Diameter Estimation in Common Carotid Artery (CCA) using Deep Learning Techniques

Nageswara Rao Gurram
University of Rochester

Ajay Anand
University of Rochester

October 9, 2020

1 Introduction

Ultrasonography of the carotid arteries is a common imaging study performed for diagnosis of carotid artery disease. There are many studies that pointed out the relationship between common carotid arteries (CCA) lumen diameter (LD) and cardiovascular diseases [1], [2]. Many studies show that decreased lumen diameter (LD) that are assessed from B-mode ultrasound images can be associated with stroke events [3], cardiovascular diseases (CVD). The changes in LD measurements over the time can be used to quantify the carotid artery stenosis. Thus, providing reliable and accurate lumen diameter gives us an attractive imaging biomarker.

Usually, medical experts perform these measurements manually while doing Ultrasonography, but this process is prone to errors, and has large observer bias [4]. There were many studies to estimate the lumen diameter in CCA by segmentation techniques. These algorithms work either by segmenting the the lumen region ([5], [6], [7]) or delineating the boundaries of vessel walls ([8], [9], [10]). P Krishna Kumar *et al.* have done a comprehensive survey on these techniques in [11].

However, these algorithms have many difficulties in segmenting the regions of interest due to artifacts such as speckle noise, acoustic shadowing, signal dropout, low contrast, curvature of arteries, presence of jugular vein above the CCA, deeper plaques protruding into the artery lumen, tortuous or tilted arteries, missing edges in carotid boundaries, occlusions caused by plaques, unpredictable bending of vessel along its major axis, variability in the artery shape and so on.

In our approach, we tried to directly assess the quantitative measurement of lumen diameter (LD) using recent developments in deep learning without any segmentation. The standard computer-vision models we've used are robust to many problems that were described above and can detect those 'interested regions' that explains ground truths. We trained deep learning models in supervised fashion by labeling the data available at Signal Processing Laboratory [12] and showed that our estimates are very close to true values.

2 Methodology

The recent advances in computer vision ([13] [14], [15]) focused on improving convolutional neural networks in obtaining the best feature representations of images. Though these models are data hungry, by using 'Transfer Learning' techniques we can apply them even to the environments where data is small and different from the original data they were trained on. It fits well to our problem, as we're trying to estimate a quantitative value (diameter) from a small set of ultrasound images. This data is very different and small when compared to our pre-trained models data i.e ImageNet [16].

The steps in our approach to this problem is illus-

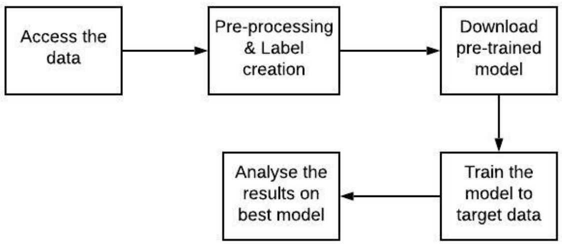

Figure 1: A step-by-step approach in solving the problem

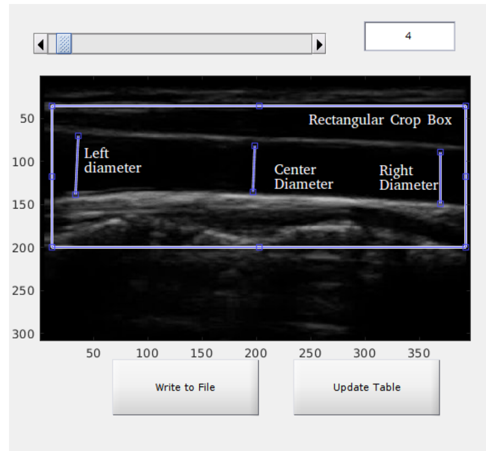

Figure 2: An example illustration of MATLAB software run on an image

trated in Figure. 1. First, the data is accessed from Signal Processing Laboratory database [12]. As this repository doesn't provide any ground truths and not particularly created for our problem, a custom MATLAB software is developed to calculate the diameter of each image. Then we've chosen three pre-trained models on ImageNet, to train and evaluate on our labeled data. Finally, the analysis on the results for estimated diameters are reported and discussed.

3 Data

The ultrasound B-mode images of common carotid artery for our experiments are taken from Signal Processing Laboratory database [12]. The details about the data can be observed from the same page.

The database contains images of common carotid artery (CCA) of ten volunteers (mean age 27.5 $\pm$ 3.5 years) with different weight (mean weight 76.5 $\pm$ 9.7 kg). Images (usually eight images per volunteer) were acquired with Sonix OP ultrasound scanner with different set-up of depth, gain, time gain compensation (TGC) curve and different linear array transducers. The image database contains 84 B-mode ultrasound images of CCA in longitudinal section. The resolution of images is approximately 390 x 330 px. The exact resolution depends on the setup of the ultrasound scanner. Two different linear array transducers with different frequencies (10MHz and 14MHz) were used. These frequencies were chosen because of their suitability for superficial organs imaging. All images were taken by the specialists with five year experience with scanning of arteries. Images were captured in accordance to the standard protocol with patients lying in the supine position and with the neck rotated to the left side while the right CCA was examined.

3.1 Label Collection

As Signal Processing Laboratory database [12] contains only the images and didn't have any label information, we had to do some manual pre-processing to get the diameters. As our approach in estimating diameter is in supervised setting, labeling information is necessary to train the models.
A custom MATLAB software is implemented for pre-processing purposes. As in standard sonograph setting, only the vessel cross-section is focused using this software. So when the ultrasound image is given as input to model, it actually takes only the rectangular cropped area of the image as in Fig. 2. The other use of this software is to capture diameter at multiple cross-sections (left, center, right) in an image. This helps to solve problem of 'varying diameter across the image' and taking the average gives smoothed value.

3.2 Pre-processing and Augmentation

The original images in the dataset [12] have different sizes due to variability of the diameter of CCA. Images with larger diameter have more height compared to other images with smaller values. So, when we cropped focused part of the image with MATLAB software, then resulting images also had varying sizes. So, we resized the original images [12] to 256x256 pixels before running ‘Label Collection’ procedure with MATLAB software.

By preserving the labels, augmenting the images can help deep learning models to avoid over-fitting. In our method we augmented our 84-image dataset by cropping each image at three sections there by creating x3 images for each. However, at test time we supply complete image to predict either average or center diameter. Thus to help model give better predictions, we assigned low probability (p=0.1) for each split (left, center, right) of image and high probability (p=0.7) for complete image while feeding to network in training phase. The value of this ’p’ is chosen based on empirical evidence that ’0.1’ gave better results. Along with this, each image is randomly flipped horizontally and then normalized to the value range [0,1].

4 Model

In recent years ‘Transfer Learning’([17], [18], [19], [20]) showed prominent results across multiple domains where the data is insufficient to train a deep learning model from scratch. In our experiments, we focused on fine-tuning the pre-trained convolutions neural networks (CNN) models to our dataset. By using pre-trained models that are trained on larger dataset such as ‘Imagenet’ [16], there is an advantage of initializing network with better parameters. This means network is already good enough at initialization to learn low level (generic) feature representations that are common to both the domains and data.

In our experiments, we have taken some of the proven networks that worked well in image recognition challenges and adapted to fit our target data. These networks themselves contain different submodules in each layer like Convolution, Pooling, Activation (ReLu or Sigmoid), Batch-Normalization, Dropout, Fully Connected. As we’re using pre-trained models, we didn’t change architecture of any network. For details about these network architectures, one can refer to respective papers ([21], [15], [22], [23], [24], [25]).

The following networks are fine-tuned or used as feature extractor for our techniques.

- ResNet-18 [15]: For our experiments, we’ve taken a pre-trained Resnet-18 model and updated last conv layer (second conv5_x block) and FC layer weights with back-propagation techniques. The decision to unfreeze only the last two layers is based on empirical results.

- VGG-19 [23]: In our method, we experimented with VGG-19 network as feature extractor and only the last fully-connected layer is re-trained for our data.

- Inception-v3 [22]: This network is also used as feature-extractor [like VGG-19] there by freezing entire network while training except last fully-connected layer.

Since the labeling is done manually for each image and also as the diameters are varying across the cross-section, the errors in this manual process can affect the model in an adverse way. For this reason, we’ve used smooth L1-loss [26] (Huber loss) which more robust to noise and outliers instead of mean squared error(squared L2-loss) while training the models.

The data split for training, validation and test are 70%, 10%, 20% respectively. We’ve explored two sets of hyper-parameters for evaluation on test set.

For choosing the best batch-size and learning-rate we’ve done a mini grid-search on these hyper-parameter space and decided to run the experiments with two configurations (Table.1). For these experiments, we’ve fixed model as ‘Resnet18’ and the prediction to center diameter.

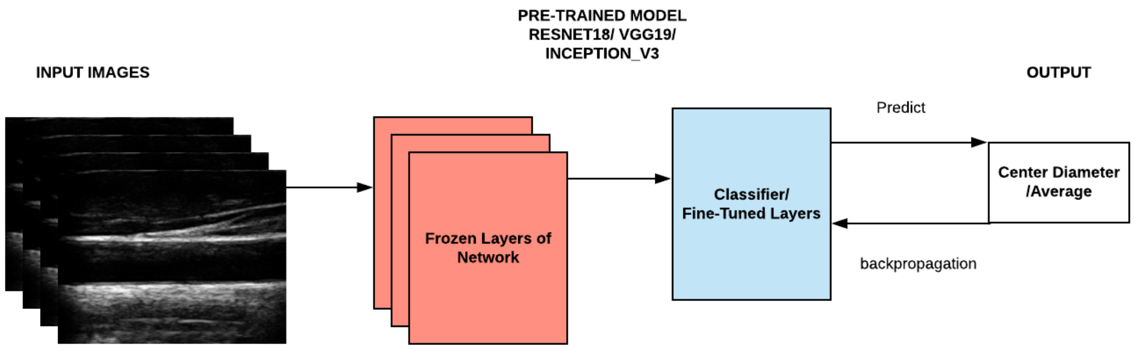

Figure 3: Model Outline

Hyper-parameter	**CONFIGa**	**CONFIGb**
Loss	Smooth L1 Loss	Smooth L1 Loss
Batch Size	4	8
Epochs	10	15
Learning rate	0.001	0.01
Optimizer Algorithm	Adam	Adam

Table 1: Two hyper-parameters configurations used all models in evaluations.

5 Results

In our experiments, we've evaluated 'mean absolute deviation' (MAD) and 'root mean square error' (RMSE) on the test dataset for various combinations of model architecture, prediction types. [refer section 3.2].

- Configuration: Two types of hyper-parameters configuration, refer Table. 1 for each configuration details.
- Prediction Type: The models are trained to predict either center diameter or average of the three (left, center, right) diameters.

Observation: The 'Resnet18' model predicted with least MAD, RMSE values while predicting either center diameter or average of three values. However, all models performed almost similar with both averaged and center diameter as prediction variable. Among the different models, Resnet18 seems to be performing better compared to others. This can be attributed to the simple network architecture of Resnet18 and its ability to learn residual errors with skip connections.

All the above results are on test dataset of split ratio=0.2 of total dataset size. Since the total data set size is only 84 images (without augmentation), the above results may not be convincing to show that model generalizes well for larger unseen data set. Cross-Validation helps in these scenarios to see whether model generalizes or not to unseen data.

We've done K-fold cross-validation with $1/K = 0.2$ [1] as parameter. The hyper-parameters and other set-

[1]Generally, during the training of neural network the best model may get converged even before running all epochs (hyper-parameter). So, to keep track of best model obtained till current epoch, we use another validation data (**val2**) which gives validation loss. Once all the epochs are done, the best model is at whichever epoch the validation loss is the least. We still do this even in K-fold Cross-Validation to select best model while training for each '(1-1/K) proportion' training data. For this reason we kept another small data set aside with split proportion as 0.1 out of this training data (**val2**)

Configuration	Model	Prediction Type	MAD	RMSE
CONFIGa	Resnet18	Center Diameter	**0.22**	**0.32**
CONFIGa	Vgg19	Center Diameter	0.53	0.69
CONFIGa	Inceptionv3	Center Diameter	0.65	0.82
CONFIGb	Resnet18	Center Diameter	**0.23**	**0.3**
CONFIGb	Vgg19	Center Diameter	1.27	1.5
CONFIGb	Inceptionv3	Center Diameter	0.7	0.91
CONFIGa	Resnet18	Average	**0.21**	**0.3**
CONFIGa	Vgg19	Average	0.6	0.87
CONFIGa	Inceptionv3	Average	0.84	1.14
CONFIGb	Resnet18	Average	**0.28**	**0.34**
CONFIGb	Vgg19	Average	1.26	1.54
CONFIGb	Inceptionv3	Average	0.69	0.8

Table 2: Test loss for different models when images resized to 256x256.

tings are chosen based on best results from Table. 2
.

Mean Absolute Deviation	**0.36**
RMSE	**0.513**

Table 3: K-fold Cross-Validation results on best fit model i.e Resnet18 with CONFIG1b setting.

The results of K-fold Cross-Validation in Table. 3 shows 'Mean Absolute Deviation' and 'Root Mean Square Error' between true diameters and predicted values. We can observe these values are relatively higher compared to the best result one test data (shown in Table. 2).

The residual error plot (Fig. 4) of K-fold CV shows that model predicted with error $> \pm$ 1 std. dev for very few images and there are only 3 samples with error $> \pm$ 2 std. dev. Almost all the predictions are with-in range of ± 0.5 of $error = 0$ line.

If we observe these images in Fig. 5, we can see two reasons for why model gave bad predictions.

- The walls of vessel may have blurry edges, this may be due to some noise caused while sonographer is using the scanner.
- The vessel may have non-uniform diameter across the image, this may confuse the network to predict wrong value.
- Enlarged vessel can have bigger diameter and if sufficient samples are not there, then model may learn it as an outlier and thus can lead to larger errors.

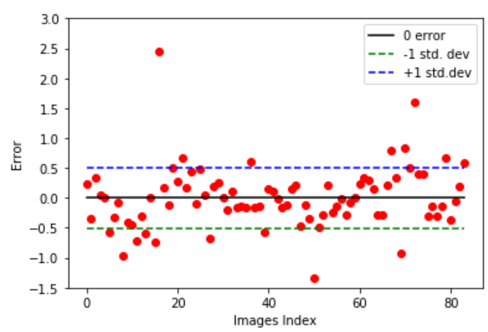

Figure 4: Errors on K-fold CV data

6 Conclusion

In this study, we tried to automate the current erroneous and inconsistent manual adjustment of diame-

in each step of K-fold Cross-Validation. So, the reader should identify the difference between **val2** (1/10-proportion of training data) and **val1**(a '1/K proportion').

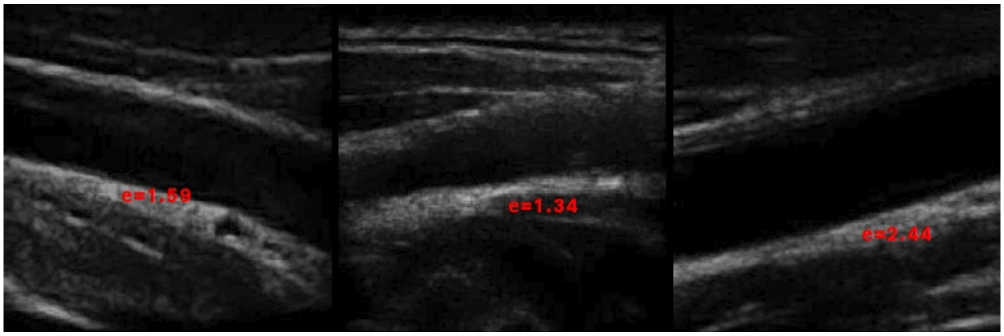

Figure 5: Images with errors > ± 2 std. dev.

ters in CCA. 'Mean Absolute Deviation (MAD)' and 'Root Mean Square Error (RMSE)' between manually calculated diameters and algorithm estimates are used as the metrics to show accuracy of model. The results indicate that for many images model predicted with very minimal error and deviations are observed for only samples with low-contrast & blurred edges and non-uniform diameters. Supplementing the model with more variant and accurately labeled data can further increase the accuracy.

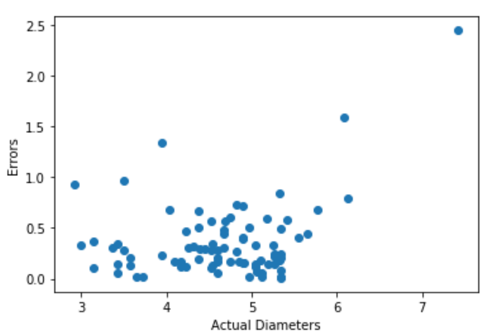

Figure 6: Errors vs Actual Diameters

Furthermore, from the Figure. 6 we can see that model predictions have large errors when actual diameters are high. This is acceptable while diagnosing stenosis since severity is high when the diameter is narrow. Stenosis severity is calculated using the ratio of narrowed diameter at stenosis area to normal diameter in carotid artery. Stenosis severity index ([27], [28]) in percentage can be formulated as:

$$SSI = (1 - \frac{LD_{narrow}}{LD_{normal}}) * 100$$

References

[1] Lewis H. Kuller Rachel H. Mackey Eric A. Wong Kim Sutton-Tyrrell Kelly D. Lloyd, Emma Barinas-Mitchell. Common carotid artery diameter and cardiovascular risk factors in overweight or obese postmenopausal women.

[2] Stéphane Laurent Gérard M. London Bruno Pannier Maryam Kavousi Francesco Mattace-Raso Oscar H. Franco Pierre Boutouyrie M. Arfan Ikram Sanaz Sedaghat, Thomas T. van Sloten and Coen D.A. Stehouwer. Common carotid artery diameter and risk of cardiovascular events and mortality.

[3] J.F. Polak. Peripheral vascular sonography.

[4] Meiburger KM Acharya UR Nicolaides A Suri JS. Saba L1, Molinari F. Inter- and intra-observer variability analysis of completely automated cimt measurement software (atheroedge^{TM}) and its benchmarking against commercial ultrasound scanner and expert readers.

[5] S. Golemati E.G. Sifakis. Robust carotid artery recognition in longitudinal b–mode ultrasound images.

[6] A. Campilho R. Rocha, J. Silva. Automatic detection of the carotid lumen axis in b-mode ultrasound images.

[7] A. Eriksson et al. M. Cinthio, T. Jansson. Evaluation of an algorithm for arterial lumen diameter measurements by means of ultrasound.

[8] H.S. Suri et al. T. Araki, P.K. Kumar. Two automated techniques for carotid lumen diameter measurement: regional versus boundary approaches.

[9] L. Sousa et al. A.M.F. Santos, J.M.R.S. Tavares. Automatic segmentation of the lumen of the carotid artery in ultrasound b-mode images.

[10] E.G. Sifakis et al. S. Golemati, J. Stoitsis. Using the hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery.

[11] Jeny Rajan John R Laird Andrew Nicolaides Jasjit S. Suri P Krishna Kumar, Tadashi Araki. State-of-the-art review on automated lumen and adventitial border delineation and its measurements in carotid ultrasound.

[12] Ultrasound image database, signal processing laboratory [online].

[13] Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner. Gradient-based learning applied to document recognition.

[14] I. Sutskever A. Krizhevsky and G. E. Hinton. Imagenet classification with deep convolutional neural networks.

[15] Shaoqing Ren Jian Sun Kaiming He, Xiangyu Zhang. Deep residual learning for image recognition.

[16] Imagenet databse. `http://www.image-net.org/`.

[17] Y. Bengio J. Yosinski, J. Clune and H. Lipson. How transferable are features in deep neural networks?

[18] T. Loupas A. A. Saad and L. G. Shapiro. Computer vision approach for ultrasound doppler angle estimation.

[19] Y. Bengio. Deep learning of representations for unsupervised and transfer learning,.

[20] Ivan Laptev Josef Sivic Maxime Oquab, Leon Bottou. Learning and transferring midlevel image representations using convolutional neural networks.

[21] Yangqing Jia Pierre Sermanet Scott Reed Dragomir Anguelov Dumitru Erhan Vincent Vanhoucke Andrew Rabinovich Christian Szegedy, Wei Liu. Going deeper with convolutions.

[22] Sergey Ioffe Jonathon Shlens Zbigniew Wojna Christian Szegedy, Vincent Vanhoucke. Rethinking the inception architecture for computer vision.

[23] Andrew Zisserman Karen Simonyan. Very deep convolutional networks for large-scale image recognition.

[24] Christian Szegedy Sergey Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift.

[25] Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov Nitish Srivastava, Geoffrey Hinton. Dropout: A simple way to prevent neural networks from overfitting.

[26] Peter J. Huber. Robust estimation of a location parameter.

[27] A.J. Fox. How to measure carotid stenosis.

[28] K.P. Kumar et al. T. Araki, A.M. Kumar. Ultrasound-based automated carotid lumen diameter/stenosis measurement and its validation system.