Vessel lumen segmentation in internal carotid artery ultrasounds with deep convolutional neural networks

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 Randy Shafritz
Division of Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, NJ, USA
shafrira@rwjms.rutgers.edu

5th Saum A. Rahimi
Division of Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, NJ, USA
rahimisa@rwjms.rutgers.edu

6th Justin W. Ady
Division of Vascular and Endovascular Therapy
Robert Wood Johnson Medical School
New Brunswick, USA
jwa60@rwjms.rutgers.edu

7th Usman 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 identification of the vessel wall, lumen, and plaque of the carotid artery. Automated machine learning methods for these tasks are highly limited. We propose and evaluate here single and multi-path convolutional U-neural network for lumen identification from ultrasound images. We obtained de-identified images under IRB approval from 98 patients. We isolated just the internal carotid artery ultrasound images for these patients giving us a total of 302 images. We manually segmented the vessel lumen, which we use as ground truth to develop and validate our model. With a basic simple convolutional U-Net we obtained a 10-fold cross-validation accuracy of 95%. We also evaluated a dual-path U-Net where we modified the original image and used it as a synthetic modality but we found no improvement in accuracy. We found that the sample size made a considerable difference and thus expect our accuracy to rise as we add more training samples to the model. Our work here represents a first successful step towards the automated identification of the vessel lumen in carotid artery ultrasound images and is an important first step in creating a system that can independently evaluate carotid ultrasounds.**

***Index Terms*—vascular ultrasounds, lumen segmentation, convolutional neural networks**

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 pulses are transmitted into your body. The sound waves travel into your body and hit the boundaries between tissues and are reflected back to the ultrasound probe. The ultrasound uses these signals to create a two dimensional image of the vessel being imaged. To evaluate a carotid ultrasound studies physicians first must 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.

Prior work in automated approaches to evaluating carotid ultrasounds is highly limited and does not use modern deep learning methods. Previously vessel identification in carotid ultrasounds with preprocessing and marker-controlled watershed transform has been explored [3]. Deep learning solutions have

been proposed for vessel segmentation in liver ultrasounds [4] and for vessel detection in femoral regions [5]. In the latter study authors also evaluate their method on carotid ultrasound images from two individuals, however their target is detection as opposed to segmentation that we seek. DeepVesselNet [6] is another deep learning model designed for vessel detection but in 3D angiographic volumes. A patch-based deep learning solution has also been proposed for 3D ultrasounds [7]. None of these are end-to-end systems that are simple to train and implement and non address vessel lumen segmentation in carotid ultrasounds that we seek here.

We present here a study to evaluate using a basic U-Net convolutional neural network to accurately identify the lumen of carotid artery in a vascular ultrasound. Our network is a simple end-to-end solution and addresses for the first time the problem of vessel lumen segmentation in carotid artery 2D ultrasounds. These ultrasounds are more affordable and common than 3D ones. Such ultrasounds are broadly used in vascular diagnosis and thus our solution has a broader impact than previous work.

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 the carotid ultrasound study of 98 patients. We utilized an automated script to crop all patient identifies from the ultrasound images and manually verified this de-identification. For this study we focused exclusively on the left and right internal carotid artery ultrasound images.

We cropped each image to obtain just the ultrasound removing all text and annotations on the image. Each images was resized to 224×224 pixels. This gave us a total of 302 images that we then manually segmented. Using RectLabel software (https://rectlabel.com/) we manually segmented the vessel lumen for each image to serve as ground truth for training and validation.

B. Convolutional neural networks

Convolutional neural networks are the current state of the art in machine learning for image recognition [8], [9], including for MRI [10]. 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.

C. Convolutional U-network

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 lesion, 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 [11].

A more powerful recent solution is the Convolutional U-Net (U-Net) [12]. 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.

D. Basic U-Net for vessel segmentation

We implemented a basic U-Net [12] in the Pytorch library [13] 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 and 1 pixel values indicating background and vessel lumen.

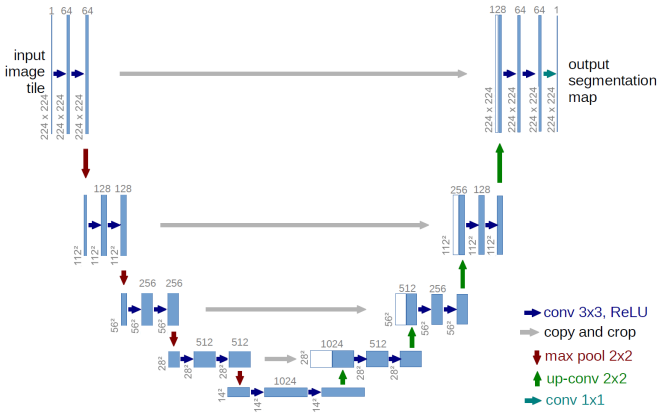

Fig. 1. Basic U-Net architecture [12] that we use in our preliminary work. Shown here are dimensions of our images in each layer and the number of convolutional and transposed convolutions per layer.

Roughly speaking, in our 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.

Inspired by our success in synthetic modalities for brain MRI systems [14] we utilized a dual-encoder model as well. We modified the original input ultrasound image by flipping it along the y-axis (called flip) and modifying the brightness and

contrast separately. These modified images were then used the as a second modality. In Figure 2 we see our dual-path encoder model that has a feature fusion for combining features from the two encoders.

Our feature fusion is a simple concatenation of features from the two encoders. In order to maintain the correct dimensions for the decoder we reduced each downsampling layer's output channels by half. In this way the concatenation restores the original dimensionality that is required by the decoder layer.

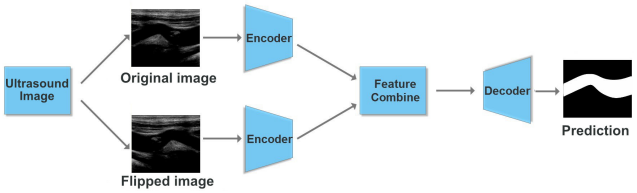

Fig. 2. Dual-encoder network that treats a modification of the input image as a synthetic modality.

E. Dice loss

The final output from our network is a 2D predicted image of dimensions 224×224. We convert each pixel value into probabilities with softmax [15] and call the resulting image p. The target ground truth r is also of the same dimensions as p and contains a 1 if the pixel is within the vessel lumen 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.

F. Model implementation and training

We implemented our system using Pytorch [13] and ran it on NVIDIA Pascal P100 and NVIDIA Titan RTX GPUs. We trained our model with 20 epochs of stochastic gradient descent [16], a learning rate of 0.03, decay step of 15, and a batch size of 1. We did not perform any normalization on the input images.

G. Post processing

We applied a simple post processing procedure to reduce potential false positives. In the final predicted segmentation we remove all disconnected components except for the largest one that is meant to be the vessel lumen. We found that this improved accuracy by a moderate margin.

H. Measure of accuracy: Dice coefficient

The Dice coefficient is typically used to measure the accuracy of predicted segmentations in medical images [17]. We convert the output image of our network into a binary mask by setting each pixel value to 1 if its softmax output is at least 0.5 and 0 otherwise. Thus we use 0.5 as the probability threshold that a pixel value is part of the vessel lumen or outside it. Starting with the human binary mask as ground truth, each predicted pixel is determined to be either a true positive (TP, also one in true mask), false positive (FP, predicted as one but zero in the true mask), or false negative (FN, predicted as zero but one in the true mask). The Dice coefficient is formally defined as

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

I. 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. Cross-validation on all images

In Table I we show the average 10-fold Dice values of vessel lumen segmentations separately for both training and validation samples. The training Dice is typically higher than validation as we see below and the validation is not far behind which suggests that our model is generalizing. We ran the 10-fold three times on our model to check for stability and found that each time our model gives a high training and validation accuracy.

Model train run	Vessel train	Vessel validation
Sample size of 234 images		
1	97.95%	93.96%
2	98.00%	93.73%
3	97.99%	93.67%
Sample size of all 302 images		
1	97.83%	94.63%

TABLE I
AVERAGE DICE COEFFICIENTS OF TRAIN AND VALIDATION SPLITS IN OUR 10-FOLD CROSS-VALIDATION ACROSS THREE DIFFERENT RUNS OF OUR MODEL.

In Figure 3 we see ultrasound images with their ground truth and predicted segmentations In (a) we see a vessel with no plaque or calcified walls. In (b) we see a vessel with thick calcified walls, in (c) we have a vessel with plaque, and in (d) we have a vessel with plaque and other regions above that could be mistaken for vessel lumen. In all four cases our model predicts the vessel lumen accurately as clear from the predicted segmentations. We found that the plaque and calcified wall does not affect the model and are contained within and outside the lumen respectively as we desire. From Figure 3(d) we also see that the model can tell the true vessel lumen from other regions that appear to be the lumen when in fact they are not.

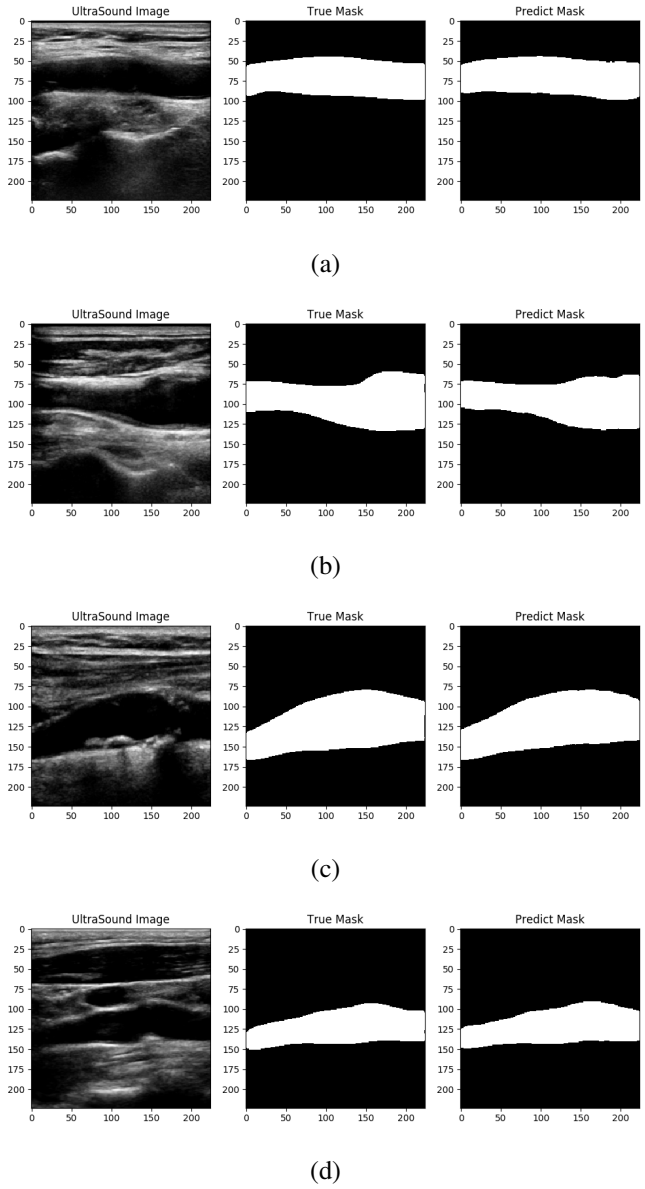

Fig. 3. Examples of ultrasound images with their manual ground truth and predicted segmentations.

B. Effect of sample size

We ran the training model with 5 different sample sizes (25, 50, 100, 234, and all 302 images). Figure 4 demonstrates the average Dice coefficient in our 10-fold experiment with the respective different sample sizes. We see that both training and validation accuracies increase as we add more samples. In fact the validation accuracy approaches the training one as the sample size increases.

C. Effect of dual-path encoder

To study the performance of the synthetic modality of our dual-encoder model, we modified the original input image by flipping it along the y-axis (called flip) and adjusting the brightness and contrast separately by a factor of 2 (using

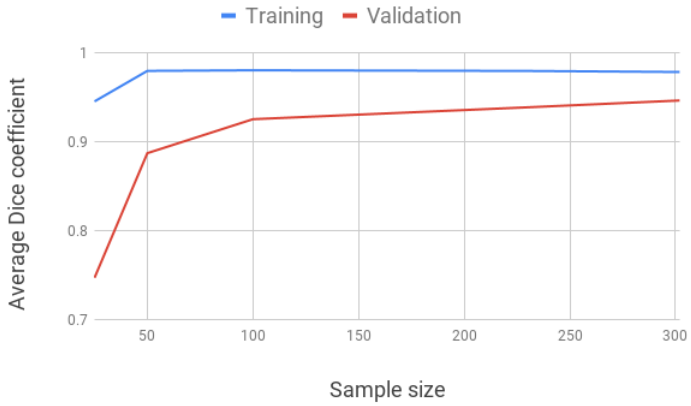

Fig. 4. As our sample size increases we see both training and validation accuracy in 10-fold cross-validation increases.

functions implemented in Pytorch and Torchvision). In Table II we show that all four synthetic modalities don't improve the validation accuracy of the single-path model.

Synthetic modality	Vessel train	Vessel validation
Flip	98.46%	91.76%
Brightness	98.48%	93.83%
Contrast	98.39%	93.7%

TABLE II
TRAINING AND VALIDATION DICE ACCURACY OF OUR DUAL-ENCODER MODEL WITH DIFFERENT SYNTHETIC MODALITIES.

D. Effect of modified Dice loss to emphasize recall

While our overall validation accuracy is at 95% our model still has difficulty in some cases. In Figure 5(a) we see that the predicted segmentation is incomplete for this image even though the image does not appear to be hard. While there are no false positives the true positive rate (recall) is low. To fix this we try a modified Dice loss below that upweights the recall component by a factor of 4.

$$D(p) = \frac{5\sum_i p_i r_i}{\sum_i p_i^2 + 4 \times \sum_i r_i^2}$$

We evaluate our model with the modified Dice loss on a subset of 234 samples on which we also report the original model's accuracy above. In Figure 5(b) we see that the modified loss improves the segmentation of this particular image. We see it also improves segmentation of the image shown in Figure 5(c) with our original Dice loss to cover the entire vessel with the modified loss in Figure 5(d).

While we see an improvement in individual images our average accuracy with the modified loss on train and validation is the same as the original one on 234 samples (see Table III below). This suggests that the new loss lowers the accuracy of other images and thus may not be the best solution to improve our recall.

IV. DISCUSSION

One of the difficulties that automated systems face in the evaluation of vascular ultrasounds is that image output and

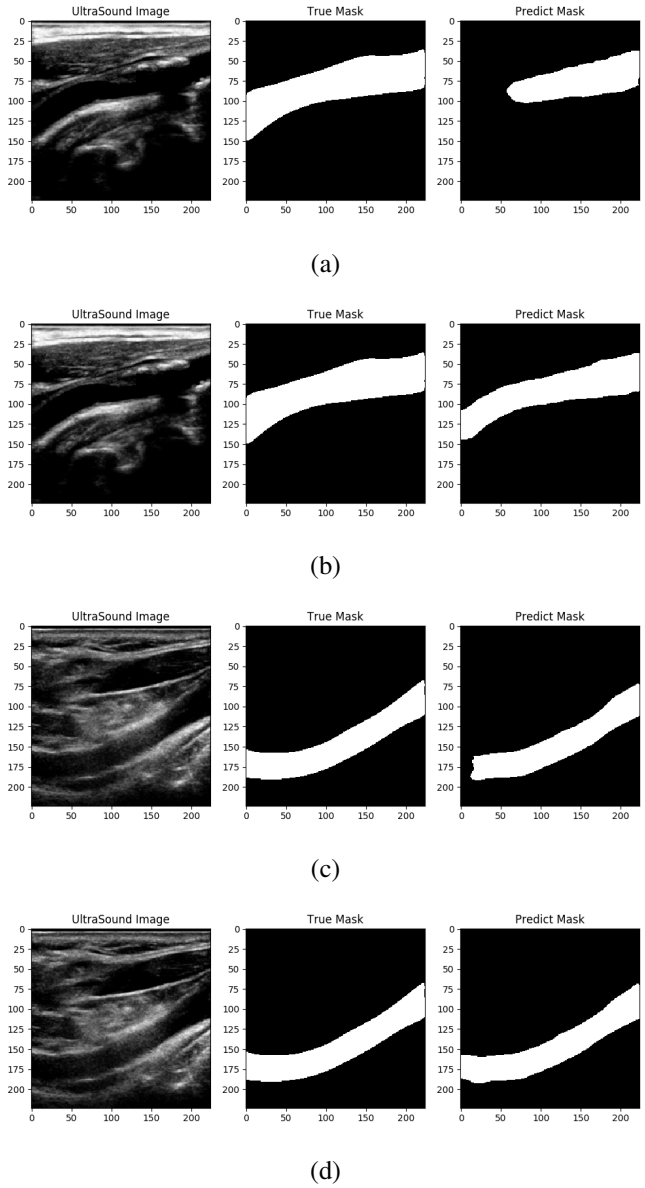

Fig. 5. Examples of ultrasound images and their manual ground truth and predicted segmentations obtained from the original Dice loss in (a) and (c) and the modified loss in (b) and (d).

Model train run	Vessel train	Vessel validation
1	97.75%	93.7%
2	97.77 %	93.51%
3	97.77 %	93.24%

TABLE III
AVERAGE DICE COEFFICIENTS OF TRAIN AND VALIDATION SPLITS IN OUR 10-FOLD CROSS-VALIDATION WITH OUR MODIFIED DICE LOSS ACROSS THREE DIFFERENT RUNS OF OUR MODEL.

quality is highly dependent upon both technician technique and the patient's body habitus and individual anatomy. Plaque and calcium in the vessel wall can result in acoustic shadows that make it difficult to visualize the posterior wall. At 95% accuracy the basic U-Net system that we have developed is able to adjust for these issues. In Figure 6 we present 4 different images where the posterior vessel wall is difficult to image and in each of these cases our system is able to accurately predict the correct vessel segmentation except for minor errors as in Figure 6(b).

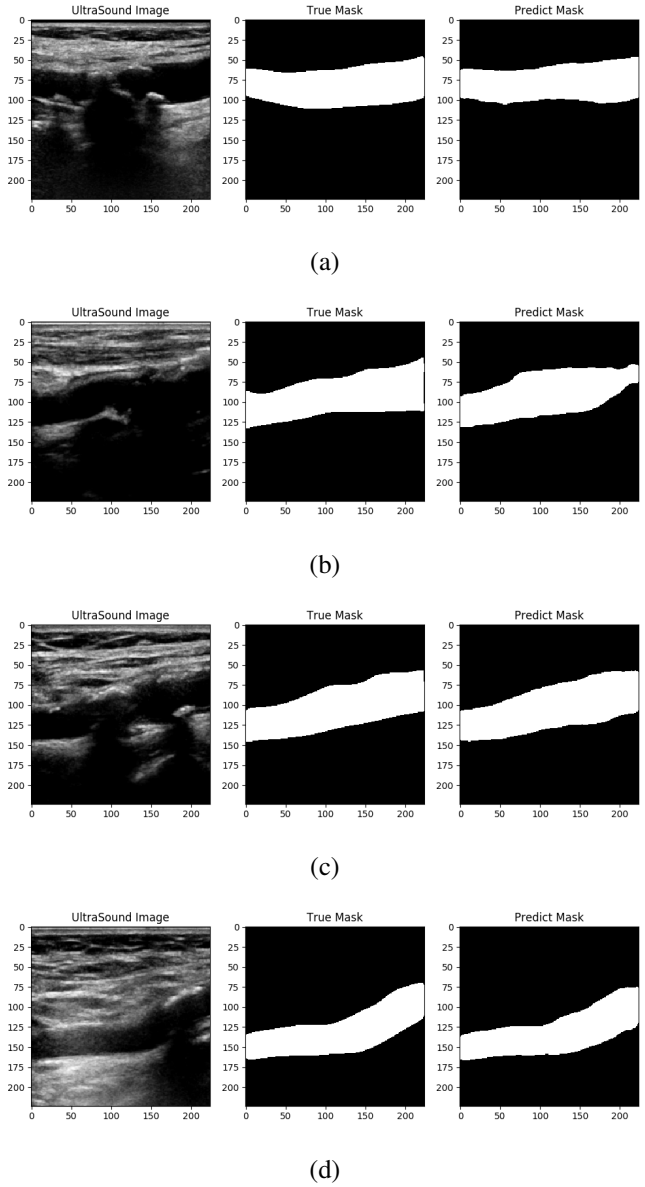

Fig. 6. Examples of ultrasound images and their manual ground truth and predicted segmentations that have part of the vessel missing from the image due to ultrasound shadowing.

We believe that the strengths of our model will be beneficial as we expand the scope of our work. While initially we focused on imaging just the internal carotid artery we next plan to expand this to the entire carotid artery study looking at the common carotid artery, external carotid artery and the carotid bifurcation. We chose the internal carotid artery initially due to the significant amount of plaque and atherosclerotic disease, making it one of the more difficult areas to get accurate predictions of the vessel lumen. Due to our successes with

internal carotid artery we expect an overall high accuracy on ultrasounds of the entire carotid artery system.

Accurate visualization of the vessel lumen is only the first step in creating a valid clinical tool that can evaluate vascular ultrasounds. Creating a system that can also identify and accurately segment atherosclerotic plaque, identify the vessel wall and accurately measure its width and identify calcification within the wall will be required. Identifying several regions of the vessel within the ultrasound is more challenging and falls under multi-class segmentation. Our encouraging results here suggest we should achieve high accuracy there as well but may need to add more images from patients just because multi-class classification typically requires more data than the binary case.

V. Conclusion

We evaluated a single and dual path convolutional neural network for vessel lumen segmentation in carotid artery vascular ultrasounds. In 10-fold cross-validation on 302 images from 98 patients we obtained 95% accuracy and expect this to rise as we add more images. Our work shows that vessel lumen segmentation can be achieved with high accuracy.

Acknowledgment

We thank NJIT Academic and Research Computing Services for computational support.

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] Philip Tamimi-Sarnikowski, Andreas Brink-Kjær, Ramin Moshavegh, and Jørgen Arendt Jensen. Automatic segmentation of vessels in in-vivo ultrasound scans. In *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10137, page 101371P. International Society for Optics and Photonics, 2017.

[4] Deepak Mishra, Santanu Chaudhury, Mukul Sarkar, Sidharth Manohar, and Arvinder Singh Soin. Segmentation of vascular regions in ultrasound images: A deep learning approach. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.

[5] 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.

[6] 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.

[7] Ran Zhou, Aaron Fenster, Yujiao Xia, J David Spence, and Mingyue Ding. Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images. *Medical physics*, 2019.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Yunzhe Xue, Meiyan Xie, Fadi G Farhat, Olga Boukrina, AM Barrett, Jeffrey R Binder, Usman W Roshan, and William W Graves. A fully 3d multi-path convolutional neural network with feature fusion and feature weighting for automatic lesion identification in brain mri images. *submitted*, 2019.

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

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

[17] 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.