation of medically relevant parameters. We shall try to improve the segmentation of the atherosclerotic plaque, which has a major effect on the parameters of interest.
Get acquainted with the available dataset of ultrasound images and segmentations and the existing segmentation algorithm.
Get acquainted with existing methods for improving segmentation quality by incorporating shape models. Choose a suitable method based for example on using autoencoders (AE) or generative adversarial networks (GAN) to model the segmentation prior. Test on available expertly segmented images Implement methods to evaluate the geometric parameters from known segmentations and test them on synthetic data and on automatic segmentation.
For non-geometric parameters, train a neural network (e.g., ResNet) to estimate them from suitable data (usually image + segmentation mask for the plaque). Try several masking strategies Test on automatic segmentations on the key frames, compare with human annotations
Evaluate the improvement of the parameter evaluation thanks to the improved segmentation method(s).

Bibliography / sources:

Valvano et al: Learning to Segment from scribbles using multi-scale adversarial attention gates. IEEE TMI, August 2021
Oktay et al: Anatomically constrained neural networks. Application to Cardiac Image Enhancement and Segmentation. IEEE TMI, Feb 2018.
Kostelanský et al. "Differentiating between stable and progressive carotid atherosclerotic plaques from in-vivo ultrasound images using texture descriptors." SIPAIM: International Symposium on Medical Information Processing and Analysis, 2021.
Kostelanský M. "Localization and segmentation of in-vivo ultrasound carotid artery images.", diplomová práce, ČVUT FEL, 2021.
Školoudík, D. et al, “Visual and digital analysis of the ultrasound image in a stable and progressive carotidatherosclerotic plaque,” Česká a slovenská neurologie a neurochirurgie (1), 38–44 (2021).

Name and workplace of master’s thesis supervisor:

prof. Dr. Ing. Jan Kybic   Biomedical imaging algorithms FEE

Name and workplace of second master’s thesis supervisor or consultant:

Date of master’s thesis assignment: **19.09.2023**   Deadline for master's thesis submission: **09.01.2024**

Assignment valid until: **16.02.2025**

prof. Dr. Ing. Jan Kybic Supervisor’s signature	Head of department’s signature	prof. Mgr. Petr Páta, Ph.D. Dean’s signature

III. Assignment receipt

The student acknowledges that the master’s thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt   Student’s signature

Acknowledgements

I would like to thank my supervisor prof. Dr. Ing. Jan Kybic for the guidance, support, and expertise which was invaluable. I also would like to thank my family for the patience they have and the support they gave me.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

In Prague, 9. January 2024

Abstract

This thesis aimed to enhance the segmentation of atherosclerotic plaques and evaluate their clinically relevant parameters based on these improved segmentation results. We selected and adapted three methods for our segmentation task: ACNN, which combines an autoencoder with a classical segmentation network; MAAG, a GAN-inspired method for scribble annotations; and SeGAN also a GAN-inspired method. These methods were implemented and adjusted to our dataset of 150 transversal ultrasound images. The MAAG method, using a U-Net segmentor and a convolutional autoencoder, achieved the highest mean IoU of 71.9%. And MAAG achieved highest accuracy of 91.4%. U-Net alone reached a mean IoU of 70% and an accuracy of 86.8%.

Segmentations produced by these models were used for the parameter estimation methods which classified homogeneity with accuracy of 63% and echogenecity with accuracy of 36% and plaque width correlation of 0.46.

Keywords: Carotid artery stenosis, Image segmentation, Anatomical Prior, Generative Adversarial Network, Convolutional Networks, Autoencoder

Supervisor: prof. Dr. Ing. Jan Kybic

Abstrakt

Cílem této práce bylo zlepšit segmentaci aterosklerotických plátů a na základě těchto zlepšených výsledků segmentace vyhodnotit jejich klinicky relevantní parametry. Pro naši segmentační úlohu jsme vybrali a upravili tři metody: ACNN, která kombinuje autoenkodér s klasickou segmentační sítí a metody MAAG a SeGAN inspirované GAN metodou. Tyto metody byly implementovány a upraveny pro naši datovou sadu 150 transverzálních ultrazvukových snímků. Metoda MAAG, využívající segmentátor U-Net a konvoluční autoenkodér, dosáhla nejvyšší průměrné hodnoty IoU 71,9%. A MAAG dosáhla nejvyšší přesnosti 91,4%. Samotná síť U-Net dosáhla průměrné hodnoty IoU 70% a přesnosti 86,8%.

Segmentace vytvořené těmito modely byly použity pro metody odhadu parametrů, které klasifikovaly homogenitu s přesností 63% a echogenitu s přesností 36% a korelaci pro šířku pláště rovnou 0,46.

Klíčová slova: Stenóza karotické tepny, Segmentace obrazu, Anatomická předloha, Generativní adversariální síť, Konvoluční síť, Autoenkodér

Překlad názvu: Zlepšení segmentace aterosklerotických plátů a odhad jejich klinicky relevantních parametrů

Contents

Figures

Tables

Chapter 1

Introduction

This thesis aims to address the challenge of enhancing the segmentation of ultrasound images of the carotid artery, specifically the atherosclerotic plaque and estimating their relevant parameters. Ultrasound imaging, a vital tool in diagnosing vascular diseases, often faces limitations due to the quality of the images as well as very expensive manual labeling, which is a challenging task even for experienced medical professionals. Improved automatic segmentation of these images is essential for accurate diagnosis and treatment planning. Furthermore the estimation of relevant parameters from ultrasound medical images is crucial to accurately diagnose and to predict future progress of atherosclerotic plaques.

In this work we explore advanced image processing techniques and machine learning algorithms, such as Autoencoders and Generative Adversarial Networks [16], in order to achieve more precise and reliable segmentations. We also explore methods to reliably estimate clinically relevant parameters of the atherosclerotic plaque images, which we will estimate from the segmentations and using classification neural networks.

Outline of the Thesis

Chapter 2 describes the medical background of the atherosclerotic plaque and stenosis. With this background we motivate the tasks of segmentation and parameter estimation in the context of stenosis prediction. And describe the available dataset of carotid ultrasound images.

Segmentation, classification and regression are well-studied topics in the realm of computer imaging. In Chapter 3, we provide a review of the literature for medical image segmentation with a focus on anatomical prior and methods that take advantage of the prior. In Chapter 3, We also review the classification and regression problems.

In Chapter 4, we describe the methods and implementation details for the task of segmentation and present the results.

In Chapter 5, we describe the methods and implementation details for the task of parameter estimation and present the results.

In the last Chapter 6, we conclude the results.

Chapter 2

Motivation

2.1 Medical background

The vascular supply to the brain is provided by two main paired arteries, the arteria carotis interna (ACI) and the arteria vertebralis, which join at the base of the skull to form a circuit that supplies all parts of the brain.

The ACI supplies up to 85% of the blood to the brain and its narrowing or stenosis, even a complete closure, can lead to very serious brain damage. The most common cause of ACI stenosis is atherosclerosis, an inflammatory disease that damages the inner layer of the blood vessel. In this disease, fatty substances, primarily cholesterol, are deposited beneath the inner layer, forming a so-called atherosclerotic plaque that arches into the interior of the vessel and leads to its narrowing. This can then lead to restriction of blood flow to the brain, complete blockage of the vessel, and even the development of a stroke. Plaques that reduce the diameter of the vessel by at least 75% are considered to be significant stenosis .

Besides fat and inflammatory cells, atherosclerotic plaques may contain varying percentages of fibrin and calcium, which leads to the formation of so-called calcifications. The more of these substances a plaque contains, the more stable it is and it can grow inwards, leading to severe stenosis.

If it contains primarily fatty substances, it is an unstable atherosclerotic plaque, which can break off and travel down the bloodstream to another organ, leading to a blockage of the supplying artery and its damage.

Risk factors that influence the development of atherosclerosis include an unhealthy lifestyle, high blood pressure, smoking, diabetes mellitus or other metabolic diseases.

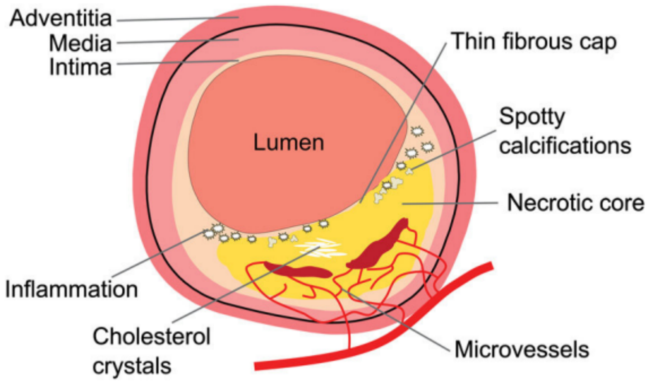


Figure 2.1: Scheme of the carotid with atherosclerotic plaque

2.2 Geometric and Non-Geometric parameters

The atherosclerotic plaques are defined by their geometric characteristics such as shape, size, or width. Using the ultrasound imaging methods, we are able to measure the parameters such as homogeneity, which characterizes uniformity and consistency of tissue. Echogenicity which characterizes the ability of tissue to reflect ultrasound waves. Calcification expresses the buildup of calcium in the plaque. The surface of the plaque is characterized as smooth or coarse. The plaque can also be ulcerated meaning, that the plaque has developed an ulcer or a break in its surface.

According to [53] risk factors for atherosclerotic progression plaquesin carotid arteries include: greater width of the atherosclerotic plaque corresponding to the percentage of stenosis of the artery and uneven, or ulcerated plaque surface.

2.3 Dataset

All of the datasets we will use in this thesis are derived from the ANTIQUE dataset. The ANTIQUE dataset was developed as part of the study titled "Atherosclerotic Plaque Characteristics Associated With a Progression Rate of the Plaque in Carotids and a Risk of Stroke" [53] conducted between 2015 and 2020. This study involved 413 patients who were monitored at the University Hospital Ostrava and the Military University Hospital in Prague. The participants, aged between 30 and 90, were all diagnosed with stenosis greater than 30%. The dataset, in its unprocessed form, comprises images from individual patient examinations, totaling 1,322 exams and encompassing 28,178 ultrasound scans. Ultrasound scans are divided into four main types: Longitudinal, Transversal, Conical and Doppler. We will work with the transversal type, example can be seen at figure 2.2

Medical professionals evaluate the parameters of the ultrasound images, these annotations contains are in their raw form available for 25765 images that belong to 1321 patients. Experts evaluate the width of the atherosclerotic plaque, homogeneity, echogenicity, type of surface, calcification the evaluation also includes information about age, weight and other factors.

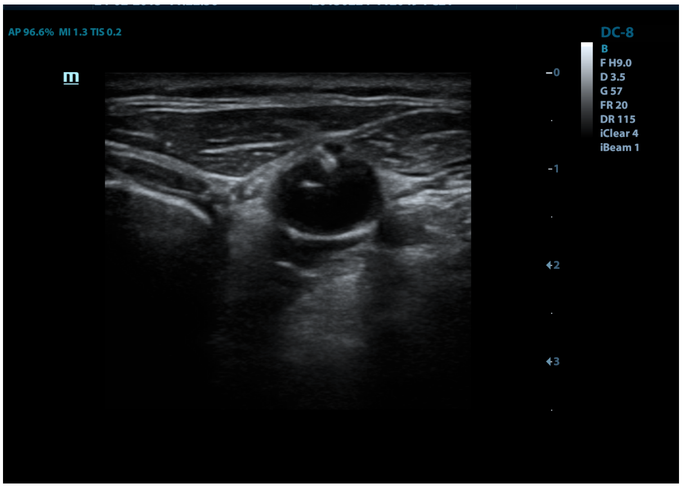


Figure 2.2: Transversal ultrasound image of carotid artery

Chapter 3

Related Work

3.1 Segmentation

Segmentation is a fundamental task in computer vision. It refers to assigning a class label to each pixel in the image. The purpose of the segmentation is to divide the image into multiple semantic parts that represent the image in an interpretable way. Segmentation is used in many fields such as autonomous driving [8] and face recognition [22]. In medical imaging, the segmentation is used to provide interpretable information and analysis for the medical images, segmentation is used to localize and detect abnormalities such as tumors [38], pathological tissue [3] or organs abnormalities [20].

3.1.1 U-Net

U-Net is a convolutional neural network architecture originally developed for medical image in 2015 [40] particularly for segmenting cellular structures in microscopic images. It is used as a benchmark method that most of the works in medical image segmentation is compared to. U-Net is defined by its U-shape architecture. This architecture allows U-Net to use high-level features from deeper layers alongside low-level features from earlier layers, resulting in precise and accurate segmentation, with ability to work with limited amount of data which is a useful characteristic for medical images.

3.1.2 Anatomical Prior

An anatomical prior refers to the pre-existing knowledge about the structure and typical arrangement of anatomical features within the human body. This prior information can be used to improve the segmentation and yield more plausible and real segmentations. For medical image segmentation, the prior is especially useful [35], as the shape and location of the organs veins and other structures are given. Incorporating anatomical priors can be done by various methods such as atlas models [2]. These methods have limitation in handling the variation in human anatomy and are usually computationally intensive for CPU's.

The anatomical prior can also be incorporated into the deep learning methods, more specifically into convolutional neural networks. CNN's overcome the problem of generalization and can be effectively executed using GPU's. One strategy to employ anatomical prior in is to postprocess the images using conditional random fields [9] this methods usually works with local constraints limiting the global context. Other methods that are able to utilize anatomical prior are frameworks using Autoencoders (AE) or Generative Adversarial Networks (GAN) to constrain the segmentation.

3.1.3 Autoencoders and Convolutional Autoencoders

Autoencoders are a family of neural networks that learn efficient data encodings. The autoencoder consists of an encoder that compresses the input into lower-dimensional latent space and a decoder that reconstructs it. The autoencoders aim to minimize the difference between the input and output images. For computer imaging methods is avantageous to use convolutional layers in the autoencoders, creating a Convolutional Autoencoder (CAE). In medical imaging the CAE's can be used to denoise the images [15]. To augment images [21] or to detect anomalies [43].

CAE Segmentation using autoencoders

Since the autoencoders learn compressed encoding, the reconstructed images are more constrained in their shape and overall characteristics. One option is to use autoencoders as a post-processing method [28]. The image is segmented using a standard supervised segmentation algorithm and then the output is processed to be more anatomically plausible using the autoencoder. Another method [10] is to train Variational Autoencoder [24] and use part of the weights in the segmentation network [10]. Another option [36] is to use the encoding as a reguralizer, where the segmentation loss is joined by the difference between encoding of the segmented image and encoding of the true label.

3.1.4 Generative Adversarial Networks

Generative Adversarial Network (GAN) is a framework for training deep learning networks proposed in 2014 by Goodfellow and his colleagues [16]. The architecture consists of two neural networks namely a generator and a discriminator. The training process is inspired by game theory. These two nets are trained simultaneously through adversarial processes. The generator's role is to create data that is indistinguishable from real data, while the discriminator's job is to distinguish between the generator's fake data and real data. The generator improves its ability to produce realistic data, and the discriminator gets better at distinguishing between real and generated data. GANs excel at generating realistic images. In medical imaging they can be used to augment images and generate new realistic images [45].

Segmentation using GANs

GANs were previously successfully used in segmentation [47]. In medical image segmentation, many research works are using the GANs framework to improve segmentation [51]. Many different architectures are possible, commonly in GAN setting the classic segmentation algorithm acts as a generator and generates segmentations from the input images. Subsequently these segmentations or masked images are fed into the discriminator, which evaluates these segmentation. In [50] the Segmentor/Critic architecture is proposed where the segmentor creates segmentations and the critic receives masked images from these segmentations. Critic proudces a multi-scale vector for real and for the fake segmentations. From these two vectors the mult-scale L1 loss is computed. UENet [41] proposes an U-Net-like segmentor and E-shapes discriminator. The advantage of GANs use in the segmentation task is that it allows weak and semi-supervised methods in [49] both of these settings are proposed using scribbles as weak segmentation labels and using unpaired segmentations in unsupervised steps. More complex GAN architectures are also used in medical image segmentation, such as pix2pix [42] or Wasserstein GAN [19]

3.2 Parameter Estimation

To estimate the parameters using neural networks, we can think of it as a classification or regression problem based on the estimated parameter. Due to this, the classification and the regression task will be reviewed separately

3.2.1 Classification

Image classification is a fundamental task in image processing, it involves labeling images into predefined classes based on their characteristics. The Deep learning methods, especially convolutional Neural Networks (CNN) made an significant impact in the field. The ImageNet [11] benchmark is a pivotal dataset in the field of image classification, consisting of millions of labeled images across thousands of categories. The most influential CNNs are AlexNet [27], VGG [44] notable for its deep architecture, and ResNet [18], which introduced the concept of residual learning. In medical imaging, the CNNs have been applied in many domains [7] such as predicting breast cancer [48] or to diagnose pneumonia [52] or to diagnose COVID-19 from X-Ray chest images [30].

3.2.2 Regression

Regression tasks using neural networks involve predicting continuous outcomes based on input data. These tasks differ from classification tasks, which predict discrete labels. Neural networks, particularly deep learning models, excel at regression due to their ability to model complex, nonlinear relationships. They are widely used in various fields, such as predicting financial trends [6], forecasting weather [5], or to predict the age from the image [12]. The difference between the regression and classification task is the output type and the loss that is computed using different loss function most notably the mean squared error (MSE) loss function is used.

3.2.3 ResNet

ResNet, short for Residual Network, is an influential neural network architecture that was introduced in 2015 [18]. What sets ResNet apart is its novel use of residual connections, which tackle the problem of vanishing gradients—a significant challenge in training very deep networks [37]. This is achieved by adding shortcut connections that skip one or more layers. The ResNet has many variants specified by the number of convolutional layers ranging from ResNet-18 to ResNet-152.

Chapter 4

Segmentation of carotid artery images

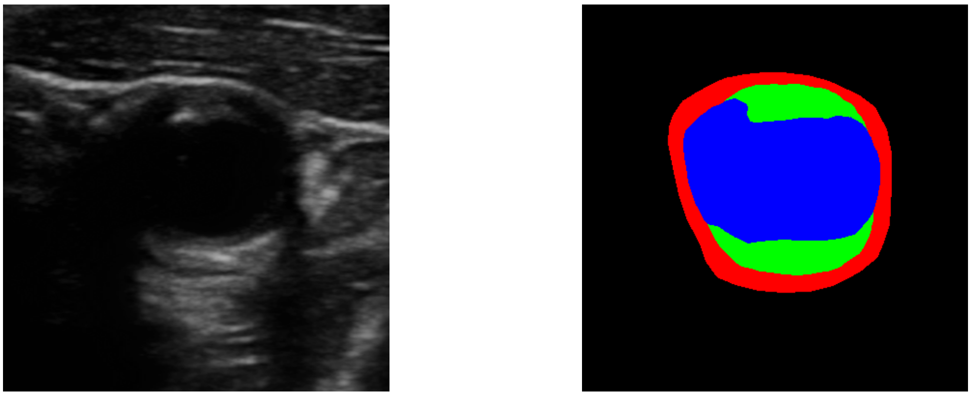
The image alone of the artery doesn't provide details on the extent of the stenosis and the characteristics of the plaque. To understand this, the carotid image is segmented into four categories namely background tissue, lumen, artery wall and the plaque.

4.1 Dataset

4.1.1 Fully Annotated dataset

Although the ANTIQUE dataset is fairly large, the Fully Annotated Dataset (FA) consists of only 150 transverse and 150 longitudinal images. This is due to the fact that the hand annotation of segments is a very difficult and time-consuming process requiring strong medical expertise from highly trained professionals.

As we can see on 4.1b the images of carotids are segmented into 4 segments: Background is black, artery wall is red, atherosclerotic plaque is green and lumen is blue.



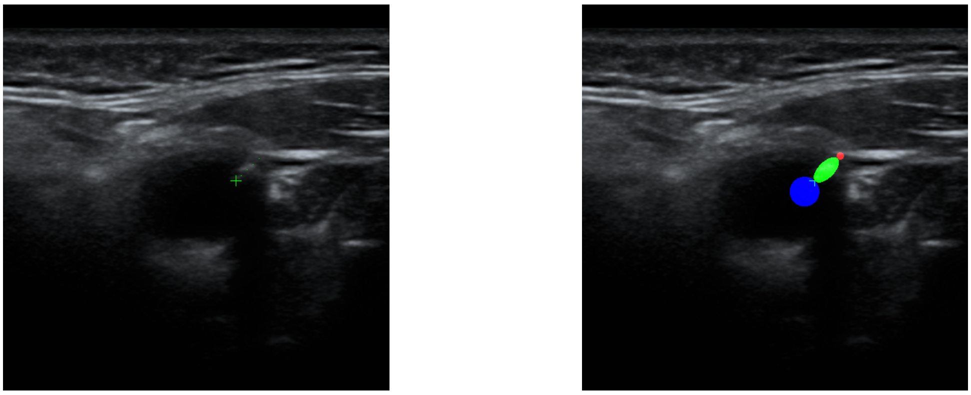
(a) : Ultrasound image

(b) : Segmentation label

Figure 4.1: Ultrasound images with paired segmentation annotation

4.1.2 Weakly annotated dataset

The ANTIQUE dataset contains a subset of images on which the plaque width was measured by the professionals. They include a green mark with dots denoting the width of the artery plaque 4.2a. Using these images weak annotations were created assuming that the area around the line is plaque, the area in front of the beginning of line segment is lumen, and the area behind the end of the line segment is wall. We can see this weak annotation in 4.2b.



(a) : Ultrasound image with plaque width annotation

(b) : Image fused with the created weak annotation

Figure 4.2: Ultrasound images with measured line of the plaque width and image fused with the created weak annotation

In total we have 1393 transversal weak annotations available, paired with the images. In order to not add additional information the green line segment annotation is removed by performing image inpainting.

4.1.3 Preprocessing

We define two image preprocessing options for the fully annotated dataset.

First is the one used by M. Kostelanský [25] [26] we use it in order to be able to compare our models to his results and to try more data preparation options. Firstly the image and segmentations are cropped by the minimal bounding box of the label segmentation, and then the cropped segmentation is padded by 5 background pixels. This cropped image is than resized to size 256×256. The advantage of this approach is that the image does not include irrelevant areas. On the other hand, the negative is that we lose the information about the absolute size of the image, and the reshaping coefficients of the x and y axes are not the same. The underlying segmentation images varies in size and scale therefore we call the dataset and models using this preprocesing as Variable-Size.

The second option is to crop a fixed-size box around the segment, we call this method Fix-Size. After inspection of the ultrasound images, we select the size of the bounding box to 380×380, which fit most of the carotid arteries and is of minimal size. Advantage of this approach is that it maintains the scale. The con is that some of the carotids are small and the background class is prevalent in the images.

The weak annotation dataset is cropped to the fixed size of 400×400, and then, the center is cropped to the size of 380×380. Thus, it is of the same scale as the fully annotated Fix-Size Dataset and the datasets can be used together.

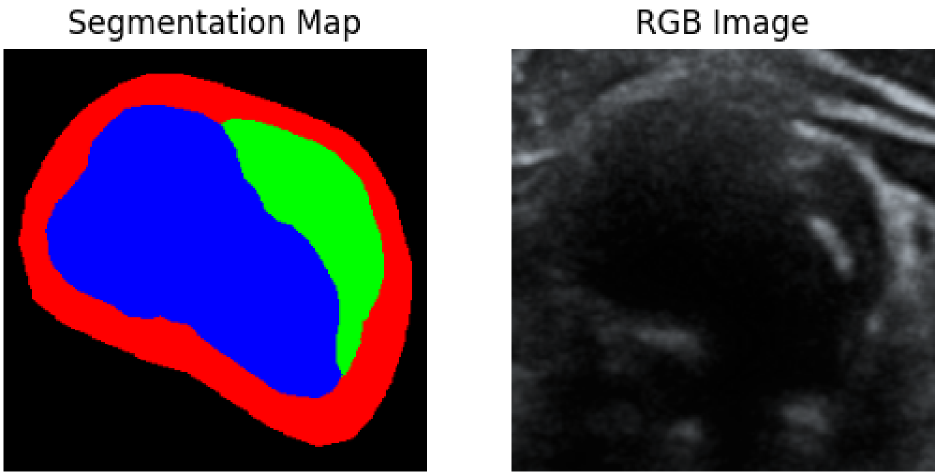


Figure 4.3: Segmentation and image produced by the Variable-Size preprocessing

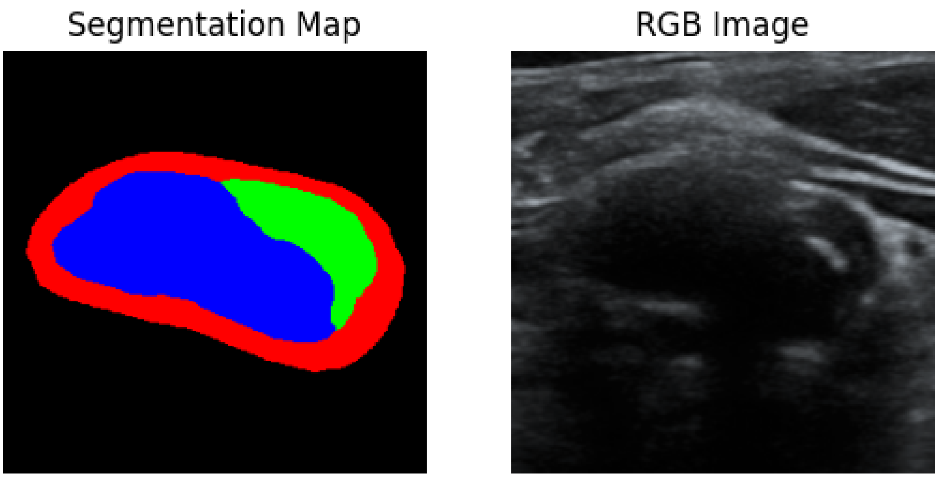


Figure 4.4: Segmentation and image produced by the Fix-Size preprocessing

augmentation	parameters
random rotation	p = 0.5 minimal angle = -10 maximal angle = 10
random contrast	p = 0.5
random brightness	p = 0.5
random horizontal flip	p = 0.5

Table 4.1: Augmentations used for training

4.1.4 Augmentation

Augmentation is a technique used to enhance the diversity and volume of training data, thereby improving the robustness and performance of the models. In medical imaging, it is especially useful because of the limited amount of data [14].

Applying to training set, we will augment the images by randomly adjusting the brightness and contrast, then perform random rotation and random horizontal flip of the image. Vertical flip will not be used as it is not invariant for ultrasound images of carotid arteries. We summarize the augmentation in Table 4.1.

4.2 Benchmark methods

In the segmentation part, we build on the master thesis of M. Kostelanský [26] and article [25]. M. Kostelanský uses U-Net with cross-entropy loss (U-

Net-CE) in the former case and U-Net with log-cosh dice loss and depth-wise convolutional filter size (U-Net-Dice) in the later case. Some notable diffences between his mnodels and the original U-Net are use of the PReLU activation layers and padding to maintain the dimension of the input. He also uses random dropout for the training.

We strictly follow his models implementation architecture details in both of the cases and use part of his code for the benchmark methods. The exception is the input size for U-Net-CE where he uses the size of 512x512, we unify it to the size of 256x256.We also preserve the train/val/test split from the master thesis.

4.3 Methods

Through the literature review, we selected from our knowledge, the most suitable methods for segmentation applicable to our task. The methods selected had been chosen by meeting following criteria. In the method proposal they are superior to classical supervised method, methods are universal and can be used for our ultrasound image domain, methods are sufficiently different from each other in order to apply a broad spectrum of methods for our segmentation problem.

The First method we selected is the method proposed in Anatomically Constrained Neural Networks (ACNN) [36]. In the proposal paper this method outperformed U-Net in the domain of 3D cardiac images. This method is vastly universal and proposes the general training strategy, which can be used with an arbitrary segmentation network. It gives us a free choice in the selection of segmentation and autoencoder architectures for our specific task.

The second method we selected is proposed in Multiscale Adversarial Attention Gates (MAAG) [49].This method outperformed U-Net on the ACDC dataset [1]. We had chose this method as it allows us to use it in a fully supervised setting and on top of that in a weakly supervised setting with partial annotation, so-called scribbles.

For the last method, we selected SeGAN [50]. SeGAN is one of the first GAN-inspired frameworks adapted specifically for the segmentation task that produces superior segmentation results compared to the fully supervised methods. In the paper, authors claim that It outperformed U-Net on several metrics in the BRATS 2015 dataset [32].

4.3.1 ACNN

Method overview

Motivation behind this method is the fact that the majority of existing segmentation methods are supervised solely through a local loss function at pixel level. To capture the global features they incorporate the shape prior to the loss function. For this task, they propose an autoencoder model to obtain a non-linear compact representation of the underlying anatomy. The autoencoder network serves as a reguralizer for the segmentation network. The overall framework proposed in the paper is depicted in figure 4.5

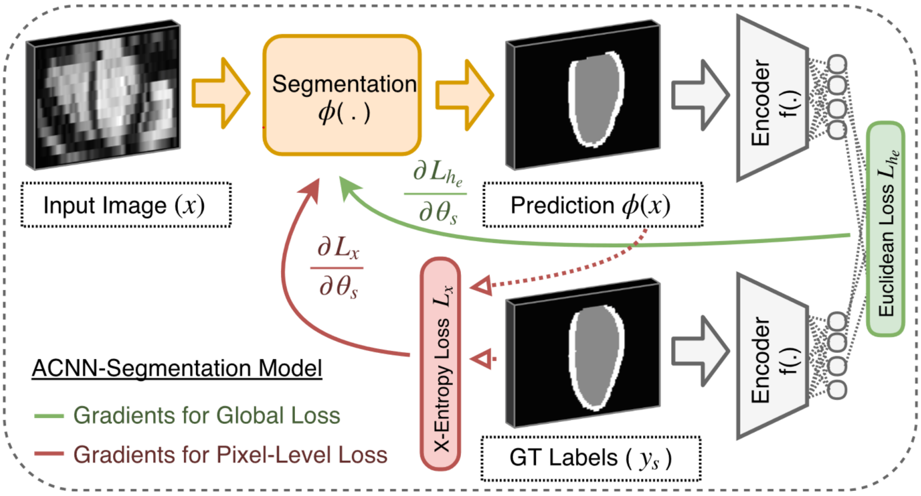

Input Image (x)
Segmentation $\phi(\,.\,)$
Prediction $\phi(x)$
Encoder f(.)
Euclidean Loss L_{h_e}
$\frac{\partial L_{h_e}}{\partial \theta_s}$
$\frac{\partial L_x}{\partial \theta_s}$
X-Entropy Loss L_x
GT Labels (y_s)
Encoder f(.)
ACNN-Segmentation Model
Gradients for Global Loss
Gradients for Pixel-Level Loss

Figure 4.5: Training scheme of the ACNN method [36]

Segmentor

The loss function for the segmentor ϕ consist of two losses. The first L_x is the supervised loss of the segmentor, for example, cross-entropy loss. The second loss L_{h_e} is a mean squared error (MSE) loss between the encoding of the label segmentation and the encoding of the created segmentation. θ_s represents the parameters of the segmentor, L_x is a loss function for the segmentor, y is the target output, and λ is a weighting parameter for the MSE loss. θ_f represents the parameters of the encoder part of the autoencoder. During the experiments we observed that the MSE loss is usually larger than the CE loss, We therefore used dynamical weighting of the MSE loss to maintain a fixed ratio between the amplitudes. Additionaly the MSE loss is weighted by hypermater k to set their contribution to the joint loss.

$$L_{h_e} = \|f(\phi(x);\theta_f) - f(y;\theta_f)\|_2^2$$

$$\min_{\theta_s} L_x(\phi(x;\theta_s), y) + \lambda \cdot L_{h_e}$$

$$\lambda = k \cdot \frac{\|L_x\|}{\|L_{h_e}\|}$$

As a segmentor network we used the U-Net-CE architecture from the benchmark method.

Autoencoder

For training the autoencoder we have tried two approaches. First method is to train the autoencoder in advance. In this approach we trained the autoencoder on the segmentation labels y using the cross-entropy loss function L_{ce}. Henceforth we took the trained encoder and used it in the segmentation model training. We optimized the encoder parameters θ_f and decoder parameters θ_g according to:

$$\min_{\theta_f,\theta_g} L_{ce}(y, g(f(y;\theta_f);\theta_g))$$

The Second approach is to train the autoencoder and segmentation network simultaneously. At each traning step autoencoder is updated first and then the segmentor is updated second. This approach gives us the option to use the created segmentation $\hat{y}$ as a input to the autoencoder instead of the labels. We optimized the autoencoder according to:

$$\min_{\theta_f,\theta_g} L_{ce}(y, g(f(\hat{y};\theta_f);\theta_g))$$

Autoencoder architectures

For the autoencoder network we constructed three different autoencoder architectures. For the notation, we define C as a set of classes, and we define encoding vector as $\mathbf{v}$.

First autoencoder architecture is taken from the [17], where autoencoder serves as an algorithm for clustering. It consist of 3 convolutional and one linear layer in the encoder and linear layer and 3 transposed convolurional layers in the decoder. Each layer is followed by the ReLu activation function.

Encoder 1
Input: 256 x 256 x $\vert C\vert$
conv5-32, stride = 2
conv5-64, stride = 2
conv3-128, stride = 2
FC-$\vert\mathbf{v}\vert$

Decoder 1
FC-131072
Reshape 32×32×28
transpose-conv3-64, stride = 2
transpose-conv5-32, stride = 2
transpose-conv5-$\vert C\vert$, stride = 2

The second autoencoder architecture is inspired by the AlexNet [27]. The encoder strictly copies architecture of the AlexNet classification network, only the input and output layers are changed. The decoder has an inverted order of encoder layers. Convolutional layers are replaced by transpose convolutional layers and maxpooling layers are replaced by nearest neighbor upsampling layers.

The third autoencoder is inspired by an autoncoder architecture used for denoising medical images [15]. The encoder consist of 4 convolutional layers each followed by ReLu activation function and Maxpooling layer. The last layer is a fully-connected layer. In decoder convolutional layers are replaced by transpose covolutional layers, and max-pooling is replaced by nearest-neighbor upsampling.

encoder 3
Input: 256 x 256 x $\vert C\vert$
conv5-64
max-pool2, stride = 2
conv3-128
max-pool2, stride = 2
conv3-256
max-pool2, stride = 2
FC-$\vert\mathbf{v}\vert$

decoder 3
FC-65536
Reshape 16×6×256
transpose-conv3-256
Upsample 2x
transpose-conv3-128
Upsample 2x
transpose-conv3-64
Upsample 2x
transpose-conv5-$\vert C\vert$
Upsample 2x

4.3.2 MAAG

Method overview

The authors created the MAAG framework primarily for weak scribble annotations which is common setting for other segmentation methods [29], however in the paper they have also showed that this framework can be used with mixed annotations by combining scribbles with fully annotated labels, which improved the results. We used this method in fully supervised settings with segmentation labels paired to the images, and aditionally we try mixed-supervision setting with available weak annotations.

The framework is trained as an adversarial game. It consists of segmentation network and discriminator network. The discriminator is trained to learn shape prior, and segmentor is trained to satisfy it. From images the segmentor produces multi-scale output segmentation, which is then fed to the multi-scale input discriminator. In the adversarial game the segmentor is trained to produce realistic-looking segmentations and discriminator is trained to correctly distinguish real annotations from the created segmentations. The overall architecture is depicted in the image 4.6 taken from the method proposal.

CNN architectures

For a segmentor we used the U-Net-CE architecture from the benchmark method and modified it by adding attention gates to the decoder blocks. The encoder part of the U-Net remained unchanged. In the decoder blocks, we added adversarial attention gates following the description in the method proposal. Each decoder block consist of two 3×3 convolutional layers producing feautre map $M^{(d)}$ followed by a 1×1 convolutional layer creating soft

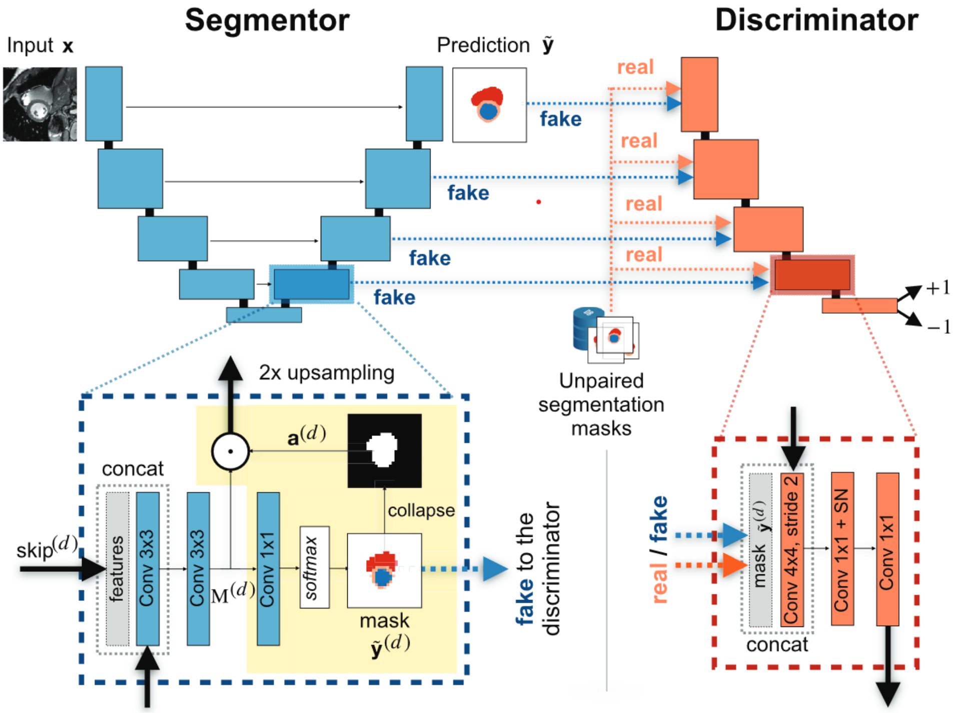


Figure 4.6: Multi-scale segmentor/discriminator architecture from the MAAG method[49]

segmentation that is used as input for the discriminator on the same depth level and is also used as an attention-map which is then multiplied with the $M^{(d)}$ using Hadamard product.

For a discriminator we strictly follow their implementation. Discriminator consist of 4 blocks that receieves input from upper block and attention from the segmentor on the same level. These two inputs are concatenated, then downsampled through a convolutional layer with a 4×4 filter size, normalized using spectral normalization [33] and then compressed using a 13 x 1 x 1 convolution network. Both convolutional layers use the *tanh* activation function. To make the adversarial game harder for the discriminator we flip the discriminator label with probability of 10% and also we apply the Gaussian noise with zero mean and standard deviation of 0.2.

Training

We adjusted the training process to fit our segmentation task. The authors worked with paired scribble annotations with images and unpaired segmentation masks and unpaired ultrasound images. The authors point out that the segmentation masks can be obtained by different medical imaging methods, therefore creatin new set of unpaired images where only the segmentation can be used. We have paired images with fully annotated segmentations as well as set of different images paired with the weak segmentation labels.

For the paired images, we used supervised loss L_{SUP} namely weighted cross-entropy loss. Class weights are scaled based on the proportion of given class.

$$L_{\text{sup}} = \left[-\sum_{i=1}^{c} w_i \cdot y_{si} \cdot \log(\hat{y}_i) \right],$$

For the paired weak annotations, we used weak supervised loss L_{weak} which is the same as the L_{SUP} but computed only on the annotated pixels. Unannotated pixels do not contribute to the loss.

In the adversarial game we minimize discriminator loss $V_{LS}(\Delta)$ and generator loss $V_{LS}(\Sigma)$ according to the Least Square GAN objective [31]

$$V_{LS}(\Delta) = \frac{1}{2}\mathbb{E}_{y\sim\mathbf{y}}\left[(\Delta(y) - 1)^2\right] + \frac{1}{2}\mathbb{E}_{x\sim\chi}\left[(\Delta(\Sigma(x)) + 1)^2\right]$$

$$V_{LS}(\Sigma) = \frac{1}{2}\mathbb{E}_{x\sim\chi}\left[(\Delta(\Sigma(x)) - 1)^2\right].$$

We train the models by iterating multiple steps. First we optimize segmentor over batch of images paired with segmentations, we minimize $L = L_{\text{sup}} + a_1 \frac{\|L_{\text{sup}}\|}{\|V_{LS}(\Sigma)\|} V_{LS}(\Sigma)$. The a_1 is a hyperparameter that makes the contribution of adversarial loss smaller as the adversarial loss only determine the plausability of the shape. Secondly, we optimize the discriminator by minimizing the discriminator loss $a_2 V_{LS}(\Delta)$. If we employ weak labels in the training, then the third step is to optimize segmentor by minimizinng the weak supervised loss $a_4 L_{weak}$.

4.3.3 SeGAN

Method overview

The authors propose SeGAN as a segmentation method inspired by the GAN framework. In SeGAN segmentation network and critic network plays min-max game where the segmentor is trying to minimize the loss and the critic is trying to maximize the loss. In this training setting the segmentor and the critic can be efficiently trained and SegAN can learn pixel dependencies on multiple scales and thus learn local and global features.

SeGAN is composed of a segmentor which is a U-Net style encoder/decoder segmentation network. We followed the authors implementation and used a network with 8 residual encoder blocks and 8 residual decoder blocks. Each residual block consists of three convolutional layers each followed by batch normalization and LeakyRelu layer.

The Critic has similar structure to the proposed segmentor decoder part. The critic consists of multiple blocks. Each block produces output which is then used as an input for the next block and is furthermore used directly in the loss function. For the critic network we follow the authors implementation of the individual blocks. Each nlock consist of convolutional layer, batch normalization and leaky ReLu layer. We use critic architecture with 6 of these blocks.

We choose the setting with one segmentor and a separate discriminator for each class. This setting provides better results than setting with only one discriminator and yields a comparable result to a setting with segmentor for each class.

Training is performed in an alternating fashion where first the critic is fixed and segmentor is trained and then the segmentor is fixed and the critic is trained. We describe process a computing loss, firstly the segmentor S produce a multi-class soft prediction map. Then for each class the image is masked by the produced soft mask and is then fed to the critic C producing hierarchical feature tensor $f_C(x)$. Afterwards the image is masked by the true label map and fed to the critic which produces a hierarchical feature tensor for label $f_C(x')$. The loss l_{mae} is a mean absolute error (MAE) between the features defined as:

$$l_{mae}(f_c(x), f_c(x')) = \frac{1}{L}\sum_{i=1}^{L}\left\| f_c^i(x) - f_c^i(x')\right\|_1$$

Where L is the number of critic layers, and $f_c^i(x)$ is the extracted feature map of image x at the ith layer of C. While training the segmentor the loss is minimized. In contrast when training discriminator the loss is then maximized.

For the training of the segmentor we also explored the possibility of using joint loss for the training. We try to add the supervised loss $Lsup$ computed directly from the segmentor output for example Dice-loss or cross-entropy loss into the segmentor loss $L = a_1 Lsup + Lmae$.

4.4 Experiments

In the experiments we train and evaluate models on the Fully Annotated dataset (FA) of transverse images. In experiments we further divide this dataset into variable size (FA-Variable-Size) and fixed size (FA-Fix-Size) datasets according to section 4.1.3. For both options 75 images are used for training, 25 for validation and 50 for testing. First we will focus on each method experiments and then we present overall results.

For the evaluation we compute Intersection over Union (IoU) for all classes and also mean IoU over classes (mIoU), also Dice Score over all classes and mean Dice Score. And we compute the overall pixel accuracy of the segmentations.

4.4.1 ACNN

In ACNN we train both the autoencoder and segmentor using ADAM optimizer [23] with beta1 = 0.5 and beta2 = 0.5, batch size = 2 and learning rate equal to $10e^{-4}$ for the segmentor we also use weight decay equal to $10e^{-3}$. We use scheduler that reduce the lr on plateau with patience 25 epochs. And we train for 300 epochs.

In the experiments we determine best autoencoder, optimal encoding size for the autoencoder and decide whether is it better to train autoencoder before the segmentor or simultaneously.

In first experiment we try different autoencoder architectures amd determine the best. For this experiment we use FA-fix dataset and encoding size equal to 256 and we train the autoencoder before segmentor. According to experiment best performing autoencoder is the deep clustering autoencder ($mIoU = 0.681$) surpasing denoising autoencoder ($mIoU = 0.668$) and the alexnet inspired autoencoder ($accuracy = 0.666$).

For the encoding vector size we tried sizes of 32, 64,128,256 and 512. Optimal encoding vector size for our use case is 256 4.2, authors of ACNN [36] use size of 64, we think that we need bigger encoding size as the segmentation maps of carotids artery varies more than the segmentation maps of abdomen with fixed position of organs and more rigid overall structure.

Encoding Size	mIoU
32	0.669
64	0.672
128	0.675
256	0.681
512	0.670

Table 4.2: Table of encoding sizes and corresponding mIoU values

Autoencoder

We inspect whether the autoencoder is able to learn the lower-dimensional representation on the space of segmentation labels. In figure 4.7 we plot input and output of the denoise clustering autoencoder after training separately the autoencoder. We can see that the autoencoder is able to learn the lower-dimensional representation.

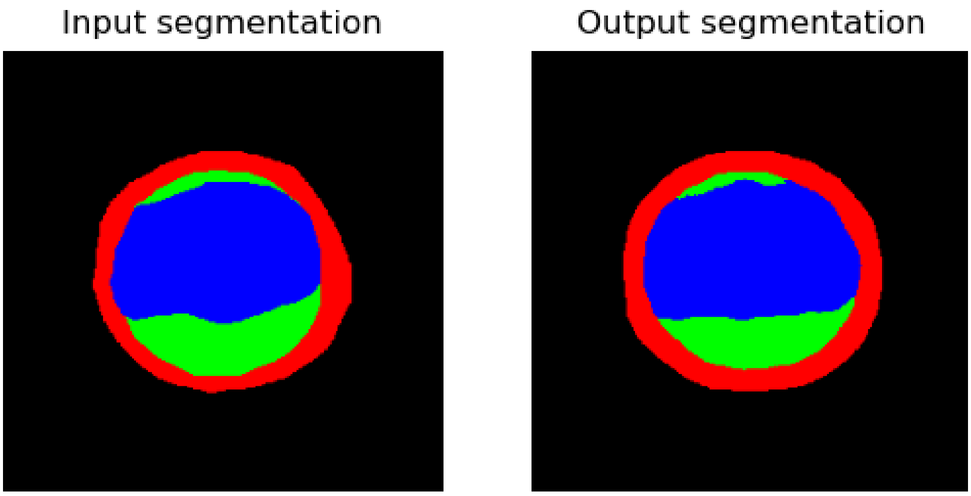
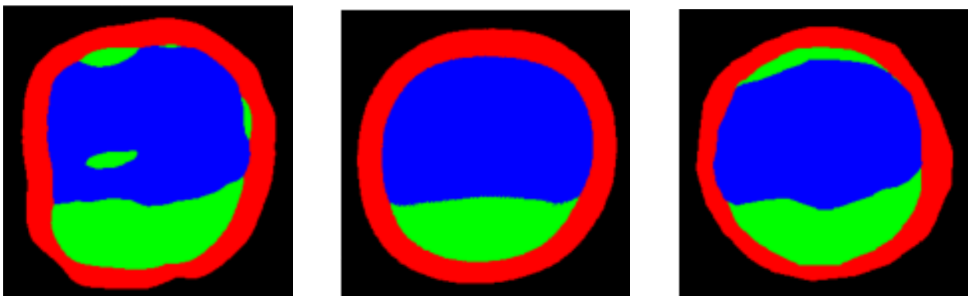


Figure 4.7: Input and output of trained autoencoder

We also plot how the autoencoder works during the training of segmentor. On figure 4.8 we show the segmentation output, output of the autoencoder performing the encoding on the segmentator output and lastly the label. We can see that the autoencoder normalizes the segmentation and output more plausible result. We observe that these results are more plausible but also loose some information about location, thus is important to overweight the MSE loss between encodings in the joint segmentor loss. During the experiments we observed that making the MSE loss one magnitude smaller which is specified in the method proposal works well.



(a) : AE Input **(b) :** AE ouput **(c) :** True segmentation

Figure 4.8: Created segmentation as input to autoencoder, output of autoencoder and true label

4.4.2 MAAG

In MAAG we train both the segmentor/generator and discriminator using ADAM optimizer [23] with beta1 = 0.5 and beta2 = 0.5, batch size = 2 and learning rate equal to $10e^{-4}$ for the segmentor we also use weight decay equal to $10e^{-3}$. In the choice of scheduler we follow method proposal and use cyclic learning rate [46].

We observe that with loss weights coefficients set as in the method proposal the segmentation outputs are very similar to each other, thus the shape constraint is too large. This result is expected as our setting uses full annotations which includes shape information in contrary to the scriblle setting with very little shape information. To surpress the shape prior we set the a_3 coefficient to be equal to 0.1 giving best.

Now we evaluate the addition of weak annotations into the framework, we take two identical settings and compare the results as we can see on table 4.3 weak segmentations slightly improve the results.

metric	with WA	no WA
mIoU	0.68	0.66
accuracy	0.93	0.91

Table 4.3: Test results on MAAG with and without addition of weak annotations

4.4.3 SeGAN

For SeGAN we innitialy tried to use the same optimizer settings as in the previous two cases, but we were unable to train the model, The critic was too succesful at maximizing the loss and the loss after few epochs exploded. We manage to train the model using RMSprop optimizer with learning rate = 0.00002 and crucially with bigger batch size of 8.

First hyperparameter we tune is the weight of the additional dice loss that is joint for the final segmentor loss with the L1 loss. We select the weight a_1 to 0.5. For showing that the whole adversarial setting results in improvement, we also try using only the supervised loss effectively removing the whole critic part. The mIoU of 0.649 is worse than the setting including critic with mIoU of 0.681

In the SeGAN paper authors show that the Segmentor/Critic architecture with $L1$ loss can be trained, in the proof the assume that the weights are clamped. We try different clamping ranges, of the discriminators weights, in experiments we observe that with some combinations of batch size and clamping range one of the discriminator (ussually for lumen class) exploded and the produced loss was extremely high. We find that restricting the discriminator weights between -0.05 and 0.05 results in relatively stable training.

4.4.4 Results

In the tables 4.4 4.5 we can see results for all the methods namely U-Net with cross-entropy loss function U-Net used in paper [25], ACNN segmentor autoencoder framework, MAAG segmentor discriminator framework and SeGAN segmentor/critic network.

metric	U-Net CE	U-Net-Dice	ACNN	MAAG	SeGAN
mIoU	0.683	0.700	**0.719**	0.714	0.681
IoU - background	0.861	0.853	**0.862**	0.858	0.807
IoU - wall	0.570	0.573	**0.604**	0.602	0.528
IoU - plaque	0.447	0.519	**0.549**	0.535	0.529
IoU - lumen	0.851	0.859	0.859	**0.860**	0.858
accuracy	0.867	0.868	0.875	**0.914**	0.848

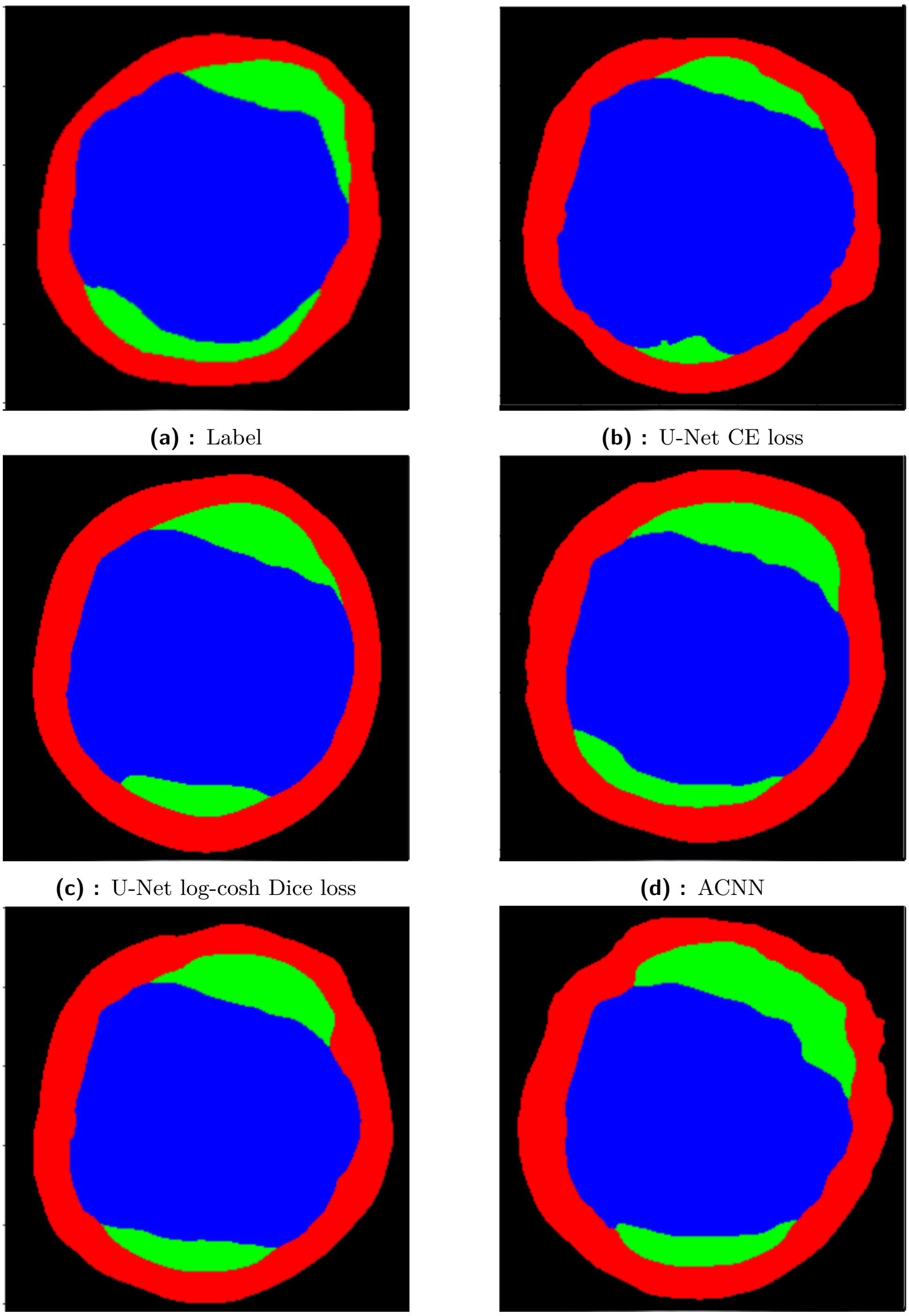
Table 4.4: Test results for all the methods on the Fully Annotated Variable-Size dataset

metric	U-Net-CE	U-Net-Dice	ACNN	MAAG	SeGAN
mIoU	0.641	0.665	**0.681**	**0.681**	0.676
IoU - background	0.953	0.955	0.952	0.947	**0.957**
IoU - wall	0.449	0.477	**0.500**	0.492	0.471
IoU - plaque	0.357	0.415	0.461	**0.482**	0.441
IoU - lumen	0.806	0.816	0.812	0.804	**0.834**
accuracy	0.930	0.935	0.936	0.930	**0.937**

Table 4.5: Test results for all the methods on the Fully Annotated Variable-Size dataset

When presenting results we have to state that we do not do much of hyperparameter tuning for the U-Net benchmark methods. We took the existing implementation by M. Kostelansky who used it on slightly different datasets. We can see that for the FA-Variable-Size the ACNN achieves best mIoU and best IoU for almost all classes, best accuracy is achieved by the MAAG. SeGAN does not achieves such results, as we can see on 4.9f the SegGAN produces least homonogeous segmentation this applies to the majority of the test images. We provide all the test prediction visualizations in the supplementary materials.

For the FA-fix dataset the ACNN and MAAG shows highest mIoU and SeGAN shows highest accuracy. For this dataset the accuracy is slightly missleading as the background class is very prevalent thus the IoU is a more suitable metric. When inspecting the results the segmentation algorithms are able to correctly segment images where the carotid is easily detectable 4.10 but at a small amount of cases the algorithm mistakes the surroundings for the carotid segments 4.11.



(a) : Label

(b) : U-Net CE loss

(c) : U-Net log-cosh Dice loss

(d) : ACNN

(e) : MAAG

(f) : SeGAN

Figure 4.9: Produced segmentations by all the methods on the Variable-Size dataset

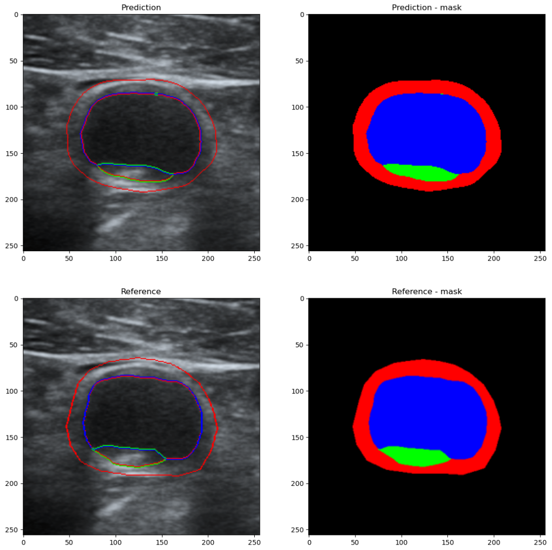


Figure 4.10: ACNN Segmentation with high accuracy

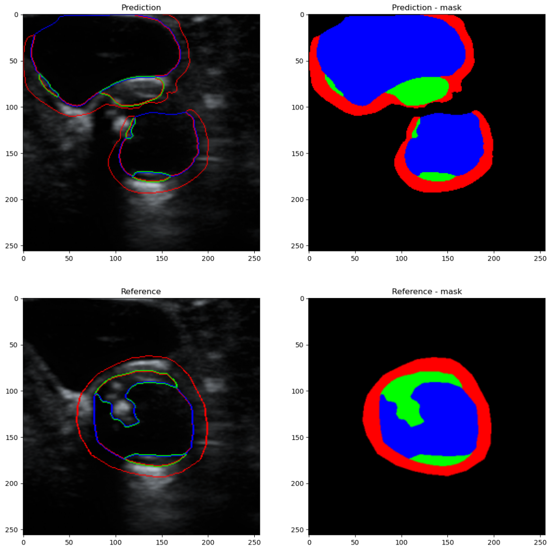


Figure 4.11: ACNN Segmentation with low accuracy

Chapter 5

Parameter Estimation

In the second task we try to estimate the parameters of the carotid plaques. As described in the section 2.2 we have available expert annotaions for geometric and non-geometric parameters. We use the created segmentations alongside the ultrasound images to predict these parameters. We use classification CNNs for this task, some of the parameters we also try to directly compute from the segmentations.

5.1 Dataset

From the ANTIQUE dataset we use the subset of images on which were performed the manual measurement of the plaque, for those images we have available the expert evaluation of multiple parameters. These images are recognizable by the green measurement marks which were produced during the measuring. We name this dataset of 1316 images as Key-Image Dataset. In the following section we describe the whole preprocessing pipeline to obtain suitable image segmentation pairs for estimating the parameters using CNN models and direct computation.

5.1.1 Preprocessing

The images from the Key-Image Dataset are containing information about the whole ultrasound scan not only the carotid and close surroundings. First step is to localize the carotid. We use the Faster R-CNN [39] model which was trained by M. Kostelansky to localize the carotid [25]. This localization net returns the bounding box and the carotid probability score. We discard all the images where the R-CNN outputs class probability score of carotid lesser then 0.9.

Processing pipeline continue with the segmentation of the localized images. In the previous chapter we trained the segmentation networks on the Fixed-Size dataset and the Variable-Size dataset. We utilize models trained on both of the datasets. For the model trained on Variable-Size dataset we crop the image directly using the bounding box obtained by the localization network. For the model trained on the Fix-Size dataset we pad or crop the bounding box in order to be of size 380×380. The final processed Key-Image dataset of ultrasound images paired with the created segmentations consists of 1528 images.

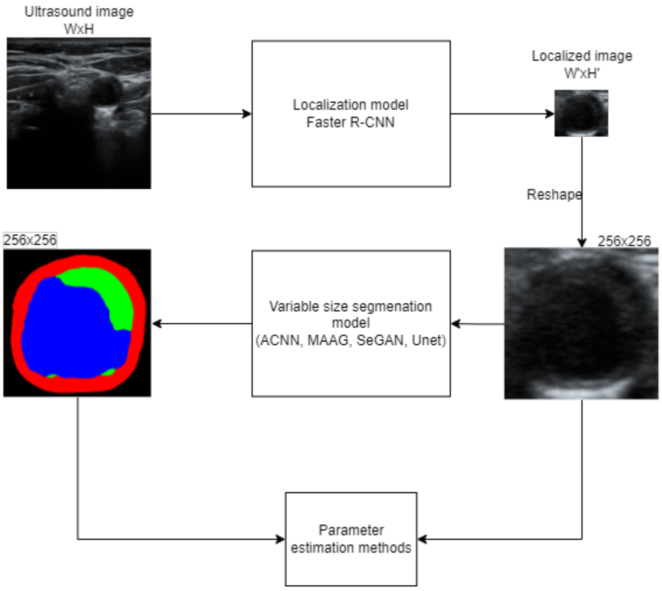


Figure 5.1: Preprocessing pipeline for Variable-Size segmentation models

5.2 Methods

We divide the parameters into two categories. First are the non-geometric parameters including homogenecity and echogenicity. These parameters are categorical. homogeneity has two possible characterizations, homogeneous and heterogeneous. Echogenicity is categorized into four classes.

Second group are the geometric parameters including only the plaque-width. Plaque-width is defined by the scalar value defining the width of the plaque segment.

5.2.1 CNN models for parameter estimation

We use a CNN models to estimate the parameters from the images with segmentations. We use ResNet as a backbone network, for categorical variables we use cross-entropy loss and for the scalar values we use Mean-Squared loss.

As a CNN classification network backbone we use ResNet18, we find out that this network is sufficiently complex and larger versions of ReNet are too complex for our task, leads to overfitting and the additional computational cost is not justified by the better performance.

ResNet and other classification methods usually take only single input. We have available images and segmentations, we try different approaches to use both images and segmentations to predict the parameters, we construct different versions of the network to combine the inputs.

Input merge model

First option is to combine the images with segmentations on the input layer of the ResNet network. This is the ussual architecture for supervised classification networks. In the ResNet architecture we only change the input layer channels. To combine the pairs we try these combinations:

- Plaque Mask - Mask ultrasound mage by the plaque segment. The resulting input tensor dimension is (height, width, 1)
- Concatenation - Concatenate the image with the 4-class segmentation. The resulting The resulting input tensor dimension is (height, width, 4)

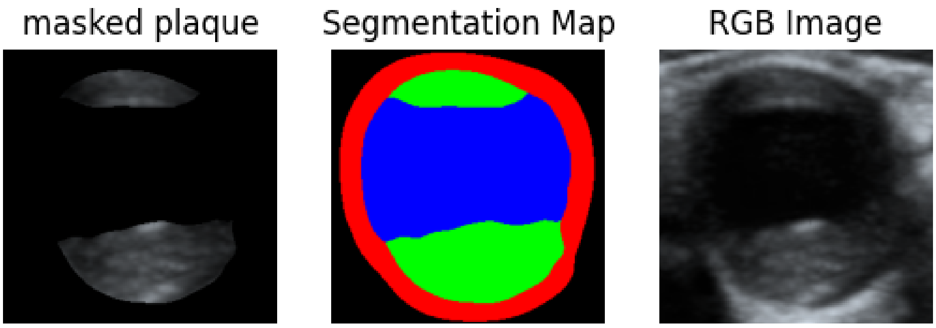


Figure 5.2: Masked plaque image

Middle layer merge model

Second method is to combine the segmentation and image at a deeper layer. This model is inspired by the [34] and [13]. The multiple inputs are processed separately in its own pipes, at defined inner layer they are merged and further processed combined.

In the used ResNet18 model we choose to process the images and segmentations separately first five layers, in the sixth layer we concatenate both tensors over the channel dimension and the concatenated tensor is processed by the remaining layers. We also try to process the image separately and in the sixth layer to mask it by separately processed plaque-mask from the other. Subsequently this masked tensor is processed by the rest of the network.

5.3 Experiments

5.3.1 Non-geometric parameters

For the non-geometric parameters we estimate echogenicity and homogeneity using the classification CNN models. We use the Adam optimizer with learning rate $5e{-}6$ weight decay $5e{-}5$ and batch size 2. Every ten epochs the learning rate is divided by 10. We train for 30 epochs. We use the cross-entropy with weights that proportionally give more weights to the underrepresented classes. In the dataset the homogeneity has classes of cardinality 442 and 1004 and the echogenicity of 283, 394, 477, 292. We also try to undersample the dataset but that approach gave inferior results.

For the experiments we take 80% images for training, 10% for validating and 10% for testing. We use the same augmentations as for the segmentation task 4.1.

Determine best segmentation option

In the first experiment we determine which set of the created segmentations, results in best classification. The more accurate the segmentations the more accurate should be the classification because the classification model recieves correct information which part is plaque whose parameters are estimated. For this experiment we use the architecture which concatenate the image and segmentation at the start of the model. The results for this experiment are shown at table 5.1 as we can see the model was unable to achieve satisfying result for two-class homogeneity classification and for four-class echogenicity classification.

Segmentation model	Homogenecity accuracy	echogenicity accuracy
Fix-Size		
U-Net	0.5	0.31
ACNN	0.63	0.31
MAAG	**0.63**	**0.36**
SeGAN	0.61	0.29
Variable-Size		
U-Net	0.57	0.24
ACNN	0.57	0.36
MAAG	0.56	0.31
SeGAN	0.56	0.36

Table 5.1: Classification of non-geometric parameters using Input Merge model, comparing different segmentation models.

Determine the best model option

In the second experiment we evaluate different multi-input models we have constructed four varying options:

- Start Concatenate - Input Merge model concatenating segmentation and image at the start of the network
- Start Mask - Input Merge model masking image by plaque segment at the start of the network
- Middle Concatenate - Middle Merge model processing separately image and segmentation and concatenating their feature maps in the middle of the network
- Middle Mask - Middle Merge model processing separately image and plaque mask and multiplying their feature maps in the middle of the network

We run this experiment using the segmentations created by the Fix-Size ACNN, which achieved best result in the previous experiment. In the table 5.2 we can see that the Start Concatenate model gives best but unsatisfactory results.

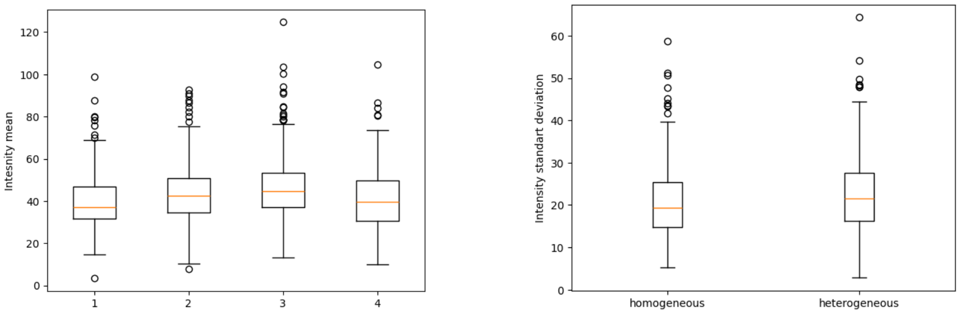
Result discussion

The problem with the low accuracies for homogeneity and echogenicity is most likely not caused by the classification model but by the inacurate

Segmentation model	Homogenecity accuracy	echogenicity accuracy
Start Concatenate	**0.63**	**0.36**
Start Mask	0.49	0.29
Middle Concatenate	0.52	0.31
Middle Mask	0.56	0.34

Table 5.2: Results for classification of non-geometric parameters using different input pair combinations

segmentations of the plaque. The echogenicity is ability to bounce an echo so the pixels defining the plaque should be brighter for higher echogenicity class, as we can see from table 5.3a it is just partially true for the plaque segments proposed by the Fix-Size ACNN segmentor, the highest echogenicity class images do not have the highest average intensity of the plaque segment. In the homogeneity, heterogeneous class (non-homogeneous) refers to a structure with dissimilar components or elements [1], and homogeneous to the opposite. In theory the heterogenous plaques should have higher standard deviation of the intensity values. We plot the standard deviation of intensity values on figure 5.3b we see that the intensity standard deviations are not much different.



(a) : Boxplot of mean intensity values for plaque pixels over the echogenicity classes

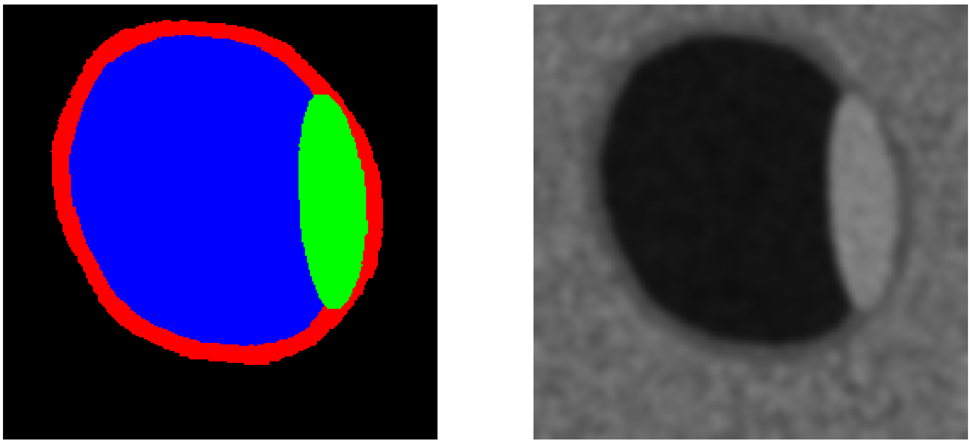
(b) : Boxplot of standard deviation intensity values for plaque pixels over the homogeneity classes

Experiments on synthetic dataset

In order to verify that the models are able to learn the features, we constuct a synthetic dataset. In the synthetic dataset we can be sure that there is no error and disturbance in the label annotaions thus the segmentation matches the actual segment. On the synthetic images we evaluate echogenicity which in our opionion can be modeled better by synthetic dataset. We model the different echogenicity classes with different means for pixel intensity generation of plaque segment.

[1] https://radiopaedia.org/articles/heterogeneous-vs-heterogenous

We create synthetic images following these steps. First step is to create and underlying segmentation. Random ellipse is created. This random ellipse represents the lumen and wall. Then the second ellipse within the first ellipse is created representing the plaque. Then the whole image is randomly rotated. Lastly, elastic transformation and grid distortion are performed. Next step is to create the image from the segmentation. To create random images from the created segmentation we sample the intensity of pixels from a normal distribution. Pixels from each segment are sampled from a distribution with a different mean and standard deviation. The last step is to apply Gaussian filter to smooth the image. For the different echogenicity classes we sample the plaque intensities with different means of the normal distributions. We show the created synthetic images and segmentation on figure 5.4



(a) : Synthetic segmentaion

(b) : Synthetic image

Figure 5.4: Created synthetic segmentation and image

We crated in total 1000 random synthetic images, 250 for each class. We present the synthetic echogenicity results on table

Segmentation model	Echogenecity accuracy
Start Concatenate	0.86
Start Mask	0.93
Middle Concatenate	0.19
Middle Mask	0.46

Table 5.3: Results for classification of synthetic echogenicity using different models

5.3.2 Geometric parameters

We can directly compute the plaque width from the segmentation we formulate the plaque width as the maximal distance over all the minimal distances from each green pixel to any blue pixel, defined as:

$$\max_{g_i \in G} \left(\min_{b_j \in B} d(g_i, b_j) \right)$$

We compute it more efficiently using the KD-tree [4]. The created line denoting the width can be seen on 5.5. The computed width is in the pixels but the annotation is in milimiters. We use approximation that 1cm is equal to 172pixels based on the scale present in the raw images.

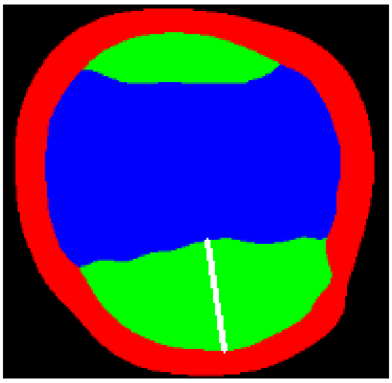


Figure 5.5: Width of the plaque denoted by white line

We try both the Fix-Size and Variable-Size datasets to compute the width, the Fix-Size has advantage that the resulting distance is indeed the factual distance in pixels. The Variable-Size images are not squares and are not of uniform size, thus the distance have to be scaled according to the difference between the image size and the segmentor output size. The advantage and reason why we use the Variable-Size datasets, is that it gives has superior IoU for the plaque class 4.4 observed during the segmentation training and testing process

We evaluate segmentations created by all the methods specified in the segmentation chapter. We use the mean absolute error (MAE) and the Pearson correlation coefficient. We show the results at table 5.4 for Variable-Size models and 5.5 for Fix-Size models

metric	U-Net	ACNN	MAAG	SeGAN
Pearson	0.226	0.279	0.247	0.200
MAE	1.6	1.58	1.66	1.52

Table 5.4: Results for direct computation of the plaque width with Variable-Size setting

metric	U-Net	ACNN	MAAG	SeGAN
Pearson	0.277	0.154	**0.461**	0.03
MAE	1.2	1.6	**0.9**	1.2

Table 5.5: Results for direct computation of the plaque width with Fix-Size setting

Chapter 6

Conclusion

The objective of this thesis was to improve the segmentation of the atherosclerotic plaques and evaluate their clinically relevant parameters based on the improved segmentation maps. During the literature review, we focused on the deep learning methods that use the anatomical prior to improve the segmentation results, mainly on autoencoders (AE) and generative adversarial networks (GAN). We have selected three suitable methods for our segmentation problem. The first was ACNN [36], which adds an autoencoder regularization network on top of the classical segmentation network. The second was MAAG [49], which is a GAN-inspired method for scribble annotations, which we used in a fully supervised setting. Lastly, SeGAN [50] was chosen as the third method. We implemented those methods, adjusting them for our setting. On the available dataset of 150 transversal ultrasound images, we have achieved a maximum mean IoU of 71.9% using the MAAG method with a U-Net segmentor and a convolutional autoencoder. The MAAG framework, using and adjusted U-Net segmentor, achieved the highest pixel-wise accuracy of 91.4%. In our settings, U-Net achieved a mean IoU of 70% and an accuracy of 86.8%. We have used all the models to produce segmentations on the key images, on which we have evaluated parameters by the medical professionals. We used the created segmentations alongside the images to estimate the parameters. Although the segmentation was slightly improved, it was not enough to reliably automatically estimate parameters. We tried to estimate non-geometric parameters of homogeneity and echogenicity using a classification network. Furthermore, we implemented several versions of the ResNet18 network, combining the input images with segmentations. The best result was achieved on a segmentation created by MAAG with an accuracy of 63% for homogeneity and 36% for echogenicity. The width of the plaque was directly computed from the created segmentations, and notably, the highest Pearson correlation coefficient of 0.46 was attained from the segmentation generated by the MAAG network. Although we slightly improved the segmentation results, they are still too inaccurate to be used for a reliable evaluation of the parameters.

Bibliography

[1] Papers with Code - ACDC Dataset.

[2] Wenjia Bai, Wenzhe Shi, Declan P. O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S. Peters, and Daniel Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. *IEEE transactions on medical imaging*, 32(7):1302–1315, July 2013.

[3] Peter Bandi, Rob Van De Loo, Milad Intezar, Daan Geijs, Francesco Ciompi, Bram Van Ginneken, Jeroen Van Der Laak, and Geert Litjens. Comparison of different methods for tissue segmentation in histopathological whole-slide images. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 591–595, Melbourne, Australia, April 2017. IEEE.

[4] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, September 1975.

[5] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538, July 2023. Number: 7970 Publisher: Nature Publishing Group.

[6] Mateusz Buczyński, Marcin Chlebus, Katarzyna Kopczewska, and Marcin Zajenkowski. Financial Time Series Models—Comprehensive Review of Deep Learning Approaches and Practical Recommendations. *Engineering Proceedings*, 39(1):79, 2023. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Lei Cai, Jingyang Gao, and Di Zhao. A review of the application of deep learning in medical image classification and segmentation. *Annals of Translational Medicine*, 8(11):713, June 2020.

[8] Senay Cakir, Marcel Gauß, Kai Häppeler, Yassine Ounajjar, Fabian Heinle, and Reiner Marchthaler. Semantic Segmentation for Autonomous Driving: Model Evaluation, Dataset Generation, Perspective Comparison, and Real-Time Capability, July 2022. arXiv:2207.12939 [cs].

[9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, May 2017. arXiv:1606.00915 [cs].

[10] Adrian V. Dalca, John Guttag, and Mert R. Sabuncu. Anatomical Priors in Convolutional Networks for Unsupervised Biomedical Segmentation. pages 9290–9299, 2018.

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. ISSN: 1063-6919.

[12] Khaled ELKarazle, Valliappan Raman, and Patrick Then. Facial Age Estimation Using Machine Learning Techniques: An Overview. *Big Data and Cognitive Computing*, 6(4):128, December 2022. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[13] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazırbaş, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning Optical Flow with Convolutional Networks, May 2015. arXiv:1504.06852 [cs].

[14] Evgin Goceri. Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*, 56(11):12561–12605, November 2023.

[15] Lovedeep Gondara. Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 241–246, December 2016. arXiv:1608.04667 [cs, stat].

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[17] Xifeng Guo, Xinwang Liu, En Zhu, and Jianping Yin. *Deep Clustering with Convolutional Autoencoders.* October 2017. Pages: 382.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. pages 770–778, 2016.

[19] Shreya Kadambi, Zeya Wang, and Eric Xing. WGAN domain adaptation for the joint optic disc-and-cup segmentation in fundus images. *International Journal of Computer Assisted Radiology and Surgery*, 15(7):1205–1213, July 2020.

[20] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia F. J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, December 2016. arXiv:1612.08894 [cs].

[21] Aghiles Kebaili, Jérôme Lapuyade-Lahorgue, and Su Ruan. Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review. *Journal of Imaging*, 9(4):81, April 2023. arXiv:2307.13125 [cs, eess].

[22] Khalil Khan, Rehan Ullah Khan, Kashif Ahmad, Farman Ali, and Kyung-Sup Kwak. Face Segmentation: A Journey From Classical to Deep Learning Paradigm, Approaches, Trends, and Directions. *IEEE Access*, 8:58683–58699, 2020. Conference Name: IEEE Access.

[23] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].

[24] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes, December 2022. arXiv:1312.6114 [cs, stat].

[25] Martin Kostelansky, Ana Manzano Rodriguez, Jan Kybic, Miroslav Hekrdla, Ondrej Dvorsky, Jiri Kozel, Patricie Baurova, David Pakizer, and David Skoloudik. Differentiating between stable and progressive carotid atherosclerotic plaques from in-vivo ultrasound images using texture descriptors. In Adam Walker, Letícia Rittner, Eduardo Romero Castro, Natasha Lepore, Jorge Brieva, and Marius G. Linguraru, editors, *17th International Symposium on Medical Information Processing and Analysis*, page 12, Campinas, Brazil, December 2021. SPIE.

[26] Martin Kostelanský. Lokalizace a segmentace in-vivo ultrazvukových obrazů karotidy. January 2021. Accepted: 2021-01-22T11:51:26Z Publisher: České vysoké učení technické v Praze. Vypočetní a informační centrum.

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[28] Agostina J. Larrazabal, Cesar Martinez, and Enzo Ferrante. Anatomical Priors for Image Segmentation via Post-Processing with Denoising Autoencoders, June 2019. arXiv:1906.02343 [cs, eess].

[29] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation, April 2016. arXiv:1604.05144 [cs].

[30] Halgurd S. Maghdid, Aras T. Asaad, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, and Muhammad Khurram Khan. Diagnosing COVID-19 Pneumonia from X-Ray and CT Images using Deep Learning and Transfer Learning Algorithms, March 2020. arXiv:2004.00038 [cs, eess].

[31] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least Squares Generative Adversarial Networks, April 2017. arXiv:1611.04076 [cs].

[32] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Çağatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, José Antonió Mariz, Raphael Meier, Sérgio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10):1993–2024, October 2015.

[33] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral Normalization for Generative Adversarial Networks, February 2018. arXiv:1802.05957 [cs, stat].

[34] Artem Moroz. Predikce parametrů aterosklerotického plátu z in-vivo ultrazvukových obrázků karotidy. May 2022. Accepted: 2022-05-31T22:52:27Z Publisher: České vysoké učení technické v Praze. Vypočetní a informační centrum.

[35] Masoud S. Nosrati and Ghassan Hamarneh. Incorporating prior knowledge in medical image segmentation: a survey, July 2016. arXiv:1607.01092 [cs].

[36] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A. Cook, Antonio De Marvao, Timothy Dawes, Declan P. O'Regan, Bernhard Kainz, Ben Glocker, and

Daniel Rueckert. Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, February 2018.

[37] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training Recurrent Neural Networks, February 2013. arXiv:1211.5063 [cs].

[38] Ramin Ranjbarzadeh, Abbas Bagherian Kasgari, Saeid Jafarzadeh Ghoushchi, Shokofeh Anari, Maryam Naseri, and Malika Bendechache. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Scientific Reports*, 11(1):10930, May 2021. Number: 1 Publisher: Nature Publishing Group.

[39] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, January 2016. arXiv:1506.01497 [cs].

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation, May 2015. arXiv:1505.04597 [cs].

[41] Xiaotong Shi, Tianming Du, Shuang Chen, Honggang Zhang, Changdong Guan, and Bo Xu. UENet: A Novel Generative Adversarial Network for Angiography Image Segmentation. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1612–1615, July 2020. ISSN: 2694-0604.

[42] Yonggang Shi, Kun Cheng, and Zhiwen Liu. Hippocampal subfields segmentation in brain MR images using generative adversarial networks. *BioMedical Engineering OnLine*, 18(1):5, January 2019.

[43] Rashmi Siddalingappa and Sekar Kanagaraj. Anomaly Detection on Medical Images using Autoencoder and Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(7), September 2021. Number: 7 Publisher: The Science and Information (SAI) Organization Limited.

[44] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, April 2015. arXiv:1409.1556 [cs].

[45] Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. GANs for Medical Image Synthesis: An Empirical Study. *Journal of Imaging*, 9(3):69, March 2023. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[46] Leslie N. Smith. Cyclical Learning Rates for Training Neural Networks, April 2017. arXiv:1506.01186 [cs].

[47] Nasim Souly, Concetto Spampinato, and Mubarak Shah. Semi Supervised Semantic Segmentation Using Generative Adversarial Network. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5689–5697, Venice, October 2017. IEEE.

[48] Fabio Alexandre Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using Convolutional Neural Networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 2560–2567, July 2016. ISSN: 2161-4407.

[49] Gabriele Valvano, Andrea Leo, and Sotirios A. Tsaftaris. Learning to Segment From Scribbles Using Multi-Scale Adversarial Attention Gates. *IEEE Transactions on Medical Imaging*, 40(8):1990–2001, August 2021.

[50] Yuan Xue, Tao Xu, Han Zhang, L. Rodney Long, and Xiaolei Huang. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. *Neuroinformatics*, 16(3):383–392, October 2018.

[51] Siyi Xun, Dengwang Li, Hui Zhu, Min Chen, Jianbo Wang, Jie Li, Meirong Chen, Bing Wu, Hua Zhang, Xiangfei Chai, Zekun Jiang, Yan Zhang, and Pu Huang. Generative adversarial networks in medical image segmentation: A review. *Computers in Biology and Medicine*, 140:105063, January 2022.

[52] Samir S. Yadav and Shivajirao M. Jadhav. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1):113, December 2019.

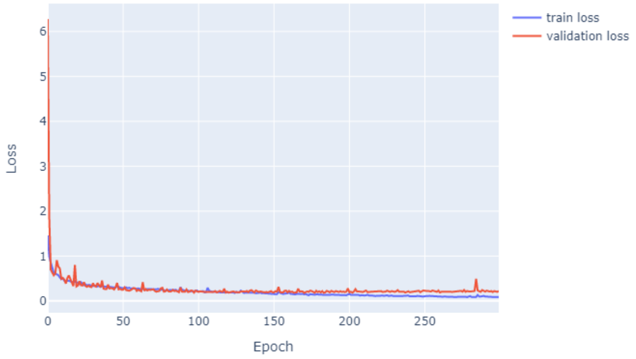
[53] David Školoudík, Petra Kešnerová, Tomáš Hrbáč, David Netuka, Jaroslav Vomáčka, Kateřina Langová, Martin Roubec, Roman Herzig, and Tomáš Belšan. Visual and digital analysis of the ultrasound image in a stable and progressive carotid atherosclerotic plaque. *Česká a slovenská neurologie a neurochirurgie*, 84/117(1):38–44, February 2021.

Appendix A

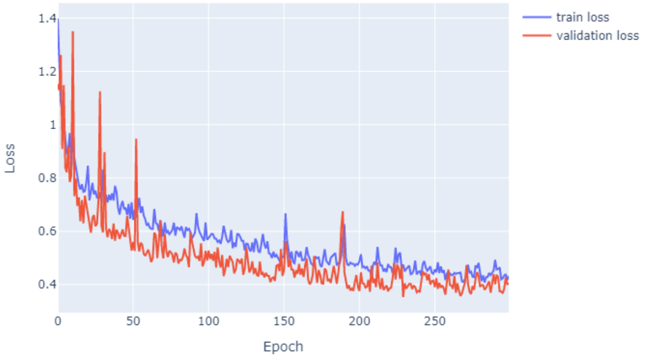
Segmentation experiment plots

We show plots related to segmentation experiments. We choose to separate them from the main experiments to avoid clutter. We show the training and validation losses in A.1 and validation IoUs for all classes after each training step in A.2

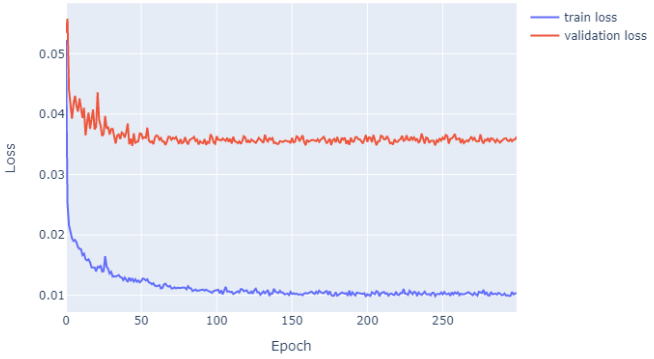
We show training and validation loss for the segmentation methods also we show validation mIoU and mean Dice score over the epochs.



(a) : ACNN train validation graphs

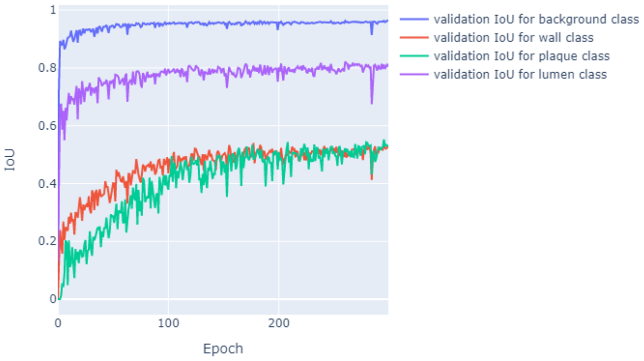


(b) : MAAG train validation graphs

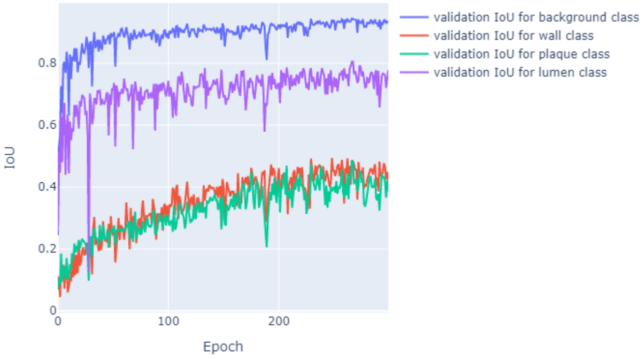


(c) : SeGAN train validation graphs

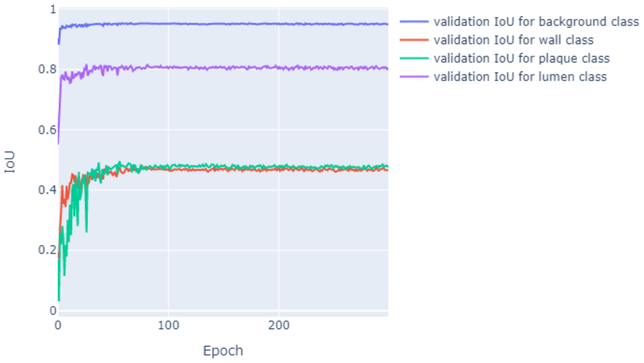
Figure A.1: Training and validation losses for FA-fix dataset



(a) : ACNN IoU for classes



(b) : MAAG IoU for classes



(c) : SeGAN IoU for classes

Figure A.2: Graph of validation IoUs for all classes over the epochs for FA-fix dataset

Appendix B

Implementation details

The project was implemented in the programming language Python and deep learning library PyTorch Code is available on GitLab. And the results and models are available on Google Drive Google Drive.

For the SeGAN method we used and adjusted pytorch code from the author's repository[1]. For the MAAG we reimplemented the author's repository[2] from Tensorflow to PyTorch. For the U-Net benchmarks we used the models from this repository[3] and for the generation of synthetic dataset we adjusted the method from this repository[4].

[1] https://github.com/YuanXue1993/SegAN

[2] https://github.com/gvalvano/multiscale-adversarial-attention-gates

[3] https://github.com/kostelansky17/carotids

[4] https://gitlab.fel.cvut.cz/morozart/bachelor-thesis