Two-stage and dual-decoder convolutional U-Net ensembles for reliable vessel and plaque segmentation in carotid ultrasound images

1st Meiyan Xie
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, USA
mx42@njit.edu

2nd Yunzhu Li
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, USA
yl744@njit.edu

3rd Yunzhe Xue
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, USA
yx277@njit.edu

4th Lauren Huntress
Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, NJ, USA
huntrela@rwjms.rutgers.edu

5th William Beckerman
Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, NJ, USA
beckerwe@rwjms.rutgers.edu

6th Saum A. Rahimi
Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, NJ, USA
rahimisa@rwjms.rutgers.edu

7th Justin W. Ady
Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, USA
jwa60@rwjms.rutgers.edu

8th Usman W. Roshan
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, USA
usman@njit.edu

***Abstract*—Carotid ultrasound is a screening modality used by physicians to direct treatment in the prevention of ischemic stroke in high-risk patients. It is a time intensive process that requires highly trained technicians and physicians. Evaluation of a carotid ultrasound requires segmentation of the vessel wall, lumen, and plaque of the carotid artery. Convolutional neural networks are state of the art in image segmentation yet there are no previous methods to solve this problem on carotid ultrasounds. We introduce two novel convolutional U-net models for both vessel and plaque from ultrasound images of the entire carotid system. We obtained de-identified images under IRB approval from 226 patients. We isolated a total of 500 ultrasound images spanning the internal, external, and common carotid arteries. We manually segmented the vessel lumen and plaque in each image that we then use as ground truth. In 10-fold cross-validation all models attain over 90% accuracy for vessel segmentation. With a basic convolutional U-Net we obtained an accuracy of 66.8% for plaque segmentation. With our dual-decoder model we see an improvement to 68.8% whereas our two-stage model falls behind at 65.1% accuracy. However, if we gave our two-stage model the true correct vessel as input its plaque accuracy rises to 81.7% suggesting that the method has potential and needs more work. We ensemble our U-Net and dual decoder U-Net models to obtain confidence scores for segmentations. By considering high confidence outputs above the 60% and 80% thresholds the accuracy of our dual decoder U-Net rises to 75.2% and 87.3% respectively. Our work here shows the potential of dual and two-stage methods for vessel and plaque segmentation in carotid artery ultrasound images and is an important first step in creating a system that can independently evaluate carotid ultrasounds.**

***Index Terms*—vessel segmentation, plaque segmentation, convolutional U-Net, dual decoder, medical AI**

I. Introduction

Stroke is the 5th leading cause of death in the United States [1]. Annually, it is responsible for billions of dollars in lost income and health care costs. For this reason there is significant effort and investment in the prevention of stroke. Ischemic strokes account for 87% of all strokes. Narrowing and deposition of plaque in the carotid arteries due to atherosclerosis is the most common cause of ischemic stroke. Carotid ultrasound is a safe, low-cost procedure that is used as a screening test in patients with risk factors for atherosclerosis [2]. It allows physicians to stratify the stroke risk of a patient and identify those patients that will most benefit from medical therapy or surgical intervention.

During a vascular ultrasound high-frequency sound waves are transmitted into your body. The sound waves are reflected back to the probe when they encounter the boundaries between different tissues in the body. This information is then utilized to create a 2D image of the vessel and surrounding tissue structures. Physicians utilize ultrasound images of the carotid artery in stroke prevention. During their evaluation physicians must first identify the vessel in the image. They then identify any atherosclerotic plaque within the wall and lumen of the vessel and finally they evaluate the physiologic impact of those plaques on the flow of blood within the vessel. This is a time intensive and resource intensive process that requires highly

skilled technicians and physicians to perform and interpret the results. As physician workload has increased and healthcare systems investigate ways to streamline processes and cut costs automating the interpretation of vascular ultrasounds has great potential.

Most prior work in automatic vessel and plaque segmentation consider 3D ultrasound images with narrowed regions of interest or MRI images, and few are based on deep or machine learning methods. For example convolutional U-Nets have been explored previously on 3D ultrasound images from the common carotid artery but their data is narrowed to a region of interest and they examined only images that contain plaque [3]. Traditional machine learning methods have been applied for plaque segmentation from B-mode ultrasounds of common carotid arteries [4]. Another study applied basic machine learning methods for plaque segmentation but on MRI images [5].

Other than the above there are plaque segmentation methods limited to video and common carotid arteries [6], histogram based methods on combined B-mode and contrast enhanced ultrasounds of common, internal, and external carotid arteries [7], and parametric and geometric deformable models followed by Bayesian classifiers on intravascular ultrasounds [8]. Previous work also includes image intensity and structure based methods on 3D ultrasounds of common carotid arteries [9], fuzzy clustering on MRI images for plaque detection only [10], [11], 3D volume-based level-set method on 3D ultrasounds of common, internal, and external carotid arteries [12], and a slice-based semi-automatic method on CTA images of common and internal carotid arteries [13].

In contrast to the above, our work considers raw ultrasound images without any pre-processing or narrowing the region of interest. These are taken directly from the hospital ward and are the same images that a trained physician would be looking at. Of note, 3D ultrasound is available only in research studies and is not commonly utilized clinically. Our proposed models are full end-to-end trainable convolutional U-Nets that allow for the segmentation of 2D ultrasounds, the most widely utilized modality. Deep learning models for vessel segmentation alone have been proposed previously on 2D and 3D carotid ultrasounds [14]–[16] but in this work we consider segmentation of both vessel and plaque.

II. Methods

A. Data collection

We obtained IRB approval from Robert Wood Johnson Medical School to use de-identified images from the Department of Vascular Surgery for this research. We manually downloaded B-mode carotid ultrasound examinations of 226 patients. We utilized an automated script to crop all patient identities from the ultrasound images and manually verified this de-identification.

We then cropped each image to obtain just the ultrasound removing all text and annotations on the image. Each images was resized to 224x224 pixels. We manually segmented the vessel lumen and plaque of 500 mages using the RectLabel software (https://rectlabel.com/). These serve as ground truth for training and validation

B. Background

1) Convolutional neural networks: Convolutional neural networks are the current state of the art in machine learning for image recognition [17], [18], including for MRI [19]. They are typically composed of alternating layers for convolution and pooling, followed by a final flattened layer. A convolution layer is specified by a filter size and the number of filters in the layer. Briefly, the convolution layer performs a moving dot product against pixels given by a fixed filter of size $k \times k$ (usually 3×3 or 5×5). The dot product is made non-linear by passing the output to an activation function such as a sigmoid or rectified linear unit (also called relu or hinge) function. Both are differentiable and thus fit into the standard gradient descent framework for optimizing neural networks during training. The output of applying a $k \times k$ convolution against a $p \times p$ image is an image of size $(p-k+1) \times (p-k+1)$. In a CNN, the convolution layers just described are typically alternated with pooling layers. The pooling layers serve to reduce dimensionality, making it easier to train the network.

2) Basic convolutional U-Net: After applying a series of convolutional filters, the final layer dimension is usually much smaller than that of the input images. For the current problem of determining whether a given pixel in the input image is part of a vessel or plaque, the output must be of the same dimension as the input. This dimensionality problem was initially solved by taking each pixel in the input image and a localized region around it as input to a convolutional neural network instead of the entire image [20].

A more powerful recent solution is the Convolutional U-Net (U-Net) [21]. This has two main features that separate it from traditional CNNs: (a) deconvolution (upsampling) layers to increase image dimensionality, and (b) connections between convolution and deconvolution layers.

We implemented a basic U-Net [21] in the Pytorch library [22] as shown in Figure 1. The U-Net is a popular choice for medical artificial intelligence work and has proven to be a successful baseline that can be built upon. The input to the model is an ultrasound image and output is an image of the same dimensions with 0, 1, and 2 pixel values indicating background, vessel lumen, and plaque.

Roughly speaking, in this model we first extract features with a series of convolutional kernels and then apply transpose convolutions to increase the dimensionality of the image up to the original. Thus we have an end-to-end network that is much simpler to train than otherwise patch-based approaches that have previously been used for segmentation.

C. Our proposed U-Net models

Aside from the basic U-net that we use as a baseline we investigate two extensions: a cascaded model with two networks and a dual decoder network with separate decoders for vessel and plaque.

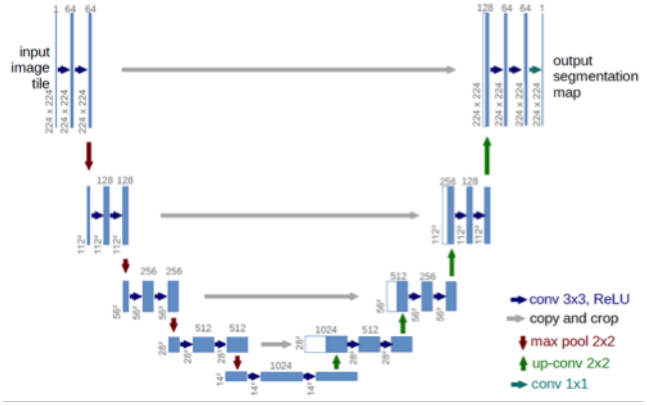

Fig. 1. Basic U-Net architecture [21] that we use as a baseline for our work. Shown here are dimensions of our images in each layer and the number of convolutional and transposed convolutions per layer.

1) Two-stage convolutional U-Net: In the two-stage approach we have a cascaded model of two convolutional U-Nets (as shown in Figure 2). In the first model we segment the vessel lumen from which we then segment the plaque with a second U-Net. For training the first network we manually segment the vessel of an additional 1661 images from our cohort of patients so as to have it be as accurate as possible. In fact we see later that the two-stage model indeed gives a better segmentation of the vessel lumen than basic U-Net and the dual decoder in the next subsection.

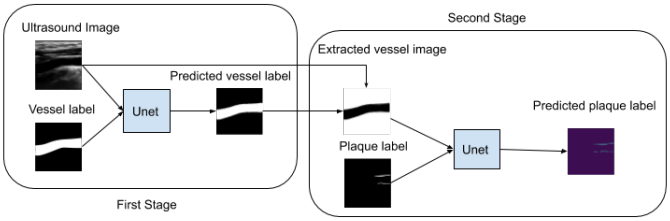

Fig. 2. Two-stage cascaded model containing two convolutional U-Nets

2) Dual decoder convolutional U-Net: The basic U-Net is a series of encoders and decoders with connections between them. In our dual-decoder we have a pair decoders in each step of the decoding (see Figure 3). In each pair one decoder is for segmenting the vessel and the other is for the plaque.

3) Ensemble models and confidence scores: We obtain confidence estimates by running each model 39 times starting with different seeds for the random number generator. From the outputs of each model we obtain a confidence estimate based on pixel frequencies as shown in Figure 4. We also obtain a majority vote prediction from the ensemble output as shown in *final prediction* in Figure 4.

4) Dice loss: The final output from the each of our models is a 2D predicted image of dimensions 224×224. We convert each pixel value into probabilities with softmax [23] and call

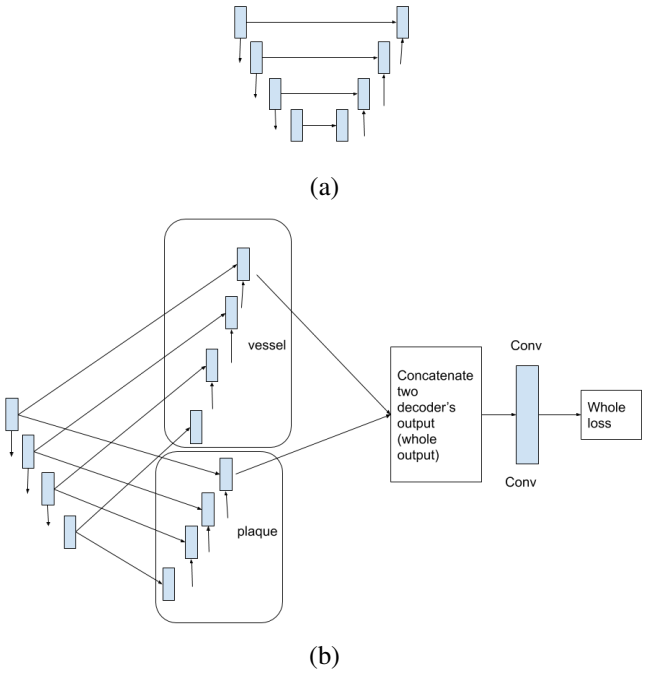

Fig. 3. Dual decoder convolutional U-Net with separate decoders for vessel and plaque. In (a) is the basic U-Net and in (b) is the dual decoder version.

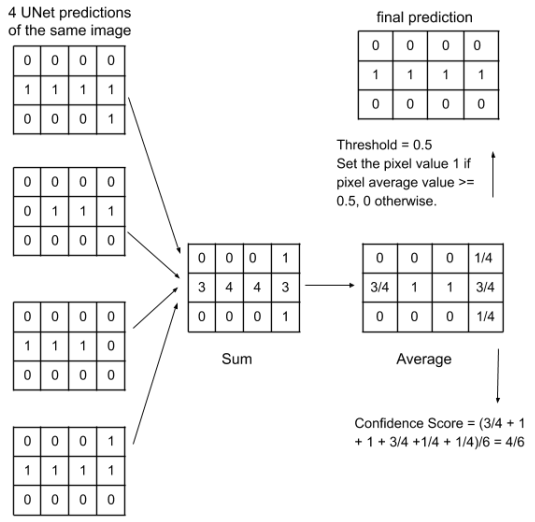

Fig. 4. Obtaining a confidence score and a final prediction from the outputs of 4 U-nets

the resulting image p. The target ground truth r is also of the same dimensions as p.

In the basic U-Net and our dual decoder the output is a segmentation of both the vessel and plaque. In the ground truth pixels we assign 1 if the pixel is within the vessel lumen, 2 if it in a plaque region, and 0 otherwise. In the two stage model we first output a vessel segmentation followed by a plaque. The ground truth for each network contains 1 if the pixel is in the vessel or plaque and 0 otherwise.

We then use the Dice loss to train our model. This is defined

to be $1 - D$ where

$$D(p) = \frac{2\sum_i p_i r_i}{\sum_i p_i^2 + \sum_i r_i^2}$$

and p_i and r_i are the i^{th} pixel values of p and r respectively.

D. Implementation, accuracy, and validation

1) Implementation: We implemented our models using Pytorch [22] and ran them on NVIDIA Pascal P100 and NVIDIA Titan RTX GPUs. We trained our models with 20 epochs of stochastic gradient descent [24], a learning rate of 0.03, decay step of 15 (with $\gamma = .1$), and a batch size of 4. We did not perform any normalization on the input images.

2) Measure of accuracy: Dice coefficient: The Dice coefficient is typically used to measure the accuracy of predicted segmentations in medical images [25]. The output from our network is a set of probabilities for each pixel indicating the class it belongs to. For example for a given pixel we have probabilities that it is background, vessel, or plaque. We assign 0, 1, or 2 depending upon the maximum probability. This then gives us an image with the same pixel values as the ground truth and allows us to calculate the Dice coefficient separately for vessel and plaque.

For vessel Dice each predicted pixel is determined to be either a true positive (TP, also one in ground truth), false positive (FP, predicted as one but zero or two in the ground truth), or false negative (FN, predicted as zero or two but one in ground truth). To calculate the plaque Dice coefficient we follow the same formula except that positive predictions have value two and negative are zero or one. After calculating these values the Dice coefficient is then formally defined as

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{1}$$

3) 10-fold cross-validation: We performed 10-fold cross-validation experiments on our data. We randomly split our dataset into ten equal parts and selected one part for validation while the remaining nine parts were used to train the model. We then rotated the validation part across the other nine parts giving us a total of 10 pairs of training validation splits. We trained the model on each split and reported the average validation and training accuracy below.

III. Results

A. Vessel and plaque segmentation

In Table I we see the vessel and plaque segmentation Dice accuracies of our models. Our dual decoder has the highest plaque accuracy followed by the basic model and then the two-stage. Interestingly if we were to use the true vessel as input to the second network in the two-stage model the plaque Dice accuracies increases to 0.82. This suggests there is room for improvement in this model: if we can get the first network to produce more accurate vessel segmentations it would in turn improve the plaque Dice accuracy o the second network.

In Figure 5 we see several ultrasound images, their ground truth segmentations, and their predicted segmentations by our dual decoder model. These are handpicked images where our model produces a visually correct vessel and plaque segmentation.

TABLE I
10-fold Dice coefficients of the basic U-Net and our two models

	Basic U-Net	Two stage	Dual decoder
Vessel Dice	.9	.95	.91
Plaque Dice	.67	.65	.69

In Figure 6 we show handpicked images with poorer segmentations. We see that in some cases our model produces false positives. Even though these images have low plaque segmentation accuracies, in some cases the segmentations are still useful than otherwise.

B. Ensemble and confidence scores

By ensembling we can obtain confidence scores for each model as described earlier. In Table II below we show the plaque Dice coefficients for different confidence score thresholds. As we raise the confidence threshold the plaque Dice coefficient increases with the dual decoder performing slightly better than the basic model. We get fewer images at higher thresholds but these are likely to be highly accurate as we see below.

TABLE II
10-fold plaque Dice coefficients of the basic U-Net and the dual-decoder model for different confidence thresholds, In parenthesis are the average number of images with confidence above the given threshold across the 10 folds.

Confidence threshold	Basic U-Net	Dual decoder
.5	.728 (34.7)	.732 (35.6)
.6	.753 (27.3)	.752 (27.9)
.7	.792 (16.2)	.795 (15.8)
.8	.859 (6.2)	.873 (5.1)

In an attempt to improve the vessel segmentation in our two-stage model we explored ensembling. We used the majority vote output (see Figure 4) of 39 models for vessel segmentation. This improved the plaque segmentation from .65 to .67 but not near the .81 accuracy we obtain with the true vessel image as input. Upon closer examination of our predicted vessel segmentations we may have found the source of our problem.

Even though vessel segmentations in the two-stage model are highly accurate (.94 Dice), if the segmentation is missing part of the vessel that contains the plaque it severely affects the plaque segmentation accuracy (see Figure 7). One way to address this is to expand the vessel segmentation to nearby pixels so that the plaque is included in the image for the plaque segmentation model. We plan to explore this in future work.

IV. Conclusion

Our work here shows the potential of dual and two-stage methods for vessel and plaque segmentation in carotid artery ultrasound images. This is an important first step in creating a system that can independently evaluate carotid ultrasounds.

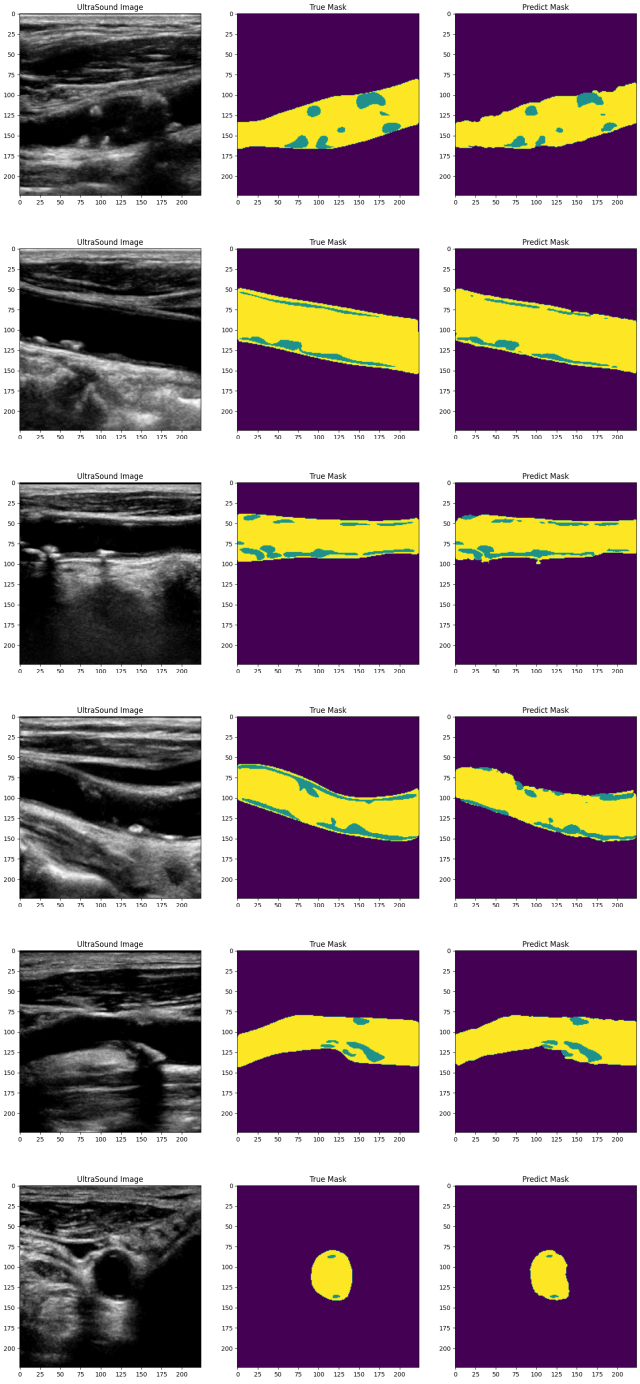

Fig. 5. Examples of ultrasound images and their true and predicted segmentations. In yellow is the vessel and green is the plaque. Our model performs very well in these handpicked images.

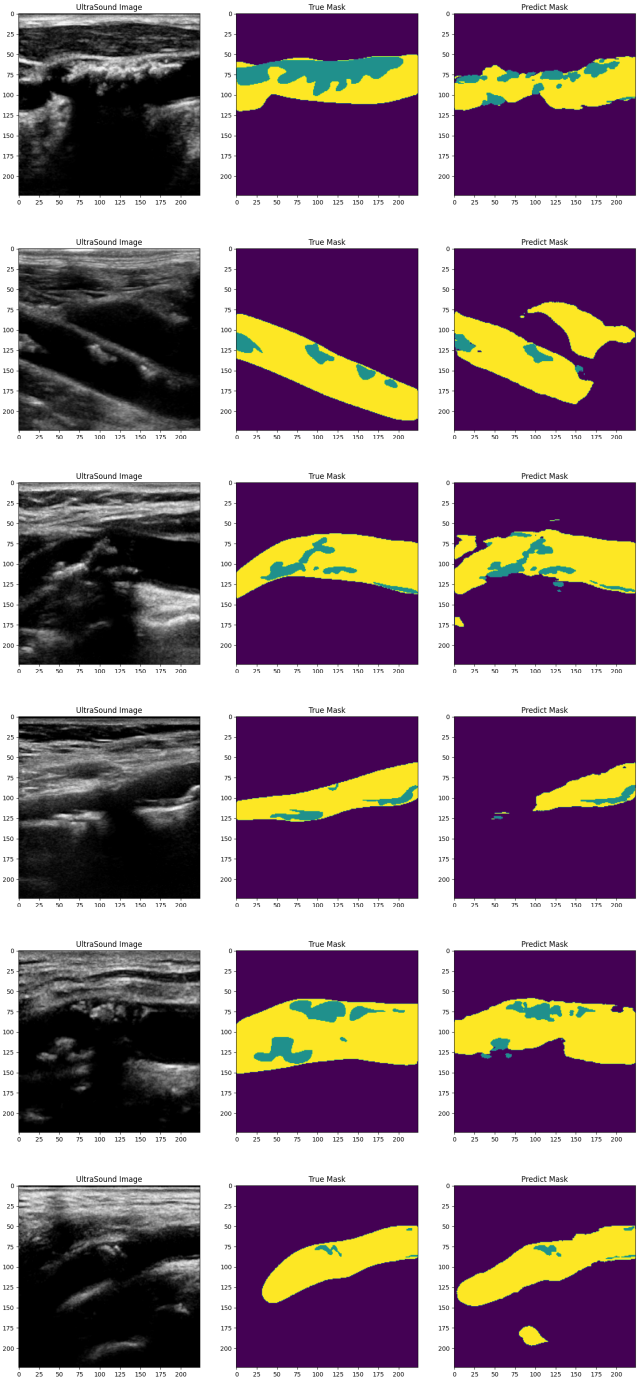

Fig. 6. Examples of ultrasound images and their true and predicted segmentations. In yellow is the vessel and green is the plaque. These are handpicked images where our model performs relatively poorly compared to images in Figure 5.

REFERENCES

[1] Dariush Mozaffarian, Emelia J Benjamin, Alan S Go, Donna K Arnett, Michael J Blaha, Mary Cushman, Sandeep R Das, Sarah De Ferranti, Jean Pierre Després, Heather J Fullerton, et al. Heart disease and stroke statistics-2016 update a report from the american heart association. *Circulation*, 133(4):e38–e48, 2016.

[2] James H Stein, Claudia E Korcarz, R Todd Hurst, Eva Lonn, Christopher B Kendall, Emile R Mohler, Samer S Najjar, Christopher M Rembold, and Wendy S Post. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the american society of echocardiography carotid intima-media thickness task force endorsed by the society for vascular medicine. *Journal of the American Society of Echocardiography*, 21(2):93–111, 2008.

[3] Ran Zhou, Wei Ma, Aaron Fenster, and Mingyue Ding. U-net based

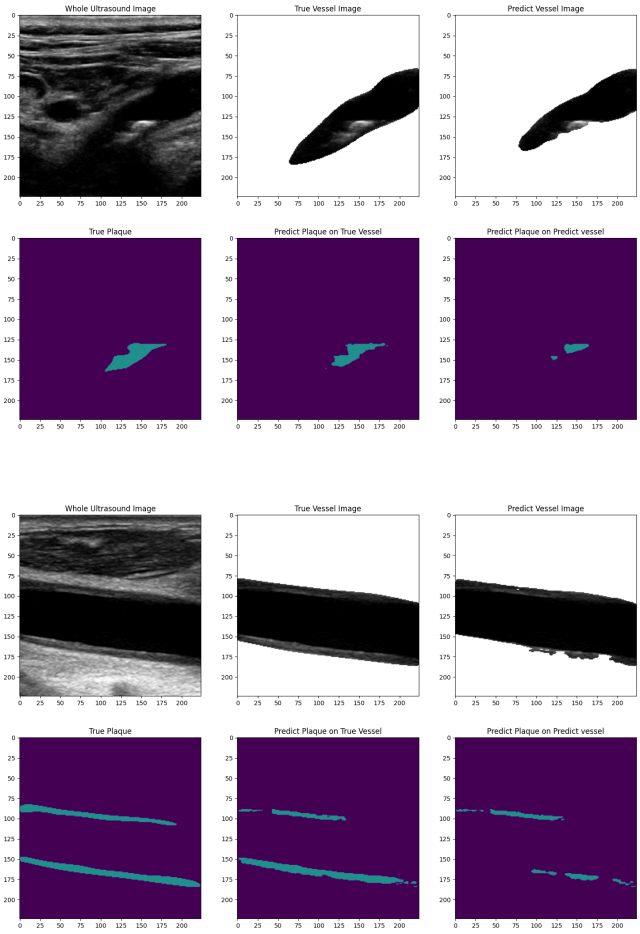

Fig. 7. Examples of ultrasound images and their true and predicted vessel segmentations from the two-stage model. Even though the overall vessel is correctly segmented some parts near the plaque are left out. The plaque segmentation model can thus never identify the plaque since it is missing altogether from the input. As a result the plaque segmentation in our two-stage model is lower than the base and dual decoder model.

automatic carotid plaque segmentation from 3d ultrasound images. In *Medical Imaging 2019: Computer-Aided Diagnosis*, volume 10950, page 109504F. International Society for Optics and Photonics, 2019.

[4] Chunjun Qian and Xiaoping Yang. An integrated method for atherosclerotic carotid plaque segmentation in ultrasound image. *Computer methods and programs in biomedicine*, 153:19–32, 2018.

[5] Qiang Zhang, Huiyu Qiao, Jiaqi Dou, Binbin Sui, Xihai Zhao, Zhensen Chen, Yishi Wang, Shuo Chen, Mingquan Lin, Bernard Chiu, et al. Plaque components segmentation in carotid artery on simultaneous non-contrast angiography and intraplaque hemorrhage imaging using machine learning. *Magnetic Resonance Imaging*, 60:93–100, 2019.

[6] Christos P Loizou, Styliani Petroudi, Constantinos S Pattichis, Marios Pantziaris, Takis Kasparis, and Andrew Nicolaides. Segmentation of atherosclerotic carotid plaque in ultrasound video. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 53–56. IEEE, 2012.

[7] Zeynettin Akkus, Nico de Jong, Antonius FW van der Steen, Johan G Bosch, Stijn CH van den Oord, Arend FL Schinkel, Diego DB Carvalho, Wiro J Niessen, and Stefan Klein. Fully automated carotid plaque segmentation in combined b-mode and contrast enhanced ultrasound. In *2014 IEEE International Ultrasonics Symposium*, pages 911–914. IEEE, 2014.

[8] Arash Taki, Zahra Najafi, Alireza Roodaki, Seyed Kamaledin Setarehdan, Reza Aghaeizadeh Zoroofi, Andreas Konig, and Nassir Navab. Automatic segmentation of calcified plaques and vessel borders in ivus images. *International Journal of Computer Assisted Radiology and Surgery*, 3(3-4):347–354, 2008.

[9] Jieyu Cheng, He Li, Feng Xiao, Aaron Fenster, Xuming Zhang, Xiaoling He, Ling Li, and Mingyue Ding. Fully automatic plaque segmentation in 3-d carotid ultrasound images. *Ultrasound in medicine & biology*, 39(12):2431–2446, 2013.

[10] Isabel M Adame, Rob J van der Geest, Bruce A Wasserman, Mona Mohamed, Johan Hans C Reiber, and Boudewijn PF Lelieveldt. Automatic plaque characterization and vessel wall segmentation in magnetic resonance images of atherosclerotic carotid arteries. In *Medical Imaging 2004: Image Processing*, volume 5370, pages 265–273. International Society for Optics and Photonics, 2004.

[11] IM Adame, RJ Van Der Geest, BA Wasserman, MA Mohamed, JHC Reiber, and BPF Lelieveldt. Automatic segmentation and plaque characterization in atherosclerotic carotid artery mr images. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 16(5):227–234, 2004.

[12] Jieyu Cheng, Yimin Chen, Yanyan Yu, and Bernard Chiu. Carotid plaque segmentation from three-dimensional ultrasound images by direct three-dimensional sparse field level-set optimization. *Computers in biology and medicine*, 94:27–40, 2018.

[13] Danijela Vukadinovic, Theo van Walsum, Sietske Rozie, Thomas de Weert, Rashindra Manniesing, Aad van der Lugt, and Wiro Niessen. Carotid artery segmentation and plaque quantification in cta. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 835–838. IEEE, 2009.

[14] Meiyan Xie, Yunzhu Li, Yunzhe Xue, Randy Shafritz, Saum A Rahimi, Justin W Ady, and Usman W Roshan. Vessel lumen segmentation in internal carotid artery ultrasounds with deep convolutional neural networks. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2393–2398. IEEE, 2019.

[15] Giles Tetteh, Velizar Efremov, Nils D Forkert, Matthias Schneider, Jan Kirschke, Bruno Weber, Claus Zimmer, Marie Piraud, and Bjoern H Menze. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *arXiv preprint arXiv:1803.09340*, 2018.

[16] Erik Smistad and Lasse Løvstakken. Vessel detection in ultrasound images using deep convolutional neural networks. In *Deep Learning and Data Labeling for Medical Applications*, pages 30–38. Springer, 2016.

[17] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[19] Jose Bernal, Kaisar Kushibar, Daniel S Asfaw, Sergi Valverde, Arnau Oliver, Robert Martí, and Xavier Lladó. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial intelligence in medicine*, 2018.

[20] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[22] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[23] Ethem Alpaydin. *Machine Learning*. MIT Press, 2004.

[24] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[25] Alex P Zijdenbos, Benoit M Dawant, Richard A Margolin, and Andrew C Palmer. Morphometric analysis of white matter lesions in mr images: method and validation. *IEEE transactions on medical imaging*, 13(4):716–724, 1994.