Biomedical Engineering Letters

Deep Learning Approach for the Segmentation of Aneurysmal Ascending Aorta

--Manuscript Draft--

Manuscript Number:	BMEL-D-20-00089R1
Full Title:	Deep Learning Approach for the Segmentation of Aneurysmal Ascending Aorta
Article Type:	Original Article
Corresponding Author:	Salvatore Pasta Universita' degli Studi di Palermo Palermo, Sicilia ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita' degli Studi di Palermo
Corresponding Author's Secondary Institution:	
First Author:	Albert Comelli
First Author Secondary Information:	
Order of Authors:	Albert Comelli
	Navdeep Dahiya
	Alessandro Stefano
	Viviana Benfante
	Giovanni Gentile
	Valentina Agnese
	Giuseppe M Raffa
	Michele Pilato
	Anthony Yezzi
	Giovanni Petrucci
	Salvatore Pasta
Order of Authors Secondary Information:	
Funding Information:	Ministero della Salute (GR-2011-02348129) — prof Salvatore Pasta
	U.S. Army (W911NF-18-1-0281) — prof Anthony Yezzi
	National Institutes of Health (R01-HL-143350) — prof Anthony Yezzi

Abstract:

Purpose

Diagnosis of ascending thoracic aortic aneurysm (ATAA) is based on the measurement of the maximum aortic diameter, but size is not a good predictor of the risk of adverse events. There is growing interest in the development of novel image-derived risk strategies to improve patient risk management towards a highly individualized level.

Methods

In this study, the feasibility and efficacy of deep learning for the automatic segmentation of ATAAs was investigated using UNet, ENet, and ERFNet techniques. Specifically, CT angiography done on 72 patients with ATAAs and different valve morphology (ie, tricuspid aortic valve, TAV, and bicuspid aortic valve, BAV) were semi-automatically segmented with Mimics software (Materialize NV, Leuven, Belgium), and

	then used for training of the tested deep learning models. The segmentation performance in terms of accuracy and time inference were compared using several parameters. Results All deep learning models reported a dice score higher than 88%, suggesting a good agreement between predicted and manual ATAA segmentation. We found that the ENet and UNet are more accurate than ERFNet, with the ENet much faster than UNet. Conclusions This study demonstrated that deep learning models can rapidly segment and quantify the 3D geometry of ATAAs with high accuracy, thereby facilitating the expansion into clinical workflow of personalized approach to the management of patients with ATAAs.
Response to Reviewers:	Manuscript Number: BMEL-D-20-00089 Article Title: Deep Learning Approach for the Segmentation of Aneurysmal Ascending Aorta Journal Title: Biomedical Engineering Letters COMMENTS FOR THE EDITOR: Dear Editor We thank the Editor for his consideration about our study submitted to BMEL. We have taken reviewer comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission. Please consider that, because of a mistake during the submission process, there is a mismatch in the author order between the text and the metadata generated by web system. We confirm that the first author is Dr. Albert Comelli, and we will fix this error on author order during the submission of the revised manuscript. Specific Editor comments Editor comments #1: Please make sure to address all comments. The potential overlap between training and test sets is a very important issue. Authors: We thanks the Editor for this point. We clarified in the manuscript that no cross-contamination between training and test sets was performed. As clarified to the Reviewer #1, the set of CT image of the testing was different to that of training set. We modified the text to clarify this critical point. Editor comments #2: Although statement on IRB approval for this study is provided, the full name of the IRB/ethics committee that approved the study should be provided in the manuscript. Also, declaration on conflict of interest for all enlisted authors should be provided in the manuscript (Giovanni Petrucci conflict is not provided). Authors: The IRB approval was given by the local committee “Comitato Etico sezionale IRCCS ISMETT” with the number IRRB_04_04. This information was added at the beginning of the Section 2.1 “Study Population”. The declaration of conflict of interest was added for Giovanni Petrucci. Reviewer #1 We thank the reviewer for his or her valuable consideration about our study. We have taken comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission. We hope that the reviewer agrees. All changes in the text are highlighted in yellow Reviewer #1: Summary: The authors present a comparison of aorta segmentation networks from contrast enhanced ECG-gated CT scans. The reference standard on 72

patients is generated via semi-automated segmentation. Experiments are set up in a 5-fold cross-validation setup. Three networks are compared: Unet, ENet and ERFNet using the Tversky cost function for network optimization. ANOVA shows that the performance of ENet and Unet are comparable and higher than that of the ERFNet, being processing time with ENet less than half of that of the Unet.

Comments:
General comment: this is a 3D segmentation problem solved in a per-axial slice basis. One could easily envision 2.5D or 3D methods. Were they considered?
Authors: We have considered 3D segmentation. However, the computational cost of 3D networks is extremely prohibitive. One approach to deal with this is to down-sample the data. Even with a good graphics card (NVIDIA QUADRO P4000 with 8 GB of RAM), a 128x128x128 input model would be the highest resolution feasible. This amount of down-sampling will seriously exacerbate the class imbalance problem and introduce interpolation artifacts as well along with extremely high computational cost. Another drawback of the 3D approach, especially in our case, is that a 3D network would then mandate expanding the training dataset. We believe, even with data augmentations our rather small dataset of 72 patients would not have been enough to prevent overfitting. Once we saw clinically significant results with only a 2D network and in light of the limitations of a 3D approach we decided to stick with 2D networks. Starting with the small Enet network, 2.5D approaches could certainly be explored in the future. Eventually, we wish to transfer our results to actual clinical use and hardware constraints definitely come into play in such scenarios. Please also consider that our approach follows 2D approach that is consistent with a similar study (Fantazzini et al CVET 2020) that was published after our submission to this journal. A comment on this interesting consideration was added in the “Discussion” section with the following text:

“In this study, 3D segmentation was not adopted as this approach requires larger dataset and is memory-demanding. One approach to deal with 3D segmentation drawbacks is to down-sample the data or adapt 3D integration of 2D convolutional neural networks trained on orthogonal planes to provide a final 3D segmentation. As this study focused on the impact of different deep learning methods on the segmentation accuracy of dilated aortas, the efficacy and accuracy of 3D approach will be investigated in future studies.”

Section 2.5.2
- It is unclear how the data is processed. According to line 20, 'we extracted individual slices from all patient cases in the training fold and then used these individual slices as input for our models'. The authors should be very careful by not including slices from the same patient into the training and test sets. Slices from the same patient are highly correlated among them. If slices from the same patient are included into the training and the test set, then the segmentation results are artificially inflated due to such correlation. Please clarify if slices from the same patient appear in each fold's training and validation.

Authors: We are sorry for having confused the Reviewer but it is clear that image slices are different between training and test datasets. The data was divided into 5 folds at the patient level and not at slice level. So there was no cross-contamination between training and test sets. Slices from the same patient were never used for both training and testing the same model. We have clarified this aspect in “Training” Section with the following text:

“In stratified five-fold cross-validation, the data set (72 patients) were divided into 5 equal patient subsets, and the holdout method repeated 5 times. Consequently, for each of the three network models, we trained each single model 5 times. Each time, one of the 5 subsets was used as the testing set, and the other 4 subsets as training set. Slices from the same patient were never used for both training and testing purpose. So, there was no cross-contamination between training and test sets.”

- Data augmentation - horizontal flips. The heart is not symmetrical with respect to the horizontal axis. I am surprised such augmentation plays any effect on the performance.

Authors: In general, data augmentations are used for preventing overfitting by

expanding the training dataset. We recognize that the heart is not symmetric with respect to the horizontal axis (or vertical flips) so that flipping of slices will likely produce the same data. However, the fact is that not symmetric data augmentation forces the model to learn to detect the same features in different orientations and this helps to stabilize and regularize the network. Nothing was changed in the manuscript if this is permitted by the Reviewer.

- Cost function (line 49) - how were the parameters \alpha and \beta selected?
Authors: The values of alpha and beta were used as suggested by the work proposed by Salehi and collaborators (see reference 20). They reported that an alpha=0.3 and beta=0.7 lead to the best results. These values are derived from deep learning of multiple sclerosis lesion segmentation. The sentence was rephrased with the following text:

“We also adopted a batch size of 8 slices for all experiments and adopted a Tversky loss function with α = 0.3 and β = 0.7 as reported by Salehi et al. [20]. These values for the loss function demonstrated to be effective for training deep neural networks for sclerosis lesion segmentation.”

Figure 3. There are errors are the end of the segmentation of the descending aorta. Is this due to the reference standard? It is also impossible to see false negatives in these images. Maybe consider adding transparency to the reference standard depiction.
Authors: Figure 6 was modified adding transparency to the models. The errors at the end of the segmentation are caused by the boundary of the standard reference segmentation while the deep learning models still recognize this end of the vessel as a region deserving segmentation. Thus, the errors are caused by the reference model rather than the proposed deep learning algorithms. The sentence at the end of “Results” section was re-written as a follow”

“Implemented algorithms segmented the whole CT sequence, while the manually reference standard stopped earlier. For this reason, the distal ends of the segmented vessels were wrong due to lack of standard references. Nonetheless, all deep learning models were able to capture the shape of the aortic valve at fully-opened shape.”

Reviewer #2

We thank the reviewer for his or her valuable consideration about our study. We have taken comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission. We hope that the reviewer agrees. All changes in the text are highlighted in yellow

Reviewer #2: The Authors compared the performance of UNet, ENet, and ERFNet on the task of segmentation of aneurysmal ascending aorta. The novelty of this work is really limited, but it can be treated as work on the application side.

Major issues:
1. It's better to add a figure to compare the structure difference of different networks.
Authors: Figure 1, 2 and 3 were added to shows the different architecture of deep learning models. These figures are related to the text of Sections 2.2, 2.3, and 2.4.

2. In terms of metrics, please show average symmetric surface distance along with Dice, since dice is always affected by the size of the target volume.
Authors: The average symmetric surface distance (ASSD) has been added (see Table 1). Moreover, the formulation to compute the ASSD was added in the text of the Section 2.5.3 as done for other metrics.

3. Section 3 Results, 2nd paragraph "Nevertheless, the computational cost highlighted that the ENet is much faster than UNet." Change the 3X3 filters to 5x5 for UNET can increase the computational cost for it. If this work is mainly about comparing the performance of several existing network, then using the original setup is better. Meanwhile for the task studied in this work, it's unclear whether this filter size increase

	is necessary or not. Authors: It was decided to use 5x5 filters instead of 3x3 in the UNet model based on the past experience of the authors in conducting segmentation tasks for various applications. We tried a version of UNet with all 5x5 convolution filters replaced by 3x3 version on the first training/testing fold. The average DICE score on test fold 1 was 92.09% with a standard deviation of 3.05%. In comparison, the 5x5 version produced an average DICE score of 93.26% with standard deviation of 2.53%. Although the results are close, the 5x5 version produces better results with less variance. In terms of computational complexity, the 3x3 version is obviously smaller. However, even with 3x3 filters, the UNet model has 1,946,338 total trainable parameters compared to ENet with only 362,992 parameters. That's still a fairly significant difference of a factor greater than 5 times. This aspect was clarified with the following text in the Result section: “It is clear that the utilize of 5x5 filters in our UNet architecture implementation is more memory demanding than the 3x3 convolution filter in the original UNet. In terms of DSC, we observed, on the first training/testing fold, that the difference on the utilize of 3x3 filters versus the 5x5 filter were 92.09±3.05% and 93.26±2.53%, respectively. Although the results are close, the 5x5 version produces better results with less variance. In terms of computational complexity, the 3x3 version is obviously smaller. However, even with 3x3 filters, the UNet model has 1,946,338 total trainable parameters compared to ENet with only 362,992 parameters. That's still a fairly significant difference with a factor greater than 5 times.” 4. Please zoom in the target region in fig 2. It's hard to see the difference Authors: The figure has been modified following the reviewer’s suggestion. 5. For the ANOVA, please report the F statistic for the whole one-way ANOVA first, and then do the multiple comparison. Also, the authors should use multiple comparison correction techniques. Authors: Table 2 has been modified to report the F statistic for the whole one-way ANOVA. Table 3 has been added to show multiple comparisons using three different correction techniques (Tukey HSD, Scheffé and Bonferroni /Holm). Minor issues: 1. No page # Authors: Page numbers was added in the manuscript.

Manuscript Number: BMEL-D-20-00089
Article Title: Deep Learning Approach for the Segmentation of Aneurysmal Ascending Aorta Journal Title: Biomedical Engineering Letters

COMMENTS FOR THE EDITOR:

Dear Editor

We thank the Editor for his consideration about our study submitted to BMEL. We have taken reviewer comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission.

Please consider that, because of a mistake during the submission process, there is a mismatch in the author order between the text and the metadata generated by web system. We confirm that the first author is Dr. Albert Comelli, and we will fix this error on author order during the submission of the revised manuscript.

Specific Editor comments

Editor comments #1: Please make sure to address all comments. The potential overlap between training and test sets is a very important issue.
Authors: We thanks the Editor for this point. We clarified in the manuscript that no cross-contamination between training and test sets was performed. As clarified to the Reviewer #1, the set of CT image of the testing was different to that of training set. We modified the text to clarify this critical point.

Editor comments #2: Although statement on IRB approval for this study is provided, the full name of the IRB/ethics committee that approved the study should be provided in the manuscript. Also, declaration on conflict of interest for all enlisted authors should be provided in the manuscript (Giovanni Petrucci conflict is not provided).
Authors: The IRB approval was given by the local committee "Comitato Etico sezionale IRCCS ISMETT" with the number IRRB_04_04. This information was added at the beginning of the Section 2.1 "Study Population". The declaration of conflict of interest was added for Giovanni Petrucci.

Reviewer #1

We thank the reviewer for his or her valuable consideration about our study. We have taken comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission. We hope that the reviewer agrees. All changes in the text are highlighted in yellow

Reviewer #1: Summary: The authors present a comparison of aorta segmentation networks from contrast enhanced ECG-gated CT scans. The reference standard on 72 patients is generated via semi-automated segmentation. Experiments are set up in a 5-fold cross-validation setup. Three networks are compared: Unet, ENet and ERFNet using the Tversky cost function for network optimization. ANOVA shows that the performance of ENet and Unet are comparable and higher than that of the ERFNet, being processing time with ENet less than half of that of the Unet.

Comments:
General comment: this is a 3D segmentation problem solved in a per-axial slice basis. One could easily envision 2.5D or 3D methods. Were they considered?
Authors: We have considered 3D segmentation. However, the computational cost of 3D networks is extremely prohibitive. One approach to deal with this is to down-sample the data. Even with a good graphics card (NVIDIA QUADRO P4000 with 8 GB of RAM), a 128x128x128 input model would be the highest resolution feasible. This amount of down-sampling will seriously exacerbate the class imbalance problem and introduce interpolation artifacts as well along with extremely high computational cost. Another drawback of the 3D approach, especially in our case, is that a 3D network would then mandate expanding the training dataset. We believe, even with data augmentations our rather small dataset of 72 patients would not have been enough to prevent overfitting. Once we saw clinically significant results with only a 2D network and in light of the limitations of a 3D approach we decided to stick with 2D networks. Starting with the small Enet network, 2.5D approaches could certainly be explored in the future. Eventually, we wish to transfer our results to actual clinical use and hardware constraints definitely come into play in such scenarios. Please also consider that our approach follows 2D approach that is consistent with a similar study (Fantazzini et al CVET 2020) that was published after our submission to this journal. A comment on this interesting consideration was added in the "Discussion" section with the following text:

"In this study, 3D segmentation was not adopted as this approach requires larger dataset and is memory-demanding. One approach to deal with 3D segmentation drawbacks is to down-sample the data or adapt 3D integration of 2D convolutional neural networks trained on orthogonal planes to provide a final 3D segmentation. As this study focused on the impact of different deep learning methods on the segmentation accuracy of dilated aortas, the efficacy and accuracy of 3D approach will be investigated in future studies."

Section 2.5.2
- It is unclear how the data is processed. According to line 20, 'we extracted individual slices from all patient cases in the training fold and then used these individual slices as input for our models'. The authors should be very careful by not including slices from the same patient into the training and test sets. Slices from the same patient are highly correlated among them. If slices from the same patient are included into the training and the test set, then the segmentation results are artificially inflated due to such correlation. Please clarify if slices from the same patient appear in each fold's training and validation.

Authors: We are sorry for having confused the Reviewer but it is clear that image slices are different between training and test datasets. The data was divided into 5 folds at the patient level and not at slice level. So there was no cross-contamination between training and test sets. Slices from the same patient were never used for both training and testing the same model. We have clarified this aspect in "Training" Section with the following text:

"In stratified five-fold cross-validation, the data set (72 patients) were divided into 5 equal patient subsets, and the holdout method repeated 5 times. Consequently, for each of the three network models, we trained

each single model 5 times. Each time, one of the 5 subsets was used as the testing set, and the other 4 subsets as training set. Slices from the same patient were never used for both training and testing purpose. So, there was no cross-contamination between training and test sets."

- Data augmentation - horizontal flips. The heart is not symmetrical with respect to the horizontal axis. I am surprised such augmentation plays any effect on the performance.

Authors: In general, data augmentations are used for preventing overfitting by expanding the training dataset. We recognize that the heart is not symmetric with respect to the horizontal axis (or vertical flips) so that flipping of slices will likely produce the same data. However, the fact is that not symmetric data augmentation forces the model to learn to detect the same features in different orientations and this helps to stabilize and regularize the network. Nothing was changed in the manuscript if this is permitted by the Reviewer.

- Cost function (line 49) - how were the parameters \alpha and \beta selected?

Authors: The values of alpha and beta were used as suggested by the work proposed by Salehi and collaborators (see reference 20). They reported that an alpha=0.3 and beta=0.7 lead to the best results. These values are derived from deep learning of multiple sclerosis lesion segmentation. The sentence was rephrased with the following text:

"We also adopted a batch size of 8 slices for all experiments and adopted a Tversky loss function with $\alpha = 0.3$ and $\beta = 0.7$ as reported by Salehi et al. [20]. These values for the loss function demonstrated to be effective for training deep neural networks for sclerosis lesion segmentation."

Figure 3. There are errors are the end of the segmentation of the descending aorta. Is this due to the reference standard? It is also impossible to see false negatives in these images. Maybe consider adding transparency to the reference standard depiction.

Authors: Figure 6 was modified adding transparency to the models. The errors at the end of the segmentation are caused by the boundary of the standard reference segmentation while the deep learning models still recognize this end of the vessel as a region deserving segmentation. Thus, the errors are caused by the reference model rather than the proposed deep learning algorithms. The sentence at the end of "Results" section was re-written as a follow"

"Implemented algorithms segmented the whole CT sequence, while the manually reference standard stopped earlier. For this reason, the distal ends of the segmented vessels were wrong due to lack of standard references. Nonetheless, all deep learning models were able to capture the shape of the aortic valve at fully-opened shape."

Reviewer #2

We thank the reviewer for his or her valuable consideration about our study. We have taken comments into careful consideration when preparing the revised manuscript and feel that the critiques led directly to an improved submission. We hope that the reviewer agrees. All changes in the text are highlighted in yellow

Reviewer #2: The Authors compared the performance of UNet, ENet, and ERFNet on the task of segmentation of aneurysmal ascending aorta. The novelty of this work is really limited, but it can be treated as work on the application side.

Major issues:

1. It's better to add a figure to compare the structure difference of different networks.

Authors: Figure 1, 2 and 3 were added to shows the different architecture of deep learning models. These figures are related to the text of Sections 2.2, 2.3, and 2.4.

2. In terms of metrics, please show average symmetric surface distance along with Dice, since dice is always affected by the size of the target volume.

Authors: The average symmetric surface distance (ASSD) has been added (see Table 1). Moreover, the formulation to compute the ASSD was added in the text of the Section 2.5.3 as done for other metrics.

3. Section 3 Results, 2nd paragraph "Nevertheless, the computational cost highlighted that the ENet is much faster than UNet." Change the 3X3 filters to 5x5 for UNET can increase the computational cost for it. If this work is mainly about comparing the performance of several existing network, then using the original setup is better. Meanwhile for the task studied in this work, it's unclear whether this filter size increase is necessary or not.

Authors: It was decided to use 5x5 filters instead of 3x3 in the UNet model based on the past experience of the authors in conducting segmentation tasks for various applications. We tried a version of UNet with all 5x5 convolution filters replaced by 3x3 version on the first training/testing fold. The average DICE score on test fold 1 was 92.09% with a standard deviation of 3.05%. In comparison, the 5x5 version produced an average DICE score of 93.26% with standard deviation of 2.53%. Although the results are close, the 5x5 version produces better results with less variance. In terms of computational complexity, the 3x3 version is obviously smaller. However, even with 3x3 filters, the UNet model has 1,946,338 total trainable parameters compared to ENet with only 362,992 parameters. That's still a fairly significant difference of a factor greater than 5 times. This aspect was clarified with the following text in the Result section:

“It is clear that the utilize of 5x5 filters in our UNet architecture implementation is more memory demanding than the 3x3 convolution filter in the original UNet. In terms of DSC, we observed, on the first training/testing fold, that the difference on the utilize of 3x3 filters versus the 5x5 filter were 92.09±3.05% and 93.26±2.53%, respectively. Although the results are close, the 5x5 version produces better results with less variance. In terms of computational complexity, the 3x3 version is obviously smaller. However, even with 3x3 filters, the UNet model has 1,946,338 total trainable parameters compared to ENet with only 362,992 parameters. That's still a fairly significant difference with a factor greater than 5 times.”

4. Please zoom in the target region in fig 2. It's hard to see the difference

Authors: The figure has been modified following the reviewer’s suggestion.

5. For the ANOVA, please report the F statistic for the whole one-way ANOVA first, and then do the multiple comparison. Also, the authors should use multiple comparison correction techniques.

Authors: Table 2 has been modified to report the F statistic for the whole one-way ANOVA. Table 3 has been added to show multiple comparisons using three different correction techniques (Tukey HSD, Scheffé and Bonferroni /Holm).

Minor issues:

1. No page #

Authors: Page numbers was added in the manuscript.

Deep Learning Approach for the Segmentation of Aneurysmal Ascending Aorta

Albert Comelli[1,2], Navdeep Dahiya [3], Alessandro Stefano[2], Viviana Benfante[2], Giovanni Gentile[4], Valentina Agnese[5], Giuseppe M Raffa[5], Michele Pilato[5], Anthony Yezzi[3], Giovanni Petrucci[6] and Salvatore Pasta[6]

[1] Ri.MED Foundation, Palermo, Italy

[2] Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Cefalù, Italy

[3] Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta GA 30332, USA

[4] Department of Diagnostic and Therapeutic Services, Radiology Unit, IRCCS-ISMETT, Palermo, Italy

[5] Department for the Treatment and Study of Cardiothoracic Diseases and Cardiothoracic Transplantation, IRCCS-ISMETT, Palermo, Italy

[6] Department of Engineering, University of Palermo, Palermo, Italy

Corresponding author:

Salvatore Pasta, PhD

Professor of Industrial Bioengineering,

Department of Engineering

University of Palermo

office: +39 09123897277

email: salvatore.pasta@unipa.it

Abstract.

Purpose: Diagnosis of ascending thoracic aortic aneurysm (ATAA) is based on the measurement of the maximum aortic diameter, but size is not a good predictor of the risk of adverse events. There is growing interest in the development of novel image-derived risk strategies to improve patient risk management towards a highly individualized level. **Methods:** In this study, the feasibility and efficacy of deep learning for the automatic segmentation of ATAAs was investigated using UNet, ENet, and ERFNet techniques. Specifically, CT angiography done on 72 patients with ATAAs and different valve morphology (ie, tricuspid aortic valve, TAV, and bicuspid aortic valve, BAV) were semi-automatically segmented with Mimics software (Materialize NV, Leuven, Belgium), and then used for training of the tested deep learning models. The segmentation performance in terms of accuracy and time inference were compared using several parameters. **Results:** All deep learning models reported a dice score higher than 88%, suggesting a good agreement between predicted and manual ATAA segmentation. We found that the ENet and UNet are more accurate than ERFNet, with the ENet much faster than UNet. **Conclusions:** This study demonstrated that deep learning models can rapidly segment and quantify the 3D geometry of ATAAs with high accuracy, thereby facilitating the expansion into clinical workflow of personalized approach to the management of patients with ATAAs.

Keywords: Deep Learning; segmentation; Aorta; Aneurysm; Aortic Valve.

1 Introduction

An ascending thoracic aortic aneurysm (ATAA) represents a permanent vessel dilatation of the aorta leading to adverse events and death. Nearly 10 out of 100,000 persons per year are affected by ATAA [1], with the congenital bicuspid aortic valve (BAV) having a reported prevalence of the aortopathy in the range of 20-84% [2]. Indeed, patients with BAV have an 80-fold higher risk of developing an ATAA than the general population with the morphologically-normal tricuspid aortic valve (TAV) [3]. If left untreated, an ATAA can lead to fatal complications such as an aortic rupture or dissection. Elective surgery to avoid aortic complications is indicated when the aortic diameter exceeds 5.5 cm since the yearly risk of dissection or rupture rises from 3 to 7% with aneurysms > 6 cm [4]. Although diagnostic imaging is an essential step to measure the critical aortic diameter of an ATAA, rupture and dissection may occur at aortic size not falling within surgical guidelines so additional metrics not based on size are needed to improve the clinical decision-making process [5]. Manual or semi-automatic segmentation from computerized tomography (CT) or magnetic resonance imaging (MRI) data is usually performed for aortic size evaluation. It is also important to remember that size is not the only important imaging marker; aortic shape matters as well as the loss of the normal "waist" of the aorta at the sinotubular junction.

Deep learning methods are emerging for vascular segmentation and remains a challenging area of research [6-10]. These techniques have shown tremendous success in the last 5 years for image classification and segmentation tasks in various fields, especially for neuroimaging for small vessel segmentation [11]. In general, the deep learning approach requires i) any medical images (e.g. MRI, or CT images), ii) a volume of interest according to the desired classification output, iii) the training and testing of the deep learning algorithm and iv) the validation of semantic segmentation. Deep learning techniques may appear simpler and more flexible than machine learning but require more quantities of labelled data for the training process and are usually more complex and less transparent (the so-called 'black box'). This has limited the widespread adoption of deep learning in

clinical practice. However, several different categories of deep learning models have been proposed for image segmentation with the UNet being the most adopted technique for biomedical imaging analysis [12]. In a different way, efficient neural network (ENet) [13] and the efficient residual factorized convnet (ERFNet) [14] are commonly used in mobile applications where hardware availability is limited and accurate segmentation is very critical.

This study aims to develop a deep learning framework for the segmentation of the aneurysmal aorta and its valve. Specifically, deep learning vessel segmentation was developed using 72 CT scans of patients with ATAAs and different aortic valve morphology (ie, BAV and TAV). Three different deep learning models including the UNet, ENet, and ERFNet were investigated to account for accurate vessel segmentation, fast training time, low hardware requirements for inference, and low training data requirements. Cross-validation strategy was applied for training. The three deep learning models were compared to reveal accurate segmentation of ATAAs with a small sample size.

2 Materials and Method

2.1 Study Population and CT Imaging

After internal review board approval (IRRB_04_04 released by Comitato Etico sezionale IRCCS ISMETT) and informed consent, a total of 72 ATAAs collected from patients underwent electrocardiographic-gated computed tomography (ECG-gated CT) angiography for aortic size evaluation were enrolled. For all patients, ECG-gated CT scans were done after intravenous injection of contrast agent to improve image quality. The CT examination was carried out on a GE VCT 64-channel scanner (GE Medical Systems, Milwaukee, Wisconsin) with gantry rotation velocity of 0.5 m/s and spiral pitch of 0.984. This allowed us to obtain 10-phase ECG-gated thoracic data sets of the entire cardiac cycle with a resolution of 512x512, and slice thickness of 0.625 mm. In this study, we selected the cardiac phase showing the aortic valve at fully opened

shape, which frequently occurs at 50–100 ms after the R peak. Aortic valve morphology (ie, BAV versus TAV) was assessed by an experienced radiologist using images reconstructed parallel to the aortic valve plane.

For training of deep learning networks, semi-automatic thresholding of the contrast-enhanced images followed by manual cropping and morphologic operations was performed to generate 3D virtual masks of whole aorta using Mimics software (Materialize NV, Leuven, Belgium). This was performed by a 10-year experienced user as previously done by our group [15-17]. All 3D segmented masks had size characterized by isotropic voxel size of 1x1x1 mm^3 and matrix resolution of 512x512. Segmentation masks were resampled using nearest neighbour interpolation and converted to binary values with 0 for background and 1. We then implemented UNet (Section 2.2), ENet (Section 2.3) and ERFNet (Section 2.4) networks and Tversky loss using Keras with Tensorflow (Section 2.5) in the open-source mathematical programming language Python (www.python.org).

2.2 UNet Model

Several changes were made to the original UNet architecture to improve segmentation results [12], as showed in Figure 1. All 3x3 convolutions were replaced by larger 5x5 convolution operators. Each convolution was followed by a drop out [18] layer with rate of 10%. Dropout layers help to regularize the network and avoid overfitting. While the original UNet architecture does not use padding when applying convolution operators, we adopted zero padding to ensure that the size of the output feature map is the same as the input size. The original UNet has a 2D size of 32x32 along with 1024 feature maps at the final layer of the contraction path. In a different way, we used an input size of 512x512 with 32 filters on the first contraction path layer, with doubling of feature maps after each max pool and stopping at 256 feature maps and 2D size of 64x64.

2.3 ENet Model

The ENet represents optimized neural network developed for fast inference and high accuracy, which typically occur in augmented reality and automotive [13]. The ENet architecture was based on building blocks of residual networks, with each block consisting of three convolutional layers. These were a 1x1 projection that reduces dimensionality, with a regular convolutional layer and a 1x1 expansion along with batch normalization and surpassing human-level performance. ENet adopted several types of convolutions to build an encoder/decoder style image segmentation network. In some layers, ENet had asymmetric convolutions characterized by separable convolutions with sequence of 5x1 and 1x5 convolutions. The 5x5 convolution had 25 parameters while the corresponding asymmetric convolution had only 10 parameters to reduce the network size. Finally, the ENet used a single initial block in addition to different variations of the bottleneck layer. Figure 2 shows the ENet architecture [13].

2.4 ERFNet Model

Inspired by residual networks and ENet, ERFNet was optimized to improve accuracy and efficiency in image segmentation with respect to ENet [14]. This leads to more accurate segmentations for urban scenario. The basic building block module of ERFNet segmentation network was referred to as a non-bottleneck-1D layer (see Figure 3) and comprised of two sets of factorized (separable or asymmetric) convolutions of size 3x1 followed by the 1x3 with rectified linear unit non-linearity. The input feature map of the main convolution path was added element-wise to the output of the convolution path, which represented the input of the next layer after applying the rectified linear unit non-linearity. Size of ERFNet input was 512x512 while the down-sampler block was similar to that of ENet architecture. This architecture was based on dilated convolutions with different sizes in the Non-bt-1D layers as well as spatial dropout as regularizer.

2.5 Training Methodology

2.5.1 Loss Function

Deep learning methods generally suffer from imbalanced data problems [19]. This problem is common in biomedical image segmentation where the anatomy of interest may be very small compared to the background consisting of connective tissue with a wide range of intensity grey values. In such imbalanced data problems, the network tends to simply predict most voxels as belonging to the background class. Loss functions can be adopted to solve the class imbalance problem and provide a large weight to foreground voxels.

To overcome imbalanced data problem, we adopted Tversky loss function [20] assuming the dice similarity coefficient (DSC) as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{1}$$

where P and G are the set of predicted and ground truth labels, respectively. To make better adjustment of the weights of false positive (FP) and false negative (FN), we adapted a penalty approach as a follow:

$$S(P, G; \alpha\ \beta) = \frac{|P \cap G|}{|P \cap G| + \alpha|P \backslash G| + \beta|G \backslash P|} \tag{2}$$

where α and β control the magnitude of penalties of FPs, and FNs and P\G is the relative complement of G on P. Therefore, the Tversky loss function can be defined as:

$$T(\alpha\ \beta) = \frac{\sum_{i=1}^{N} p_{0i}\ g_{0i}}{\sum_{i=1}^{N} p_{0i}\ g_{0i} + \alpha \sum_{i=1}^{N} p_{0i}\ g_{1i} + \beta \sum_{i=1}^{N} p_{1i}\ g_{0i}} \tag{3}$$

where the output of the final layer of the network (soft-max layer), p_{0i} is the probability of voxel *i* to be part of ATAA wall and p_{1i} is the probability of it belonging to the background. Also, the ground truth training label g_{0i} is 1 for ATAA wall and 0 for everything else (background) and *vice-versa* for

the g_{1i}. By adjusting the parameters, α and β, the trade-off can be controlled between FPs and FNs. Setting $\alpha = \beta = 0.5$ leads to the familiar DSC while setting $\alpha + \beta = 1$ leads to a set of F_β scores, β's larger than 0.5 weight recall higher than precision by placing more emphasis on FNs leading to better segmentation in slices with small foreground area.

2.5.2 Training

In stratified five-fold cross-validation, the data set (72 patients) were divided into 5 equal patient subsets, and the holdout method repeated 5 times. Consequently, for each of the three network models, we trained each single model 5 times. Each time, one of the 5 subsets was used as the testing set, and the other 4 subsets as training set. Slices from the same patient were never used for both training and testing purpose. So, there was no cross-contamination between training and test sets.

Data augmentation was used to train neural network models to reduce overfitting. This was applied by randomly rotating and translating in both x and y directions, and then applying shearing, horizontal flip and zooming to the input training image slices. Additionally, data standardization or normalization was performed as a pre-processing step to prevent the weights from becoming too large and thus avoid numerical instability. For each fold, 2D pixel-wise mean and standard deviation were computed using all training data. Specifically, each patient data was standardized subtracting the mean and dividing by the standard deviation. An initial set of 16 patients for determining the best learning rates for each of three models was used. A learning rate of 0.0001 for ENet model and 0.00001 for ERFNet and UNet models with Adam optimizer were adopted [21]. We also adopted a batch size of 8 slices for all experiments and adopted a Tversky loss function with $\alpha = 0.3$ and $\beta = 0.7$ as reported by Salehi et al. [20]. These values for the loss function demonstrated to be effective for training deep neural networks for sclerosis lesion segmentation.

All models were trained with a maximum of 100 epochs; particularly, an automatic stopping criterion ending the training step when loss decreased upon 10 epochs was implemented.

A high-end HPC system equipped with a GPU (NVIDIA QUADRO P4000 with 8 GB of RAM) was used to train all networks and run inference.

2.5.3 Data Analysis

For each clinical case, sensitivity, positive predictive value (PPV), Dice score (DSC), volume overlap error (VOE), relative volume difference (VD), and average symmetric surface distance (ASSD) were computed to compare the performance of each deep learning network [22, 23].

$$Sensitivity = \frac{TP}{TN + FN} \tag{4}$$

$$PPV = \frac{TP}{TP + FN} \tag{5}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{6}$$

$$VOE = 1 - \frac{TP}{TP + FP + FN} \tag{7}$$

$$VD = \frac{|FN - FP|}{2TP + FP + FN} \tag{8}$$

$$ASSD(X,Y) = \frac{\{ASD(X,Y) + ASD(Y,X)\}}{2} \tag{9}$$

where:

$$ASD(X,Y) = \sum_{x \epsilon X} min_{y \epsilon Y} \frac{d(x,y)}{|X|} \quad (10)$$

Analysis of variance (ANOVA) on the DSC was used to assess statistical differences among network. Statistical significance was considered for $\alpha \leq 0.05$.

3. Results

Table 1 shows the performance of ATAA segmentation computed using the ENet, UNet and ERFNet methods. It can be observed that the ENet had the highest DSC module (91.2±8.9%) as compared to those of both UNet (91.1±10.2%) and ERFNet (88.4±9.9%).

At analysis of variance, the p-value corresponding to the F-statistic of one-way ANOVA was lower than 0.05, suggesting that the one or more treatments were significantly different (see Table 2). This was demonstrated by the statistical difference in the DSC comparison between deep learning methods (see Table 2, and Table 3 where three different multiple comparison correction techniques were used). Nevertheless, the computational cost highlighted that the ENet is much faster than UNet; Table 4 shows the comparison of computational complexity and performance for each model.

Figure 4 shows the profiles of training DSC and Tversky loss function for one fold. Both DSC and Tversky loss profiles indicate that the ENet model converges much faster than both the ERFNet and UNet, with the ENet model reaching a training DSC of nearly 0.9 in less than 15 epochs. This suggests that training could be obtained faster with ENet. The fact that the training loss of UNet model is lower than both the ENet and ERFNet suggests the presence of overfitting, although the number of UNet filters in each layer was reduced.

It is clear that the utilize of 5x5 filters in our UNet architecture implementation is more memory demanding than the 3x3 convolution filter in the original UNet. In terms of DSC, we observed, on the first training/testing fold, that the difference on the utilize of 3x3 filters versus the 5x5 filter

were 92.09±3.05% and 93.26±2.53%, respectively. Although the results are close, the 5x5 version produces better results with less variance. In terms of computational complexity, the 3x3 version is obviously smaller. However, even with 3x3 filters, the UNet model has 1,946,338 total trainable parameters compared to ENet with only 362,992 parameters. That's still a fairly significant difference with a factor greater than 5 times.

Figure 5 displays contours of automatically segmented aorta for different height of the CT axial plane of the aneurysmal aorta. Differences in the capability of all models to segment the whole aorta can be observed. For one representative patient case, Figure 6 highlights segmented dilated aorta and its valve at fully opened shape as obtained by each deep learning models. Implemented algorithms segmented the whole CT sequence, while the manually reference standard stopped earlier. For this reason, the distal ends of the segmented vessels were wrong due to lack of standard references. Nonetheless, all deep learning models were able to capture the shape of the aortic valve at fully-opened shape. The regions of pronounced curvature changes exhibited geometrical changes between predicted and manually-segmented ATAAs.

4. Discussion

In this study, the feasibility and efficacy of three deep learning models for the segmentation of the aneurysmal ascending aorta was assessed accounting for accurate vessel segmentation, fast training time, low hardware requirements for inference, and low training data requirements. Using ECG-gated CT angiography of 72 patients with ATAA and different valve morphologies, all deep learning models were able to accurately segment the dilated aortas when compared to those obtained by manual segmentation. Among tested deep learning models, the key differences are 1) the ENet and UNet result more accurate than ERFNet, with the ENet faster than UNet; 2) the ENet model converges faster than both the ERFNet and UNet. Although validation in large image dataset is warrant, the clinical application of deep learning may revolutionize the way we diagnose the

aneurysmal ascending aorta and open the way towards clinical decision-support system for risk stratification and patient management.

A deep learning model for multiple structures would be consistent and fast to reproduce the same result every time. It is recognized that certain segmentation outputs of our ATAA models revealed lower DSC, but on further inspection, this was due to ground truth annotation error by human readers or artifacts caused by the high heart rate that is common in the bicuspid patient population [2]. Furthermore, the agreement between deep learning predictions and manual segmentations is comparable to that usually reported for the inter- and intra-reader agreements by manual operators [24]. To the best of our knowledge, this is the first study that adopted ENet and ERFNet for the cardiovascular medical imaging analysis. These models are developed for real-time applications and are therefore smaller and faster than the UNet model used in other studies for cardiovascular segmentation [8, 9, 25, 26]. The ENet model has an order of magnitude fewer parameters than both ERFNet and UNet while ERFNet has less than half the number of parameters compared to UNet. Using a fair GPU hardware, we found that the ENet needs only 15.2 s for the ATAA segmentation as compared to the slower 39.1 s shown by the UNet. However, when computations are performed on CPU, the size of the ATAA model has a remarkable impact on the performance of deep learning model, with ENet and UNet, which respectively employ on average about 122.5 s and 1398.2 s to segment the CT data set of a patient with ATAA.

The fact that the proposed deep learning models were able to accurately segment the dilated aortas on the basis of a small training dataset is due to an *ad-hoc* pre-processing that was previously developed by our group [23, 27, 28, 22]. In general, deep learning methods suffer when applied to class imbalanced data and tends to predict most voxels as belonging to a background class. To overcome this issue, we used a custom version of Tversky loss function [20] to provide a larger weight to the target voxels and thus to learn the foreground object representation more effectively.

Moreover, a five-fold cross-validation strategy using 2D slices from all patient cases as model input was used to overcome the limit of the small training dataset while the overfitting was reduced by six different types of data augmentation techniques. In other deep learning studies, the number of data for training was remarkable [26, 8, 9, 6]. With regards to cardiovascular anatomies, Baskaran et al. [9] applied a UNet-inspired deep learning model to segment cardiac structures and great vessels from 206 patients who underwent coronary CT angiography. They obtained an overall median DSC of 0.820. For the abdominal aorta, Roth et al [26] trained a deep learning model using 331 CT scans to obtain a DSC of 0.79. Another study based on convolutional neural networks segmented three parts of the thoracic aorta, with DSC ranging from 0.83 to 0.88 [25]. Using an *ad-hoc* pre-processing strategy, we were able not only to reduce the need for a big training dataset but also improved the segmentation accuracy as the DSC was greater than 0.88 with all deep learning models.

The clinical decision-making process for the management of patients with ATAAs is based on the maximum aortic size normalized by the patient body size index or height. However, size is not a good predictor of aortic rupture or dissection [4]. Phenotypic classification has evinced that ATAAs confined to the aortic root grow differently by aneurysm shapes involving the tubular portion of the ascending aorta, thereby demonstrating the lack of predictive capability of the single aortic size measurement [29]. There is therefore an emerging interest in the development of image-derived strategies to improve ATAA risk definition to highly individualized level [30]. These novel strategies rely on flow analysis computed by *in-vivo* 4D Flow MRI [31], computational predictions based on rupture potential indices of the ATAA wall [32], combination of computational analyses and plasma-based biomarkers [15, 33, 34]. Recently, few research groups have proposed machine learning and statistical shape analysis to investigate the relationship between shape features and numerically predicted risk variables of ATAAs [35, 36]. On the other hand, aortic strain for stiffness-based risk predictions is increasing the interest of many researchers because this metric

can be easily obtained by echocardiographic imaging without the need of *in-silico* simulations and assumptions on ATAA material properties. We recently developed a mathematical algorithm to quantify the full-field aortic strain of the ATAA wall from ECG-gated CT angiography and predicted the aneurysm risk by a stiffness-based parameter [37]. This approach was time-consuming because it was based on manual segmentations of the ATAA wall at ten cardiac phases. The combination of the proposed deep learning models with mathematical algorithms for strain analysis can be easily implemented in the clinical framework to provide stiffness-based risk prediction and tailor personalized approach to ATAA management. Therefore, this study adds another brick towards the implementation of fully-automatic risk strategies for patients with ATAAs.

There are a number of limitations in this work. The number of patients used for training and validation may have limited the accuracy of deep learning models. The different patterns of aortic dilatation (ie, aortic root vs tubular aortic dilatations) and bicuspid phenotypes (ie, anterior vs posterior) may have increased the variability in the investigated CT image dataset. As more patients will be recruited, the training and validation of deep learning models will be re-evaluated by grouping patients according to similar shape features or aortic valve phenotypes. In this study, 3D segmentation was not adopted as this approach requires larger dataset and is memory-demanding. One approach to deal with 3D segmentation drawbacks is to down-sample the data or adapt 3D integration of 2D convolutional neural networks trained on orthogonal planes to provide a final 3D segmentation. As this study focused on the impact of different deep learning methods on the segmentation accuracy of dilated aortas, the efficacy and accuracy of 3D approach will be investigated in future studies. Finally, the accuracy of trained deep learning models is likely confined to our CT scanner, and extension to healthy non-aneurysmal aortas could be not straightforward.

5. Conclusion

This study demonstrated the feasibility and efficacy of deep learning for the segmentation of ATAAs as collected from ECG-gated CT angiography. The tested deep learning models highlighted a good segmentation accuracy with DSC of 88% in all models (ie, UNet, ENet and ERFNet), with differences related to the training time and data requirements. The clinical application of deep learning for automatic vessel segmentation can improve not only the diagnosis of ATAAs but can also improve the management of patients towards personalized risk strategies.

Acknowledgements

This work was partially supported by a grant (W911NF-18-1-0281) from USA Army Research Office (ARO) to Anthony Yezzi, by a grant (R01-HL-143350) from National Institute of Health (NIH) to Anthony Yezzi, and by a grant (GR-2011-02348129) from the Italian Ministry of Health to Salvatore Pasta.

Compliance with Ethical Standards:

Conflict of interest: Albert Comelli declares that he has no conflict of interest. Navdeep Dahiya declares that he has no conflict of interest. Alessandro Stefano declares that he has no conflict of interest. Viviana Benfante declares that she has no conflict of interest. Giovanni Gentile declares that he has no conflict of interest. Valentina Agnese declares that she has no conflict of interest. Giuseppe M Raffa declares that he has no conflict of interest. Michele Pilato declares that he has no conflict of interest. Anthony Yezzi declares that he has no conflict of interest. Giovanni Petrucci declares that he has no conflict of interest. Salvatore Pasta declares that he has no conflict of interest.

Ethical standards: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study. Additional informed consent was obtained from all patients for which identifying information is included in this article.

Informed consent: Informed consent was obtained from all individual participants included in the study

References

1. Davies RR, Goldstein LJ, Coady MA, Tittle SL, Rizzo JA, Kopf GS et al. Yearly rupture or dissection rates for thoracic aortic aneurysms: Simple prediction based on size. The Annals of Thoracic Surgery. 2002;73(1):17-28.

2. Verma S, Siu SC. Aortic dilatation in patients with bicuspid aortic valve. N Engl J Med. 2014;370(20):1920-9. doi:10.1056/NEJMra1207059.

3. Michelena HI, Khanna AD, Mahoney D, Margaryan E, Topilsky Y, Suri RM et al. Incidence of aortic complications in patients with bicuspid aortic valves. JAMA. 2011;306(10):1104-12. doi:10.1001/jama.2011.1286.

4. Pape LA, Tsai TT, Isselbacher EM, Oh JK, O'Gara PT, Evangelista A et al. Aortic diameter >= 5.5 cm is not a good predictor of type A aortic dissection - Observations from the international registry of acute aortic dissection (IRAD). Circulation. 2007;116(10):1120-7. doi:Doi 10.1161/Circulationaha.107.702720.

5. Borger MA, Fedak PWM, Stephens EH, Gleason TG, Girdauskas E, Ikonomidis JS et al. The American Association for Thoracic Surgery consensus guidelines on bicuspid aortic valve-related aortopathy: Full online-only version. J Thorac Cardiovasc Surg. 2018;156(2):e41-e74. doi:10.1016/j.jtcvs.2018.02.115.

6. Mohammadi S, Mohammadi M, Dehlaghi V, Ahmadi A. Automatic Segmentation, Detection, and Diagnosis of Abdominal Aortic Aneurysm (AAA) Using Convolutional Neural Networks and Hough Circles Algorithm. Cardiovasc Eng Technol. 2019;10(3):490-9. doi:10.1007/s13239-019-00421-6.

7. Lopez-Linares K, Aranjuelo N, Kabongo L, Maclair G, Lete N, Ceresa M et al. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks. Med Image Anal. 2018;46:202-14. doi:10.1016/j.media.2018.03.010.

8. Baskaran L, Al'Aref SJ, Maliakal G, Lee BC, Xu Z, Choi JW et al. Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. Plos One. 2020;15(5):e0232573. doi:10.1371/journal.pone.0232573.

9. Baskaran L, Maliakal G, Al'Aref SJ, Singh G, Xu Z, Michalak K et al. Identification and Quantification of Cardiovascular Structures From CCTA: An End-to-End, Rapid, Pixel-Wise, Deep-Learning Method. JACC Cardiovasc Imaging. 2020;13(5):1163-71. doi:10.1016/j.jcmg.2019.08.025.

10. Dahiya N, Yezzi A, Piccinelli M, Garcia E. Integrated 3D Anatomical Model for Automatic Myocardial Segmentation in Cardiac CT Imagery. Computer methods in biomechanics and biomedical engineering Imaging & visualization. 2019;7(5-6):690-706. doi:10.1080/21681163.2019.1583607.

11. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP. Deep Learning in Neuroradiology. AJNR American journal of neuroradiology. 2018;39(10):1776-84. doi:10.3174/ajnr.A5543.

12. Ronneberger O, Fischer P, Broz T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2015.

13. Paszke A, Chaurasia A, Kim S, Culurciello E. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. ArXiv. 2016;1606.02147.

14. Romera E, Alvarez JM, Bergasa LM, Arroyo R. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. IEEE Transactions on Intelligent Transportation Systems. 2018;19(1):263 - 72.

15. Pasta S, Gentile G, Raffa GM, Bellavia D, Chiarello G, Liotta R et al. In Silico Shear and Intramural Stresses are Linked to Aortic Valve Morphology in Dilated Ascending Aorta. Eur J Vasc Endovasc Surg. 2017;S1078-5884(17):30331-3.

16. Rinaudo A, Raffa GM, Scardulla F, Pilato M, Scardulla C, Pasta S. Biomechanical implications of excessive endograft protrusion into the aortic arch after thoracic endovascular repair. Computers in biology and medicine. 2015;66:235-41. doi:10.1016/j.compbiomed.2015.09.011.

17. Pasta S, Gentile G, Raffa GM, Scardulla F, Bellavia D, Luca A et al. Three-dimensional parametric modeling of bicuspid aortopathy and comparison with computational flow predictions. Artif Organs. 2017. doi:10.1111/aor.12866.

18. Srivastava N, Hinto G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting Journal of Machine Learning Research. 2014;15(56):1929−58.

19. Masko D, Hensman P. The Impact of Imbalanced Training Data for Convolutional Neural Networks. 2015.

20. Salehi SS, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017.

21. Kingma D, Ba J. A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations 2015.

22. Comelli A, Stefano A, Russo G, Sabini MG, Ippolito M, Bignardi S et al. A smart and operator independent system to delineate tumours in Positron Emission Tomography scans. Comput Biol Med. 2018;102:1-15. doi:10.1016/j.compbiomed.2018.09.002.

23. Comelli A, Bignardi S, Stefano A, Russo G, Sabini MG, Ippolito M et al. Development of a new fully three-dimensional methodology for tumours delineation in functional images. Comput Biol Med. 2020;120:103701. doi:10.1016/j.compbiomed.2020.103701.

24. Maffei E, Messalli G, Martini C, Nieman K, Catalano O, Rossi A et al. Left and right ventricle assessment with Cardiac CT: validation study vs. Cardiac MR. Eur Radiol. 2012;22(5):1041-9. doi:10.1007/s00330-011-2345-6.

25. Ecabert O, Peters J, Walker MJ, Ivanc T, Lorenz C, von Berg J et al. Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. Med Image Anal. 2011;15(6):863-76. doi:10.1016/j.media.2011.06.004.

26. Roth HR, Oda H, Zhou X, Shimizu N, Yang Y, Hayashi Y et al. An application of cascaded 3D fully convolutional networks for medical image segmentation. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society. 2018;66:90-9. doi:10.1016/j.compmedimag.2018.03.001.

27. Comelli A, Stefano A, Bignardi S, Russo G, Sabini MG, Ippolito M et al. Active contour algorithm with discriminant analysis for delineating tumors in positron emission tomography. Artificial intelligence in medicine. 2019;94:67-78. doi:10.1016/j.artmed.2019.01.002.

28. Comelli A, Stefano A, Russo G, Bignardi S, Sabini MG, Petrucci G et al. K-nearest neighbor driving active contours to delineate biological tumor volumes. Engineering Applications of Artificial Intelligence. 2019;81:133-44.

29. Della Corte A, Bancone C, Dialetto G, Covino FE, Manduca S, Montibello MV et al. The ascending aorta with bicuspid aortic valve: a phenotypic classification with potential prognostic significance. Eur J Cardiothorac Surg. 2014;46(2):240-7; discussion 7. doi:10.1093/ejcts/ezt621.

30. Cosentino F, Scardulla F, D'Acquisto L, Agnese V, Gentile G, Raffa G et al. Computational modeling of bicuspid aortopathy: Towards personalized risk strategies. Journal of molecular and cellular cardiology. 2019;131:122-31. doi:10.1016/j.yjmcc.2019.04.026.

31. Bieging ET, Frydrychowicz A, Wentland A, Landgraf BR, Johnson KM, Wieben O et al. In Vivo Three-Dimensional MR Wall Shear Stress Estimation in Ascending Aortic Dilatation. Journal of Magnetic Resonance Imaging. 2011;33(3):589-97. doi:Doi 10.1002/Jmri.22485.

32. Farzaneh S, Trabelsi O, Avril S. Inverse identification of local stiffness across ascending thoracic aortic aneurysms. Biomech Model Mechanobiol. 2018. doi:10.1007/s10237-018-1073-0.

33. Gallo A, Agnese V, Coronnello C, Raffa GM, Bellavia D, Conaldi PG et al. On the prospect of serum exosomal miRNA profiling and protein biomarkers for the diagnosis of ascending aortic dilatation in patients with bicuspid and tricuspid aortic valve. International journal of cardiology. 2018;273:230-6. doi:10.1016/j.ijcard.2018.10.005.

34. Pasta S, Agnese V, Gallo A, Cosentino F, Di Giuseppe M, Gentile G et al. Shear Stress and Aortic Strain Associations with Biomarkers of Ascending Thoracic Aortic Aneurysm. The Annals of thoracic surgery. 2020. doi:10.1016/j.athoracsur.2020.03.017.

35. Liang L, Liu M, Martin C, Elefteriades JA, Sun W. A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm. Biomech Model Mechanobiol. 2017;16(5):1519-33. doi:10.1007/s10237-017-0903-9.

36. Cosentino F, Raffa GM, Gentile G, Agnese V, Bellavia D, Pilato M et al. Statistical Shape Analysis of Ascending Thoracic Aortic Aneurysm: Correlation between Shape and Biomechanical Descriptors. Journal of personalized medicine. 2020;10(2). doi:10.3390/jpm10020028.

37. Pasta S, Agnese V, Di Giuseppe M, Gentile G, Raffa GM, Bellavia D et al. In Vivo Strain Analysis of Dilated Ascending Thoracic Aorta by ECG-Gated CT Angiographic Imaging. Ann Biomed Eng. 2017. doi:10.1007/s10439-017-1915-4.

Figure Legends

Figure 1: Comparison between the original UNet architecture (a) [12] and our UNET implementation (b). Each blue box is a multi-channel feature map with the number of channels denoted at the top of the box. The x-y size is denoted at the bottom lower left edge of the box. White boxes represent copied low-resolution features.

Figure 2: ENet architecture [13]. (a) ENet initial block with 2x2 max pooling with a stride of 2 and convolution has 15 filters, summing to 16 feature maps after concatenation. (b) ENet bottleneck module. 'conv' is either a regular, dilated, or full convolution (deconvolution) with 3x3 filters, or a 5x5 convolution decomposed into two asymmetric (separable) ones.

Figure 3: Basic building block layer of ERFNet network called Non-bottleneck-1D (Non-bt-1D) [14].

Figure 4: Plot the training DSC and loss function Tversky loss for each of three models for one particular fold.

Figure 5: Comparison of segmentation performance for the three architectures in 8 different slices. The manual segmentation (yellow), ENet (red), ERFNet (blu) and U-Net (green) are superimposed.

Figure 6: Comparison of 3D segmentation of prostate using the three Net architectures. The manual segmentation (yellow), ENet (red), ERFNet (blu) and U-Net (black) are superimposed.

Table 1: Performance segmentation using the ENet, UNet and ERFNet methods.

	Sensitivity	PPV	DSC	VOE	VD	ASSD
ENet						
Mean	92.69%	90.67%	91.22%	15.22%	2.83%	4.46
± std	11.13%	9.49%	8.97%	11.62%	14.47%	4.54
± CI (95%)	2.57%	2.19%	2.07%	2.68%	3.34%	1.05
UNet						
Mean	91.63%	91.79%	91.09%	15.30%	0.12%	5.48
± std	12.14%	7.21%	10.18%	11.76%	16.09%	4.67
± CI (95%)	2.80%	1.66%	2.35%	2.72%	3.72%	1.08
ERFNet						
Mean	89.01%	88.94%	88.41%	19.56%	0.92%	5.48
± std	12.46%	10.33%	9.94%	13.88%	16.87%	4.67
± CI (95%)	2.88%	2.39%	2.30%	3.21%	3.90%	1.08

Table 2: ANOVA on the DSC showed statistical differences between segmentation methods.

ANOVA	F value	F critic value	P-value
ENet vs ERFNet vs UNet	3.667	3.038	0.027
ENet vs ERFNet	5.520	3.907	0.020
ERFNet vs UNet	4.474	3.907	0.036
ENet vs UNet	1.270	3.907	0.261

Table 3: Tukey HSD, Scheffé and Bonferroni/Holm tests (post-hoc tests) were used as multiple comparison correction techniques.

Tukey HSD	**Q-statistic**	**P-value**
ENet vs ERFNet	2.451	0.019
ERFNet vs UNet	2.3413	0.022
ENet vs UNet	0.109	0.189
Scheffé	**T-statistic**	**P-value**
ENet vs ERFNet	1.733	0.022
ERFNet vs UNet	1.656	0.026
ENet vs UNet	0.077	0.199
Bonferroni/Holm	**T-statistic**	**P-value**
ENet vs ERFNet	1.733	0.025
ERFNet vs UNet	1.656	0.030
ENet vs UNet	0.077	0.282

Table 4: Comparison of computational complexity and performance of the three models.

Model Name	**Number of Parameters**		**Size on disk [MB]**	**Inference Times [s]**	
	Trainable	Non-Trainable		CPU	GPU
Enet	362992	8352	5.8	122.56	15.23
ERFNet	2056440	0	25.3	157.53	16.64
Unet	5403874	0	65.0	1398.23	39.11

Fig. 1

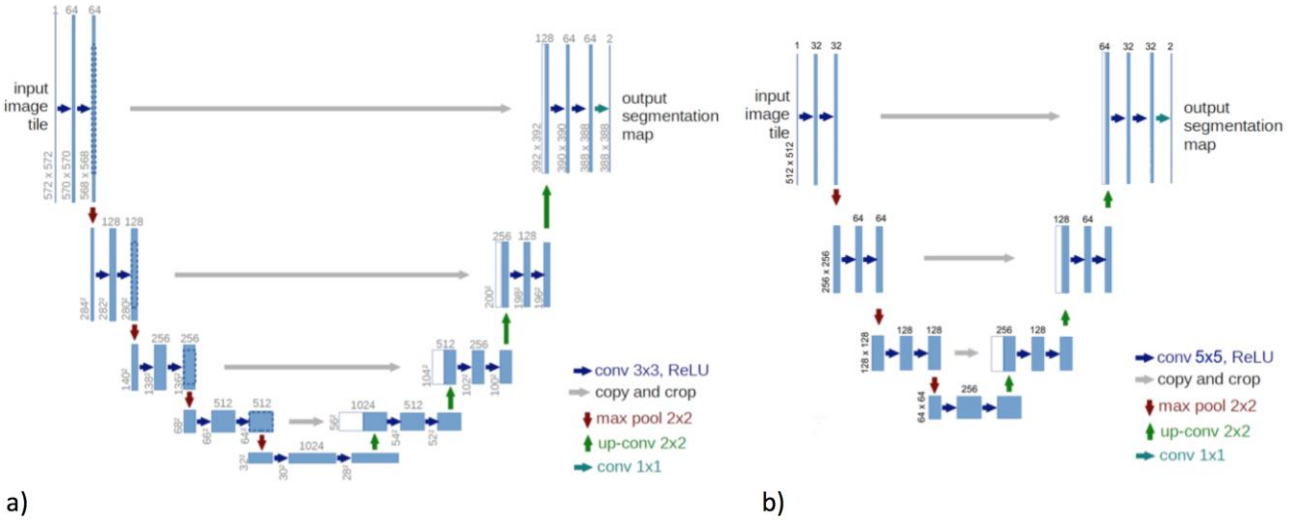
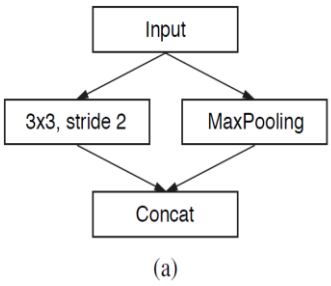
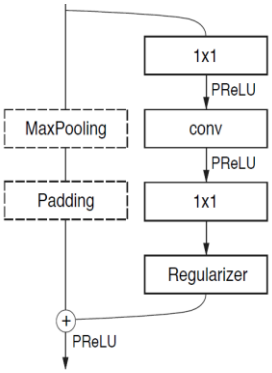

input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1
a)
input image tile
output segmentation map
conv 5x5, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1
b)

Fig. 2


Input
3x3, stride 2
MaxPooling
Concat

(a)


1x1
PReLU
MaxPooling
conv
PReLU
Padding
1x1
Regularizer
PReLU

(b)

Fig. 3

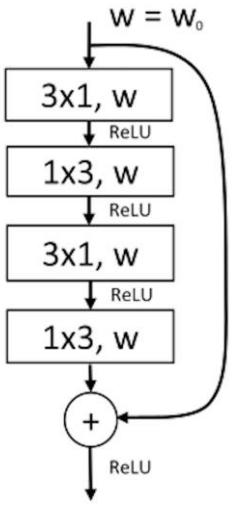


Fig. 4

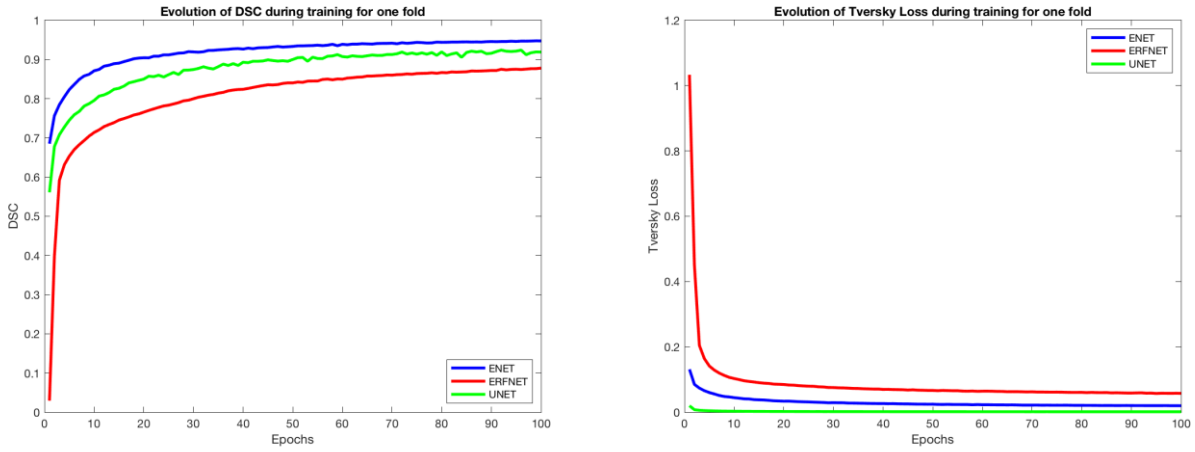

Evolution of DSC during training for one fold
DSC
Epochs
ENET
ERFNET
UNET
Evolution of Tversky Loss during training for one fold
Tversky Loss
Epochs
ENET
ERFNET
UNET

Fig. 5

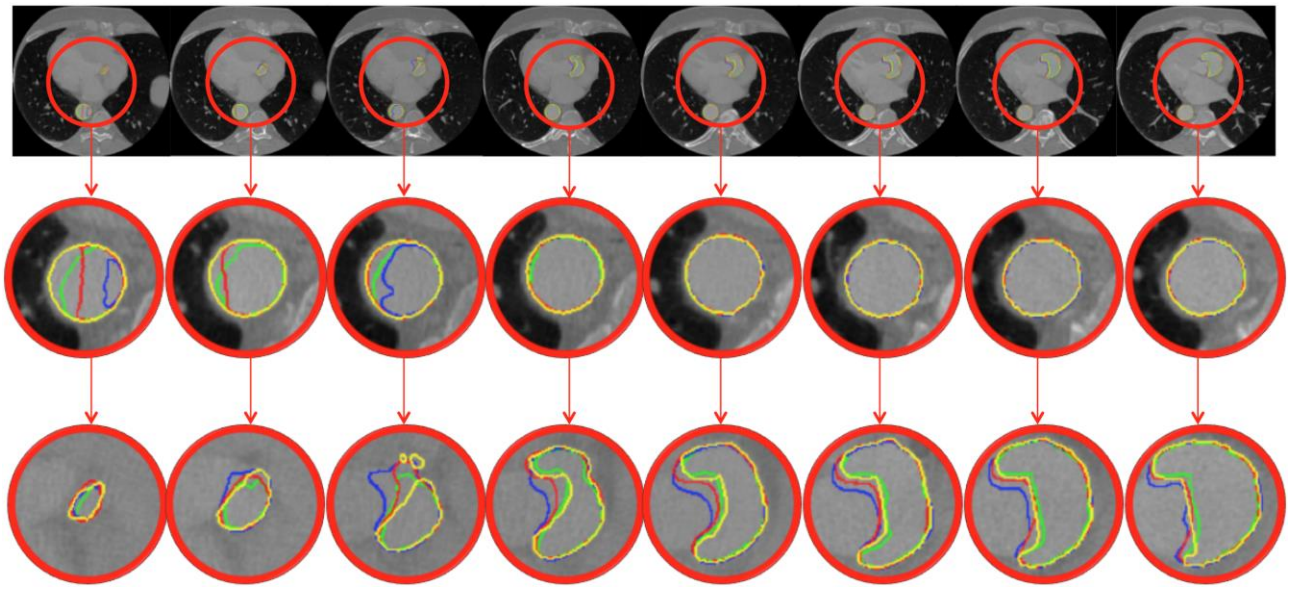


Fig. 6