2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Convolutional Neural Network based Segmentation of Abdominal Aortic Aneurysms

Anish Salvi[1], Ender Finol[2], Prahlad G Menon[1,2]

Abstract— **Abdominal aortic aneurysms (AAAs) are balloon-like dilations in the descending aorta associated with high mortality rates. Between 2009 and 2019, reported ruptured AAAs resulted in ~28,000 deaths while reported unruptured AAAs led to ~15,000 deaths. Automating identification of the presence, 3D geometric structure, and precise location of AAAs can inform clinical risk of AAA rupture and timely interventions. We investigate the feasibility of automatic segmentation of AAAs, inclusive of the aorta, aneurysm sac, intra-luminal thrombus, and surrounding calcifications, using 30 patient-specific computed tomography angiograms (CTAs). Binary masks of the AAA and their corresponding CTA images were used to train and test a 3D U-Net - a convolutional neural network (CNN) - model to automate AAA detection. We also studied model-specific convergence and overall segmentation accuracy via a loss-function developed based on the Dice Similarity Coefficient (DSC) for overlap between the predicted and actual segmentation masks. Further, we determined optimum probability thresholds (OPTs) for voxel-level probability outputs of a given model to optimize the DSC in our training set, and utilized 3D volume rendering with the visualization tool kit (VTK) to validate the same and inform the parameter optimization exercise. We examined model-specific consistency with regard to improving accuracy by training the CNN with incrementally increasing training samples and examining trends in DSC and corresponding OPTs that determine AAA segmentations. Our final trained models consistently produced automatic segmentations that were visually accurate with train and test set losses in inference converging as our training sample size increased. Transfer learning led to improvements in DSC loss in inference, with the median OPT of both the training segmentations and testing segmentations approaching 0.5, as more training samples were utilized.**

I. INTRODUCTION

Abdominal aortic aneurysms (AAAs) are balloon-like dilations in the descending aorta associated with high mortality rates**.** AAAs can result in death due to dissections or ruptures involving blood leaking between artery walls or a portion of the artery itself bursting [1]. In the United States, between 2009 and 2019, ~28,000 people died from reported ruptured AAAs while ~15,000 people died from reported unruptured AAAs [2]. Due to the high mortality rate of ruptured AAAs, better characterization of AAA structure is needed. The first step in this process is accurate 3D segmentation of medical images which can be difficult for image processing pipelines involving region-growing annotation techniques underpinned on parametric curve evolution.

Medical image segmentation requires manual effort from clinical staff, familiarity with segmentation software, and is time consuming. We propose a convolutional neural network (CNN) based machine learning approach of adopting a 3D U-Net architecture to classify voxel-wise probabilities of a given voxel belonging to a vascular region that associates with an AAA [3]. Our goal is to segment the AAA structures accurately and consistently from computed tomography angiograms (CTAs) in a fraction of the time that it takes using manual approaches or traditional segmentation pipelines. Automatic segmentation allows for rapid determination of the presence of AAAs. Further, segmentation identifies the AAA structure and geometry to inform decisions on clinical care, especially those regarding elective surgery.

The U-Net we implemented has demonstrated differences from past machine learning approaches for segmentation. López-Linares *et al.* [4] reported a CNN capable of segmenting AAA intraluminal thrombus (ILT). Another CNN pipeline generated automatic segmentations of 2D aorta images for concatenation into a 3D aorta structure [5]. However, our U-Net segments not only the ILT but the entire AAA structure, inclusive of the aorta, in 3D. We selected a 3D architecture because 3D models can learn more organized and precise patterns in volumetric data than their 2D or 2.5D counterparts [6]. Furthermore, our training approach offers a different perspective on AAA segmentation. A prior 3D U-Net study explored the segmentation of AAAs using cropped CTAs, inclusive of only the AAA itself, to compare its accuracy with that of the original U-Net [7]. However, our 3D U-Net was trained on a CTA dataset, inclusive of varying regions of the body, not just the AAA region of interest. Thus, we evaluated our CNN for identification performance more comparable in a clinical setting in which CTA acquisition methods vary. We hypothesized that our trained CNN's segmentation performance would improve as the number of training masks increased, indicating consistency in learning.

II. METHODS

A. Imaging Dataset

30 patient-specific AAAs were manually segmented from CTAs using ITK-SNAP's 3D geodesic active contour segmentation module, applying corrections where needed [8]. The descending aorta, aneurysm sac, ILT, and calcifications (if present) of each AAA were segmented separately at first and later merged into 30 single, unified binary masks using an interactive voxel labeling method in ParaView (Kitware Inc.) [9]. The volumes, in cubic millimeters, of each segmentation feature were analyzed descriptively for relationships between aneurysm, ILT, and calcification volumes. The patients were separated into elective surgery and non-elective surgery groups to stratify rupture risk

[1] Department of Bioengineering, University of Pittsburgh; [2] Department of Mechanical Engineering, University of Texas at San Antonio

consistent with clinical decisions for repair. AAAs requiring elective surgery were assumed to have an elevated rupture risk compared to the non-elective surgery group. Non-elective surgery patients were considered as not having an anticipated need for surgery. Aneurysm, ILT, and calcification volumes were compared between groups.

B. U-Net based AAA Segmentation

The CTAs were of size 512x512xdim(z) where dim(z) was a patient specific length ranging from ~ 100-1000. AAA binary masks and CTAs were down sampled to 64x64x64 voxels using a SimpleITK image resampling pipeline. B-spline interpolation was adopted for CTAs whereas nearest-neighbor interpolation was adopted for masks [10]. We fit separate models using 10, 15, and 20 AAA binary masks and their respective CTAs (leaving the remaining CTAs in each case for out-of-sample testing). We ran each fitting exercise for 300 and 400 epochs and saved the resulting fit models. Six unique models differing in number of training samples and epochs run were developed to produce thirty automatic segmentations of each patient-specific AAA (including training and testing sets). The trained models predicted a voxel-level probability of belonging to the foreground mask, as defined by our training masks (viz. descending aorta and AAA sac, including ILT and calcification regions), in 3D.

We adopted Pytorch for our CNN design and instance normalization (as an alternative to batch normalization) to account for small batch sizes. The leaky rectified linear unit (LeakyReLU) activation function was used as an adaption to the original U-Net (which employed ReLU activation) [3]. The CNN consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a CNN but culminates in a bottleneck layer (viz. a feature map / space) before expanding back into a mask image that is identical in size to the input to the contracting path. As such, the expansive path facilitates upsampling of the feature map. To assess the models' accuracy, the Dice Similarity Coefficient (DSC) score was computed for each of the models' automatic segmentations. The DSC was calculated according to Eq. (1),

$$DSC = 2 \cdot (Y \cap P) / (Y + P) \quad \ldots (1)$$

where Y is the ground truth and P is the prediction. A higher DSC indicates a more accurate segmentation whereas we implemented a loss-function based on the DSC (i.e. 1 - DSC) that was iteratively minimized to fit the U-Net.

The U-Net's optimum probability thresholds (OPTs) for automatic segmentation were refined by colormap thresholding in ParaView, in lieu of receiver-operator-characteristic analysis [9]. Each voxel in the automatic segmentation was assigned a probability based on the likelihood it represented the AAA structure pursuant to neural inference using each of our 6 trained segmentation models. Determined by visual inspection, the model-specific probability threshold which created the most complete AAA representations in our testing set was defined as the OPT.

C. Determining Consistency with Transfer Learning

To assess the trained CNN's segmentation performance as the number of training samples increased, the model trained for 400 epochs with 15 binary masks was selected for transfer learning. This particular model was chosen owing to the consistency for binarizing voxel-level probability maps, indicated by its interquartile range of OPTs in training and corresponding testing sets. Transfer learning involves adapting the model's pre-trained weights by introducing new samples. The U-Net was trained on 5 new masks for 400 epochs followed by 5 new masks for another 400 epochs. The two resulting models, Dice Losses, OPTs, and automatic segmentations were collected.

III. Results

A. Imaging Dataset

ILT volume was observed to increase as aneurysm sac volume increased (adjusted $r^2 = 0.16$; $p = 0.02$). Calcification volume was found to increase with ILT volume ($p < 0.01$; adjusted $r^2 = 0.39$). The elective surgery group had a heavy right tailed distribution of aneurysm volume vis-a-vis the non-surgery group with similar heavy-tailed distributions in ILT and calcification volumes.

B. U-Net based AAA Segmentation

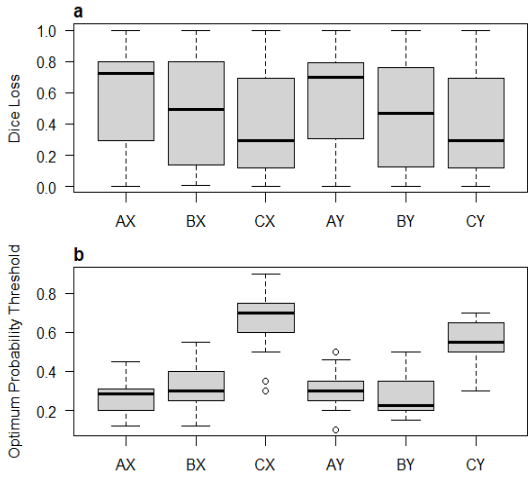

Figure 1. a) Box plots of Dice Losses (1 – DSC from Eq. (1)) from model-specific inference using each of the 6 trained U-Nets. Letters A, B, and C denote models that emerged from 10, 15 and 20 AAA based training runs, respectively, whereas X and Y denote the models that were trained for 300 and 400 epochs, respectively; b). Box plots of optimum probability thresholds for automatic segmentation using each of the 6 trained models. Dice Losses and OPTs per model reported above consist of both training and testing sets.

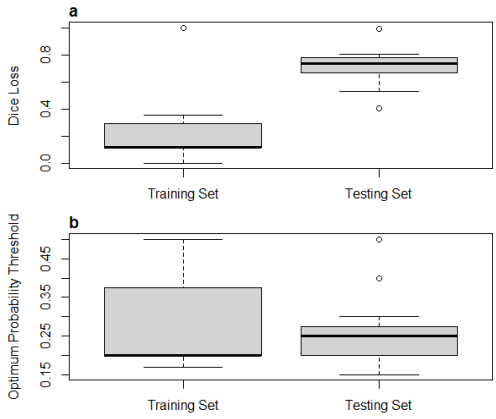

Figure 2. a) Box plots of Dice Losses (1 – DSC from Eq. (1)) of training and testing segmentations produced by the BY model. BY refers to the model trained with 15 binary masks for 400 epochs; b) Box plots of the optimum probability thresholds for the training and testing segmentations produced by the BY model.

The CNN models trained with 10, 15, and 20 masks for 300 epochs produced respective median Dice Losses of 0.72, 0.50, and 0.30. The CNN models trained with 10, 15, and 20 masks for 400 epochs produced respective median Dice Losses of 0.70, 0.47, and 0.30. Dice Loss decreased as the number of training masks increased from 10 to 20 masks. Dice Loss remained fairly constant as the number of epochs run increased from 300 to 400 epochs (Figure 1a). The CNN models trained with 10, 15, and 20 masks for 300 epochs produced respective median OPTs of 0.29, 0.30, and 0.70. The CNN models trained with 10, 15, and 20 masks for 400 epochs produced respective median OPTs of 0.30, 0.23, and 0.55. The OPT increased as the number of training masks increased from 10 to 20 masks, approaching 0.5 for 400 epochs run (Figure 1b).

The BY model which was trained with 15 masks for 400 epochs achieved a median Dice Loss of 0.12 for its training set and 0.74 for its testing set (Figure 2a). The median Dice Losses for the training sets were significantly lower than those for the testing sets in all six of the models generated. The same BY model achieved a median OPT of 0.2 for its training set and 0.25 for its testing set with a complete interquartile range overlap (Figure 2b). This overlap was the most consistent amongst all models fitted, when comparing training and testing set-specific OPTs.

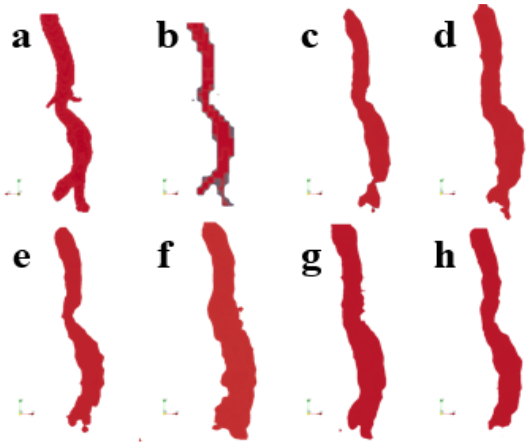

Figure 3. a) Manual Binary Segmentation; b) Binary Testing Mask; Automatic Segmentation from c) AX; d) BX; e) CX; f) AY; g) BY; h) CY. A, B, and C refer to the models trained on 10, 15, 20 binary masks, respectively. X and Y refer to the models trained for 300 and 400 epochs, respectively. This patient specific AAA was part of the testing set for all six models. Visualization was accomplished in ParaView [8].

Upon visual inspection, AAA automatic segmentations were similar to the corresponding manual segmentation (Figure 3). However, some automatic segmentations contained artifacts that could not be removed by increasing or decreasing the OPT.

C. Determining Consistency with Transfer Learning

Model BY, the model trained with 15 samples for 400 epochs, was chosen to determine the impact of new data introduced by transfer learning. Segmentation performance was assessed by Dice Loss and optimal thresholding of voxel-level probabilities.

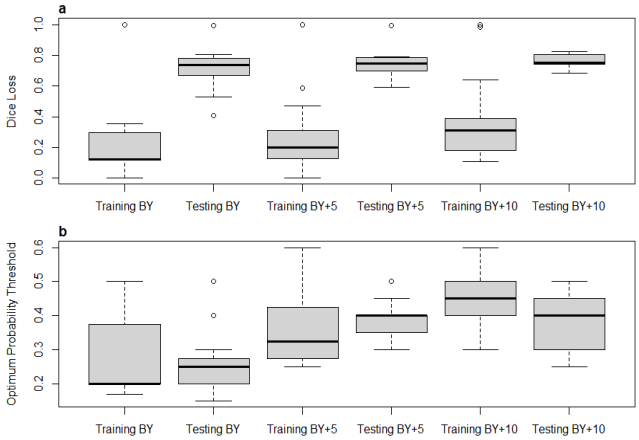

Figure 4. a) Box plots of the Dice Losses (1 – DSC from Eq. (1)) for the transfer learning trained models (testing and training sets). BY refers to the model trained with 15 binary masks for 400 epochs. BY+5 refers to the pre-weighted BY model trained for an additional 5 binary masks, 400 epochs. BY+10 refers to the pre-weighted BY+5 model trained for an additional 5 binary masks, 400 epochs. b) Box plots of the optimum probability thresholds for the transfer learning trained models (testing and training sets).

The CNN model trained with 15, 20, and 25 masks for 400 epochs produced respective median Dice Losses of 0.12, 0.20, and 0.31 for its training set and 0.74, 0.75, 0.75 for its testing set. Dice Loss increases for the training set and remains consistent for the testing set as the number of training masks increases (Figure 4a). The CNN model trained with 15, 20, and 25 binary masks for 400 epochs produced respective median OPTs of 0.20, 0.33, 0.45 for the training set and 0.25, 0.40, 0.40 for the testing set. The median OPT increases as the number of training masks increases for both training and testing sets (Figure 4b).

IV. Discussion

A. Imaging Dataset

Using $\alpha = 0.05$ for statistical significance, aneurysm volume was found to be proportionate to ILT volume as was ILT volume to calcification volume. The relationship between aneurysm and ILT volume is supported with literature implying that high volume aneurysms are associated with large amounts of ILT [11]. The non-elective surgery group had aneurysm sac, ILT, and calcification volumes that followed a normal distribution, while the same quantities for the elective surgery group had a heavy right tailed distribution. This positive skew of volumes quantified in the elective surgery group is consistent with the notion that larger AAAs are more likely to be electively repaired. Descriptive characteristics of our dataset are consistent with AAA literature. A 2011 review indicated AAA size correlated with risk of AAA rupture [12]. Our assessment is further substantiated with evidence that ILT volume is greater in ruptured AAAs compared to unruptured AAAs [13]. Buijs *et al.* [14] claimed calcification is more prevalent in patients with ruptured AAAs than those who took preventative measures, which is consistent with our outcomes.

B. U-Net based AAA Segmentation

Increasing the number of binary masks provided for CNN training had a greater impact in reducing Dice Loss and increasing OPT as compared to the impact of increasing epochs run (Figure 1). The training of more robust U-Nets was correlated with increasing the number of samples. While the difference between the BY model's median Dice Loss between testing and training sets was 0.62, the BY model's OPT interquartile ranges overlapped for the training and testing sets suggesting that the BY model, which was trained with 15 masks for 400 epochs, was the most robust U-Net produced (Figure 2). The high Dice Loss of the testing set remained an issue in our models, nevertheless, the similarity in train v/s test OPT suggests consistency in performance. A marginal visual improvement in AAA structure was achieved because of model inference, when a greater number of samples was used for training (Figure 3). Nevertheless, some automatic segmentations still included artifacts that could not be removed by thresholding.

C. Determining Consistency with Transfer Learning

As the pre-trained BY model, which was previously trained with 15 masks for 400 epochs, was trained with an increasing number of samples at a constant 400 epochs, the median Dice Loss and its interquartile range of the training set increased to approach that of its testing set (Figure 4a). However, the models improved consistently with additional training data because the difference between the median Dice Loss of the training set and testing set reduced as the number of training images increased (Figure 4b). Further, as the pre-trained model was trained with additional binary masks, the OPT increased over the three trials runs, approaching 0.5.

D. Improvements

The U-Net model could improve its segmentation performance by training on varying binary masks and CTAs. Most binary masks and CTAs used for training were from an elective surgery group and most non-elective surgery binary masks formed the testing group. Therefore, our CNN may have overfitted for the elective surgery group. The elective surgery group had larger aneurysms, more ILT, and more surrounding calcifications than the non-elective surgery group, which is supported by AAA literature. To confirm the CNN did not overfit based on increased rupture risk AAAs, selecting masks in equal proportions from both elective surgery and non-elective surgery may improve the model's performance by having equal subgroups within the training and testing sets. To increase the number of training and testing samples while keeping the number of patient specific AAAs constant, data augmentation could be performed on the CTAs and their corresponding binary masks. These transformations could include rotating, flipping, and shifting the binary masks and their respective CTAs. The increase in training masks may decrease the Dice Loss and will increase the OPTs of our U-Net's segmentations. Another area of improvement would be obtaining the OPTs programmatically. Implementing a K-means or Otsu's thresholding algorithm could reduce the time required for manually determining thresholds and improve inter-reader reliability [15].

V. Conclusions

We demonstrate the feasibility of a 3D U-Net which automatically segments AAAs to accelerate clinical decision making regarding elective repairs and biomedical research based on patient-specific geometric characteristics. While the performance of our 3D U-Net can improve by increasing binary training masks, our pre-trained models segmented AAAs in out-of-sample CTAs with visual accuracy.

References

[1] Aortic Aneurysms. Centers for Disease Control and Prevention.

[2] Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[4] López-Linares, K., Aranjuelo, N., Kabongo, L., Maclair, G., Lete, N., Ceresa, M., ... & Ballester, M. A. G. (2018). Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using deep convolutional neural networks. *Medical image analysis*, *46*, 202-214.

[5] Fantazzini, A., Esposito, M., Finotello, A., Auricchio, F., Pane, B., Basso, C., ... & Conti, M. (2020). 3D Automatic Segmentation of Aortic Computed Tomography Angiography Combining Multi-View 2D Convolutional Neural Networks. *Cardiovascular engineering and technology*, *11*(5), 576-586.

[6] Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, *32*(4), 582-596.

[7] Habijan, M., Galić, I., Leventić, H., Romić, K., & Babin, D. (2020, September). Abdominal Aortic Aneurysm Segmentation from CT Images using Modified 3D U-Net with Deep Supervision. In *2020 International Symposium ELMAR* (pp. 123-128). IEEE.

[8] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 2006 Jul 1;31(3):1116-28.

[9] Ayachit, Utkarsh, The ParaView Guide: A Parallel Visualization Application, Kitware, 2015, ISBN 978-1930934306

[10] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, R. Beare, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research", J Digit Imaging., doi: 10.1007/s10278-017-0037-8, 31(3): 290-303, 2018.

[11] Harter, L. P., Gross, B. H., Callen, P. W., & Barth, R. A. (1982). Ultrasonic evaluation of abdominal aortic thrombus. *Journal of Ultrasound in Medicine*, *1*(8), 315-318.

[12] Aggarwal, S., Qamar, A., Sharma, V., & Sharma, A. (2011). Abdominal aortic aneurysm: A comprehensive review. Experimental & Clinical Cardiology, 16(1), 11.

[13] Singh, T. P., Wong, S. A., Moxon, J. V., Gasser, T. C., & Golledge, J. (2019). Systematic review and meta-analysis of the association between intraluminal thrombus volume and abdominal aortic aneurysm rupture. Journal of vascular surgery, 70(6), 2065-2073.

[14] Buijs, R. V., Willems, T. P., Tio, R. A., Boersma, H. H., Tielliu, I. F., Slart, R. H., & Zeebregts, C. J. (2013). Calcification as a risk factor for rupture of abdominal aortic aneurysm. European Journal of Vascular and Endovascular Surgery, 46(5), 542-548.

[15] López-Linares, K., Lete, N., Kabongo, L., Ceresa, M., Maclair, G., García-Familiar, A., ... & Ballester, M. Á. G. (2018, April). Comparison of regularization techniques for DCNN-based abdominal aortic aneurysm segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 864-867). IEEE.