3D Inception U-Net for Aorta Segmentation using Computed Tomography Cardiac Angiography

Savitha Rani Ravichandran[1,2], Balaji Nataraj[1,2], Su Huang[2], Zhiliang Qin[2], Zhongkang Lu[2], Akie Katsuki[3], Weimin Huang[2]*, Zeng Zeng[2]

[1]*National University of Singapore,* [2]*Institute for Infocomm Research, A*STAR, Singapore,* [3]*GE Healthcare Japan Corporation, Japan*

***Abstract*— Computed Tomography Coronary Angiography (CTCA) is an effective imaging technique used for diagnosis and surgical planning. Segmentation of the aorta from the CTCA can be used clinically for interpretation, aortic valve measurement for intervention and identification of coronary structures. The process of segmentation done manually is tedious and time consuming. In this paper we propose an approach to automatic aorta segmentation from cardiac CTCA scans using deep learning. A dataset of 20 pairs of CT and mask is used, with the manually segmented masks as ground truth. The dataset is split into a training set of 10 pairs, validation set of 3 pairs and test set of 7 pairs. The proposed framework uses two different U-Net models trained for location and refined segmentation tasks. The U-Net architecture and multiple variants of the U-Net are adopted, and their performances are compared. The paper also explores different types of data preprocessing and augmentation that can be performed to improve the performance of the deep learning model. The results show that 3D Inception U-Net performs the best in localization and segmentation tasks. The Localization test DSC is 0.77 and Segmentation test DSC is 0.81. Qualitative testing shows that predicted masks are quite similar to the original.**

Keywords—Cardiac CT scan, Aorta Segmentation, Inception U-Net, 3D deep learning, medical imaging

I. INTRODUCTION

Computed Tomography Cardiac Angiography (CTCA) is widely used for both diagnostic and surgical procedures [1]. However, to get accurate measurement of quantities relating the valves, it is required to segment the aorta from the rest of the cardiac CT. Though it is possible to do it manually, the process is tedious and time consuming. Therefore, automatic methods of cardiac aorta segmentation need to be used. The segmentation method needs to be accurate and robust to work in any setting, clinical or surgical.

Many automatic aorta segmentation methods have been proposed in the past. Zheng et al. [2] use a part-based aorta model by splitting the aorta into four parts: aorta root, ascending aorta, aortic arch and descending aorta. Marginal space learning (MSL) is used to detect the aorta root and the aortic arch. A 2D circle detector using haar wavelet features and boosting learning is used to detect aortic circles. Kitasaka et al. [3] use shape features of the blood vessels and model fitting technique to segment the aorta and other blood vessels. The edge voxels are detected using the standard deviation of the CT. Then Euclidean transformation is applied on non-edge voxels. Isgum et al. [4] use multiple manually labelled target images known as atlases to segment the aorta in CT scans. The segmentation is performed based on a fusion of locally determined decision weights. Duquette et al. [5] modify the interactive graph cut methodology to the problem.

The use of U-Nets has been popular for segmentation tasks in the medical domain. The original 3D U-Net [7] is used for segmenting Xenopus kidney. Li et al [15] use a modified version of the 3D U-Net called the H-Dense U-Net for automatic liver and tumour segmentations. Xiao et al [16] show that weighted residual U-Net is effective at high quality retina vessel segmentation. These methods show that deep learning techniques such as U-Nets is effective for segmentation.

We propose the use of deep learning techniques to automatically segment the aorta from CTCA. In our approach, we use two different U-Net models, one for localization of aorta and another for the refined segmentation. We compare the segmentation results produced by 3D U-Net, 3D Inception U-Net, 3D Attention U-Net and 3D Residual U-Net. Given limited labelled data, we train 3D U-Nets by only using 10 pairs of CT and masks with data augmentation, which still yields encouraging results.

The rest of the paper is organized as follows. Section II describes the methodology and the data preparation methods. Section III describes the experimental results and Section IV concludes the paper.

II. METHODOLOGY

In our approach, to obtain the best performing model for the segmentation task, we implement different models with different pre-processing and augmentation.

A. Dataset

CTCA dataset with manual labels is collected to develop and validate deep learning algorithms. The CTCA was acquired using CT scanners from GE healthcare. The data set has 50 Patients and each CT scan comes with resolution in 0.39x0.39x0.62 mm. However only 20 sets contain a manually labelled aorta part, of which 10 were used for training, 3 for validation and 7 for testing. We have used ITK-SNAP tool [17] for visualizing the CT images and results.

B. Data Pre-processing

The original data contains CTs and masks in the shape (512, 512, L) where L is the slice number that is variable depending on the scan. Since it is variable and too large to use for training the model directly, it has to be processed to reduce the size and keep the size uniform across all data. It was found that the size (64, 64, 64) was the maximum size that could be comfortably used to balance memory, speed and accuracy by the GPUs used for training. Two different pre-processing methods were implemented here. One was to resize the

The work is partially supported by IAF grant 152 187 0013/IAF311014M., A*Star Singapore.

*Corresponding author: wmhuang@i2r.a-star.edu.sg

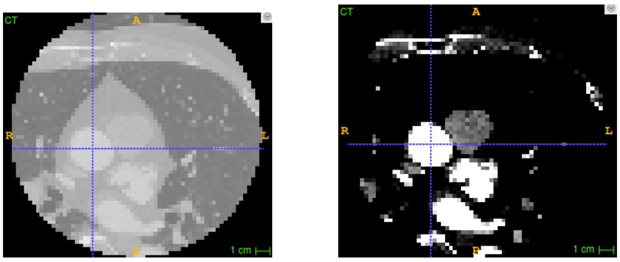
original image to a constant shape of (64, 64, 64) using continuous interpolation. Apart from reducing the size, we also employed histogram normalization. From the examination of multiple CTs, it was found that the cardiac aorta is consistently visible when viewed at intensity values between 50 and 350. Using 50 as the lower threshold and 350 as the higher threshold, the histogram of a CT is normalized. The process is defined by:

$$CT[i] = \begin{cases} \mathrm{lt}, if\ CT[i] < lt \\ ut, if\ CT[i] > ut \\ \dfrac{(CT[i] - lt) * (ut - lt)}{ut - lt} + lt, otherwise \end{cases}$$

where, lt = 50, ut = 350 and CT[i] is the current CT voxel being processed.

This considerably improves the visibility of the aorta and distinguishes it from the other parts of the heart. Fig 1 shows a comparison before and after the thresholding and histogram stretching processes.

To prepare the CTs to be processed by the U-Net, the data was further normalized to have zero mean and unit standard deviation.



(a) Before (b) After

Fig 1. Preprocessing of CT slices

C. Data Augmentation

The dataset used to train the deep learning system only contains 10 pairs of CT and mask. However, to avoid overtraining and improve generalization in a deep learning model, it requires a large dataset [13]. In our approach, we augment the CTs and masks by rotation by a small angle in all three dimensions. For each pair of data during training, we give a 50% chance that rotation will be applied on it. If rotation is applied, three random angles between -8 degrees and 8 degrees are chosen at each of the three dimensions and the CTs and masks are rotated accordingly. Augmentation through this approach assisted in improving the validation accuracy by a significant factor.

D. Deep Learning Models

The approach can be divided into two tasks: localization and segmentation. The localization involves the use of a model to predict an approximate mask of the aorta. The predicted mask is then used to crop the original CTs and masks to provide smaller volume that mostly consist of the aorta in a higher resolution. The training of the localization model is done using the resized data of size (64, 64, 64) from (512,512, L). The data go through the pre-processing and augmentation processes before training. The masks predicted by this model are poor in accuracy due to the large resizing factor of 1:8, which causes a large loss in resolution. To find the rough aorta volume to crop, the coordinates of all activated voxels in the predicted mask from the localization model are checked. The minimum activated coordinate (x_l, y_l, z_l) in each dimension and maximum activated coordinate (x_h, y_h, z_h) in each dimension are located. These two coordinates represent the cuboid containing the whole aorta. To obtain a cubic crop, values are added on either side of the two shorter dimensions. The values are calculated as follows:

$$(x_{diff}, y_{diff}, z_{diff}) = (x_h - x_l, y_h - y_l, z_h - z_l)$$

This gives the length of each edge in the cuboid.

$$(k_x, k_y, k_z) = \max(x_{diff}, y_{diff}, z_{diff}) - (x_{diff}, y_{diff}, z_{diff})$$

Here, (k_x, k_y, k_z) is the extra length to be added to each dimension. Half of these lengths are added on either side of each edge to keep the aorta centred. In case one of the values is odd, the values added to each side are skewed by one voxel. Apart from this, another extra value of 15 voxels is added on both sides of each dimension to give the final coordinates as follows:

$$Lower = (x_l - \frac{k_x}{2} - 15, y_l - \frac{k_y}{2} - 15, z_l - \frac{k_z}{2} - 15),$$

$$Higher = (x_l + \frac{k_x}{2} + 15, y_h + \frac{k_z}{2} + 15, z_h + \frac{k_z}{2} + 15).$$

These coordinates are rescaled from (64, 64, 64) to (512, 512, L) to determine the cropping coordinates from original CTs and masks. In many cases, the z-dimension, which is smaller than the other two dimensions, cannot be cropped to the required size as it is too small. In this case, the crop is limited to the edge of the CT and a cuboid is cropped which is converted to a cube later by the resizing operation.

Once the original CTs and masks are cropped, the cropped data is resized to (64, 64, 64) to be used to train a model for refined segmentation. The resized cube is pre-processed, and data augmentation is performed. Then the segmentation model is trained to provide the accurate mask for the aorta. The complete process of the approach is shown in Fig 2. The figure also shows how the CTs and masks change with each process.

To find the best U-Net architecture for the segmentation task, different variations of the 3D U-Net were tried. The variants of the U-Net are implemented using Residual nets [8], Attention nets [9] or the Inception nets [10]. These variants of the networks have been shown to work well in 2D segmentation cases, so this work studies their effectiveness in 3D segmentation.

The original U-Net was proposed by Ronneberger et al. [13] for the application of biomedical image segmentation. The 3D U-Net as proposed by Cicek et al [7] has been used here as the basic network to perform the aorta segmentation.

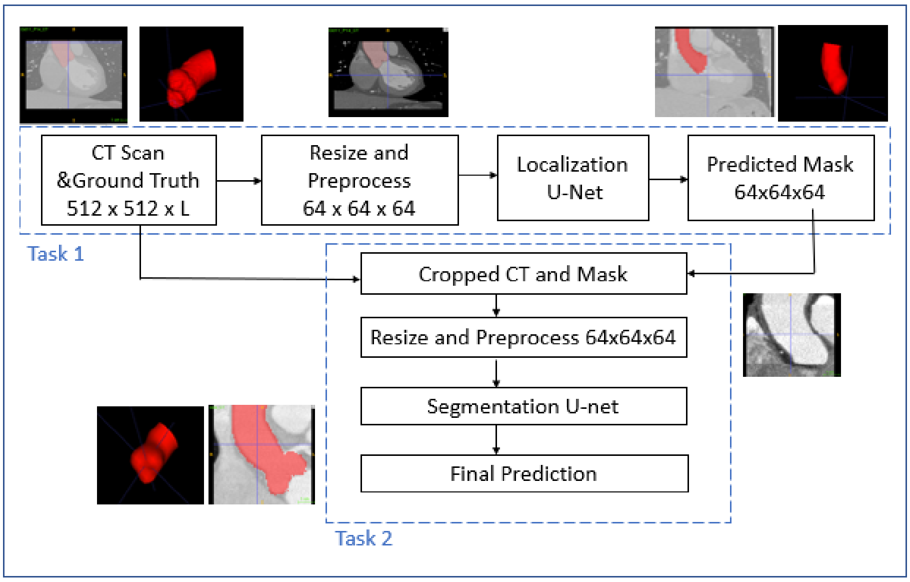


Fig 2. Process flow of the proposed approach consisting of two tasks: Cropping and Segmentation

The residual U-Net [8] is a variant of the U-Net that includes a residual unit with identity mapping instead of just the convolution and Relu layers that are usually present in a U-Net. This increases the total number of parameters to approximately 26 million. The residual blocks allow information propagation without degradation and ease the training of the network. A 3D version of the residual U-Net has been implemented for aorta segmentation

The attention U-Net [9] includes attention gates that filter the features propagated through the U-Net skip connections. This facilitates feature selectivity so that only the important features are propagated.

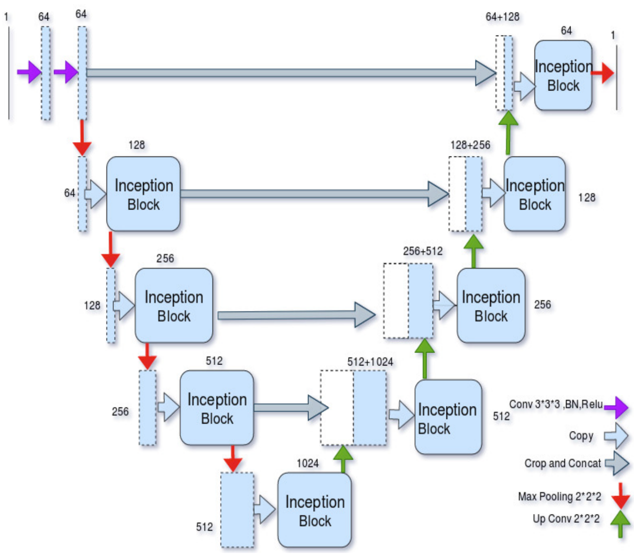


Fig 3. Block Diagram of Inception U-Net.

The Inception U-Net includes the Inception blocks [10] after the convolutions in the original U-Net. This increases the total convolutions to 42, keeps the number of max pools and up convolutions constant, and decreases the total number of parameters to approximately 2 million. The use of these blocks with the U-Net prevents overfitting and improves generalization. The Inception U-Net is shown in Fig 3. A 3D version of the Inception U-Net was implemented for our task.

The four different models given were used for the segmentation task and their performances were compared using the Dice Coefficient (DSC) [18]. The Dice ranges between 0 to 1 where 1 represents perfect match with ground truth.

III. Experiments and Results

Experiments were done for both the localization and the segmentation to determine the best configuration of the processes. For the localization and segmentation, four different architectures were trained: 3D U-Net, 3D Residual U-Net, 3D Attention U-Net and 3D Inception U-Net. The results of the localization models are given in Table 1. It is seen that the 3D Inception U-Net performs better than all other U-Nets, so it was chosen as the localization model.

TABLE I. DSC Comparison of Models Used for Cropping Task

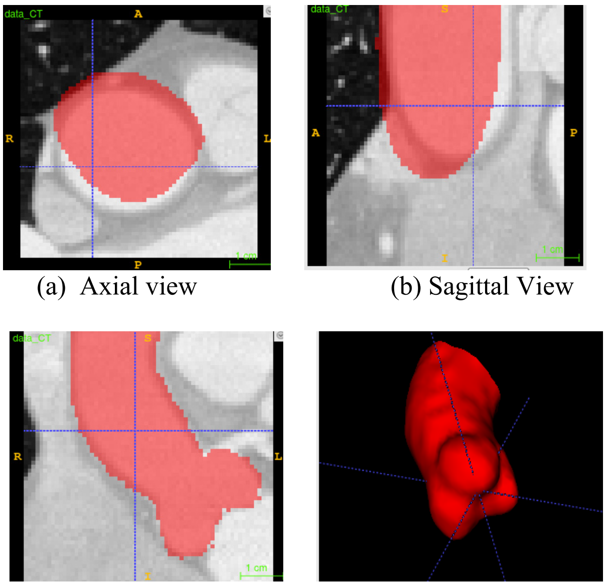
Model	Training Dice coefficient 64×64×64	Validation Dice coefficient 64×64×64	Testing Dice Coefficient 64*64*64
3D U-Net	0.8077	0.7362	0.7152
ResidualU-Net	0.7963	0.7612	0.7553
AttentionU-Net	0.7727	0.7493	0.7321
Inception U-Net	0.8263	0.7859	0.7759

The results of the Segmentation models are given in Table 2. The 3D Inception U-Net again shows the best performance and is chosen as the segmentation model.

The training configuration used for the models of localization and segmentation are similar. The models are run for maximum of 150 Epochs. The learning rate is set to reduce by half every 10 epochs and when the validation loss does not improve for 20 epochs. The optimizer used is Adam with batch size of nine. The training was done on three NVIDIA TITAN V GPUs.

TABLE II. DSC COMPARISON OF MODELS USED FOR SEGMENTATION TASK

Model	Training Dice coefficient 64×64×64	Validation Dice coefficient 64×64×64	Testing Dice Coefficient 64*64*64
3D U-Net	0.8586	0.7490	0.7378
ResidualU-Net	0.8376	0.7804	0.7751
AttentionU-Net	0.7804	0.7551	0.7403
Inception U-Net	0.8891	0.8376	0.8106



(a) Axial view (b) Sagittal View

(c) Coronal View (d) 3D Volume

Fig 4. The mask predicted by the segmentation task shown from the three different dimensions and as 3D Volume

The results showed that using the 3D Inception U-Net for both tasks resulted in the best DSC. So, using this configuration, predictions are obtained for the testing data. These predicted masks obtained as the output of the segmentation are assessed quantitatively and qualitatively. Importance is given to the shape of the predicted mask and the clarity in defining the aortic valve. The predicted mask for one test case is shown in Fig 4. The figures show that the predicted mask resembles the ground truth mask to a good extent. The predicted mask is also able to segment the three aortic valves considerably well.

IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed an approach for accurate and automatic segmentation of aorta from 3D CTCA. It is shown that a two-step process consisting of a location and segmentation can provide good results by maintaining the resolution of the training images for segmentation. It is also shown that the 3D Inception U-Net performs the best for both the localization and segmentation. It achieves the best DSC among all models. Qualitative assessment shows that the predicted masks are quite similar to original mask in shape and position. We believe that the models could be improved by increasing the training data to improve generalization. Clinically an accurate 3D aorta segmentation can be used to analyse the function of aortic valves, for valve implantation and help automatically identify the coronary arteries.

REFERENCES

[1] Stagnaro N, Della Latta D, Chiappino D, “Diagnostic accuracy of MDCT coronary angiography in patients referred for heart valve surgery,” Radiol Med. 2009 Aug;114(5):728-42.

[2] Zheng, Yefeng, et al. "Automatic aorta segmentation and valve landmark detection in C-arm CT: application to aortic valve implantation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg, 2010.

[3] Kitasaka, Takayuki, et al. "Automated extraction of aorta and pulmonary artery in mediastinum from 3D chest X-ray CT imageswithout contrast medium." Medical Imaging 2002: Image Processing. Vol. 4684. International Society for Optics and Photonics, 2002.

[4] Isgum, Ivana, et al. "Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans." IEEE transactions on medical imaging 28.7 (2009): 1000-1010.

[5] Duquette, Anthony Adam, et al. "3D segmentation of abdominal aorta from CT-scan and MR images." Computerized Medical Imaging and Graphics 36.4 (2012): 294-303.

[6] F. Rousseau, P. Habas, and C. Studholme, “A supervised patch-based approach for human brain labeling,” IEEE Trans. Med. Imaging 30(10), 1852–1862 (2011).

[7] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation.arXiv preprint arXiv:1606.06650

[8] Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. "Road Extraction by Deep Residual U-Net."arXiv preprint arXiv:1711.10684(2017)

[9] O. Oktayet al., "Attention U-Net: Learning Where to Look for the Pancreas,"arXiv preprint arXiv:1804.03999,2018

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,V. Vanhoucke,A.Rabinovich et al. “Going deeper with convolutions.”Cvpr, 2015.

[11] Deng, Xiang, et al. "Graph cut based automatic aorta segmentation with an adaptive smoothness constraint in 3D abdominal CT images." Neurocomputing (2018).

[12] Cha, Jungwon, et al. "3D segmentation of the ascending and descending aorta from CT data via graph-cuts." Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. Vol. 10578. International Society for Optics and Photonics, 2018.

[13] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[14] Bui, Vy & Hsu, Li-Yueh & Chang, Lin-Ching & Y Chen, Marcus. (2018). An Automatic Random Walk Based Method for 3D Segmentation of the Heart in Cardiac Computed Tomography Images. 10.1109/ISBI.2018.8363822.

[15] Li, Xiaomeng, et al. "H-DenseU-Net: Hybrid Densely Connected U-Net for Liver and Tumor Segmentation from CT Volumes." IEEE Transactions on Medical Imaging (2018).

[16] Xiao, Xiao, et al. "Weighted Res-U-Net for High-Quality Retina Vessel Segmentation." 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2018.

[17] 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.

[18] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016.