Octahedron-shaped Convolution for Refining Aorta Semantic Segmentation

Xi Xiang, Gongning Luo, Pengfei Zhao, Wei Wang and Kuanquan Wang

June 30, 2021

Octahedron-shaped Convolution for Refining Aorta Semantic Segmentation

1st Xi Xiang
Faculty of Computing
Harbin Institute of Technology
Harbin, China
19S103133@stu.hit.edu.cn

2ndGongning Luo
Faculty of Computing
Harbin Institute of Technology
Harbin, China
992087795@qq.com

3rd Pengfei Zhao
Faculty of Computing
Harbin Institute of Technology
Harbin, China
19S103186@stu.hit.edu.cn

4th Wei Wang
School of Computer Science and Technology
Harbin Institute of Technology
Shenzhen, China
wangwei2019@hit.edu.cn

5th Kuanquan Wang
Faculty of Computing
Harbin Institute of Technology
Harbin, China
wangkq@hit.edu.cn

***Abstract*—Refining 3D aorta segmentation is essential for clinical aorta analysis. The small tubular diameter of the aorta branches and the discontinuity of neighbouring information make it difficult to get a continuous semantic segmentation map. In this paper, we proposed a novel adaptive octahedron-shaped convolution (AOSC) based on VNet and signed distance map(SDM). AOSC aimed to aggregate more contextual information for each sample point in the aortic branches with smaller tubular diameters. The weights of feature fusion introduced SDM as auxiliary information to measure the similarity of neighbouring points. Furthermore, we embedded the learned 3D offset field into AOSC to avoid inaccurate segmentation on the region around the narrow tubular structures. The AOSC module prolonged the predicted length of small aorta branches and then improved the tubular continuity of the aorta segmentation map. We evaluated the AOSC module on our-collected dataset and MICCAI ASOCA2020 coronary artery dataset. Our method achieved the state-of-the-art results in terms of Dice and Jaccard metrics. The code will be available at this link(******).**

***Index Terms*—aorta segmentation, tubular diameter, aorta branches, contextual information, tubular continuity.**

I. Introduction

In Computed Tomography(CT) images, the diameters of aorta vary greatly in different positions. Different points in CT images contain different semantic information. Consequently, it is challenging to reduce the discontinuities around the small tubular branches region.

As shown in Fig.1, different from the aorta trunk, aorta branches with small tubular diameters extend to various directions in 3D space. The end of the aortic branches are almost surrounded by a large number of background points, so there are interferences around. Moreover, in a certain CT image, the gray values of the foreground points and their neighbouring background points are very similar. So it increases the difficulty of the network to classify points.

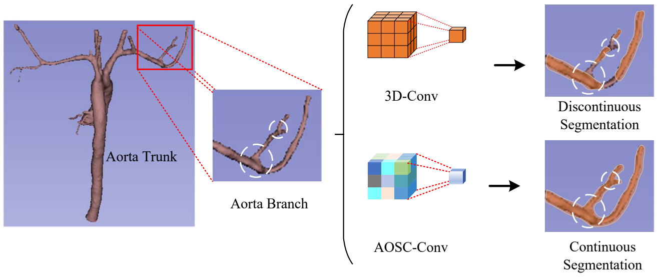

Fig. 1. At the aorta branches with small tubular diameter, inadequate contextual information increases the discontinuities of segmentation.

Current methods for refining cardiac substructure segmentation, such as aorta, left atrium and so on, can be categorized into two classes: shape-aware anatomical structure based methods and local fine-tuning based methods. The **shape-aware anatomical structure** based methods focus on modelling the shape of the target organ. In recent years, SDM [1]–[5] describes the relationship between spatial structure and distance field. After a Heaviside step function [6], the calculated SDM can be easily converted into a binary segmentation map, but the threshold of the Heaviside function is a hyperparameter that is difficult to choose. Moreover, the confidence of a single point in a segmentation map is relatively smaller than the average confidence of grouped points. The **local fine-tuning** based methods are inclined to rectify segmentation map at the boundary and edge. Cheng et al. [7] define a Direction Field to exploit the directional relationship between points. Chu et al. [8] propose to learn an edge detector to locate the discontinuity and add additional supervision on these areas. However, due to the complexity of the boundary information, fine-tuning boundary is very difficult and is still an open challenge.

To address the discontinuity of the aortic bifurcation, we

Wei Wang is corresponding author.

proposed a new convolution to improve the segmentation results. This convolution helps obtain rich spatial information in semantic segmentation. The main contributions of this paper are three folds: 1) we proposed the novel AOSC convolution to aggregate contextual information for each sample point, and avoid inaccurate segmentation on the region around the narrow tubular branches; 2) we introduced SDM as a weight map to measure the similarity of points, and the learned 3D offset field was embedded with AOSC; 3) we proposed a method having strong generalization, and it can be extended to segmentation tasks easily. Our proposed method improve the final aorta segmentation map, especially for the narrow tubular branches.

II. Method

As is depicted in Fig.2, the framework is based on V-Net [9]. It mainly consists of two parts: the SDM module and the AOSC module. The SDM module provides auxiliary information regarding spatial distance between adjacent points. This spatial information is significant to identify appropriate features for discontinuous points; The AOSC proposed a novel feature fusion mechanism to attenuate extreme features for points. So it improves the prediction accuracy for the points around narrow tubular branches. The two parts can be embedded in other 3D segmentation tasks easily.

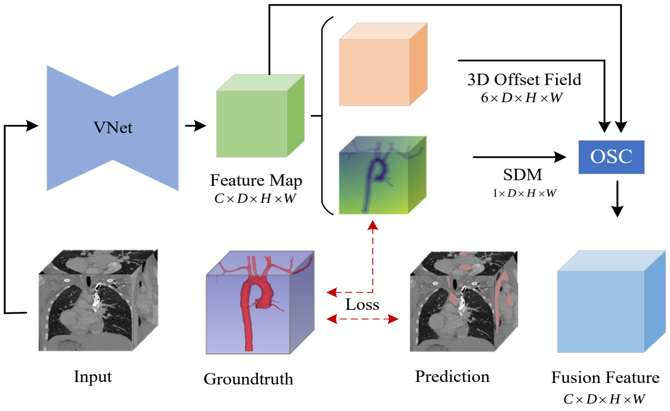
VNet
Feature Map
$C\times D\times H\times W$
3D Offset Field
$6\times D\times H\times W$
OSC
SDM
$1\times D\times H\times W$
Loss
Input
Groundtruth
Prediction
Fusion Feature
$C\times D\times H\times W$

Fig. 2. The framework. AOSC sampled features at discontinuous regions. We determined the weight of features according to SDM, and the produced new features help smooth the segmentation maps.

Our octahedron-shaped Convolution is shown in Fig.3. We explain how OSC is embedded in the VNet network, and the way of 3D feature sampling.

A. Structure Related Signed Distance Map

The SDM module is proposed to distinguish foreground and background points while simultaneously provide an elaborate weight map for the **feature fusion** in AOSC module. Noted that different classes of points are intertwined for discontinuous points in the predicted binary map. To better predict those discontinuous points, the SDM module introduced an effective guidance for the prediction of the segmentation results by measuring the similarities of adjacent points. Moreover, the SDM module also provided a weight map, which is inversely proportional to the difference between SDM values of two points, and it offers valuable guidance for the feature fusion in the AOSC module especially for the discontinuous points.

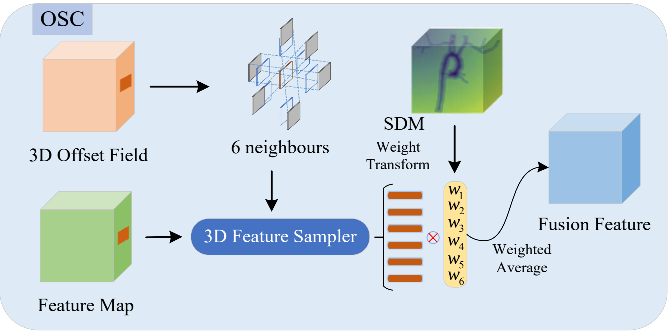

Fig. 3. The way of 3D feature sampling and fusion from a learned 3D offset Field.

Given the ground truth label T, let S be the surface of the aorta, which is defined as:

$$S = \{\mathbf{s} \in \Omega_{in} \mid T_{\mathbf{s}} = 1, \exists \mathbf{q} \in \mathcal{N}(\mathbf{s}), T_{\mathbf{q}} \in \Omega_{out}\} \quad (1)$$

where $\mathcal{N}(\mathbf{s})$ represents the 26-neighbour points of s in 3D space. Signed Distance Map (SDM) [1] which maps R^3 to R is defined as:

$$D(x) = \begin{cases} 0, x \in \mathcal{S} \\ -\inf_{y\in\mathcal{S}} \|x-y\|_2, x \in \Omega_{\text{in}} \\ +\inf_{y\in\mathcal{S}} \|x-y\|_2, x \in \Omega_{\text{out}} \end{cases} \quad (2)$$

where Ω_{in} and Ω_{out} denote the region inside and outside of the aorta respectively. We adopt Euclidean distance to calculate the distance from each point to its nearest surface S, because the Euclidean distance is robust to the tubular structure. Therefore, SDM ensures the continuity and the surface smoothness to some extent.

B. Octahedron-shaped Convolution

Many factors cause discontinuities in segmentation. One of the most important factors is that the gray values from foreground points and that from background points are approximate, particularly when foreground points and background points tend to be adjacent. In this section, we proposed an Octahedron-shaped Convolution (OSC) to decrease the discrepancy of intra-class points to avoid the segmentation discontinuity.

As shown in Fig.4, we try to capture relevant features around each center point, and further fuse these relevant features based on SDM to achieve more representative feature embedding. Moreover, the embedding features sampled from the feature maps are averaged with rich contextual information. We regard points around each center point as vertices, and the constructed geometry is shaped like an Octahedron. In this work, based on the OSC, we further propose two variants: FOSC and AOSC, which are illustrated as follows.

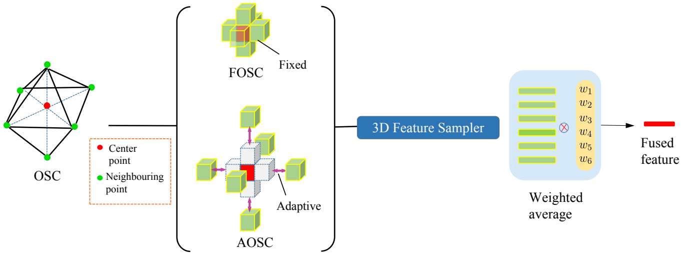

Fig. 4. The central point finds the most similar points from 6 directions with different steps. In FOSC, six steps are all set to 1, while in AOSC six steps are learned adaptively.

1) Fixed Octahedron-shaped Convolution: To capture fixed relevant features, we proposed the first variant of OSC——FOSC. Features that are directly adjacent have high feature similarity, because they are more likely to be from the same category. So we define basic offset $\mathcal{R}$ as:

$$\mathcal{R} = \left\{\Delta p | \Delta p \in (M_{k_x}, M_{k_y}, M_{k_z})\right\} \tag{3}$$

where $(M_{k_x}, M_{k_y}, M_{k_z})$ is a unit vector, and for $(M_{k_x}, M_{k_y}, M_{k_z})$,$M_{k_i} = c, M_{k_j} = 0, i = x, y, z, j \neq i$. In this unit vector, c is a constant, which is set to 1 or -1, representing the offset of the current point to its immediate neighbors. Then we map every single vector Δp in set $\mathcal{R}$ to a three-dimensional shifting grid. The shifting grid is used to preserve the spatial information at places where the discontinuity of prediction results often occurs. Consequently, the discontinuity of segmentation around aorta branches is significantly alleviated. The FOSC module is formalized as below:

$$F_1(p_v) = \frac{1}{\sum_{\Delta p \in \mathcal{R}} \mathcal{W}(p_v)} \sum_{\Delta p \in \mathcal{R}} \mathcal{W}(p_v) \cdot \mathbf{B}(p_v + \Delta p) \tag{4}$$

Function **B** represents bi-linear interpolation because the coordinates of points we intend to sample are not always integers. $\mathcal{W}(p_v)$ represents the SDM differences between p_v and its 6-neighbour points, which are denoted as the following two forms:

- direct combination of 6 immediate neighbors:

  $$\mathcal{W}(p_v) = \frac{1}{|D\left(p_v\right) - D\left(p_v + \Delta p\right)|} \tag{5}$$

  where function $D(x)$ is SDM shown as Equation 2.
- combination of some of the points with same SDM symbols:

  $$\mathcal{W}(p_v) = 1 - |D\left(p_v\right) - D\left(p_v + \Delta p\right)| \tag{6}$$

  For continuous tubular structures, points with same SDM symbols are often assigned same labels in prediction. Function $D(x)$ is SDM shown as Equation 2, too.

2) Adaptive Octahedron-shaped Convolution: Each separated point aggregates several feature points that are similar to the current point adaptively. We introduce a 3D offset field to record the coordinates of the most similar features in a grid. Then the discontinuous points are reduced so as to decrease the discontinuities of segmentation around aorta branches. We detail the notation of the 3D offset field: for each point p_v, we find its neighbour points $m_i (i = 0, 1, ..., 5)$ which are several unit steps far from p_v. The final offset vector $\overrightarrow{p_v m_i}$ is the product of the learned variable $\boldsymbol{\gamma}$ and unit vector $(M_{k_x}, M_{k_y}, M_{k_z})$(which is defined in II-B1). Offset vectors $\overrightarrow{p_v m_i}$ are denoted as:

$$\begin{aligned}\overrightarrow{p_v m_i} &= p_v + \gamma \Delta p \\ &= \{(x_v + \gamma_x M_{k_x}, y_v + \gamma_y M_{k_x}, z_v + \gamma_z M_{k_x})\}_{k=1}^{N}\end{aligned} \tag{7}$$

where γ is a positive float number in the range of $[0, 1]$ learned by the network, and the distance of the entire feature map is normalized to float between 0 and 1. $N = 6$, because there are six fused features of point p_v. Then we formalized the fusion features of AOSC as:

$$F_2(p_v) = \frac{1}{\sum_{\Delta p \in \mathcal{R}} \mathcal{W}(p_v)} \sum_{\Delta p \in \mathcal{R}} \mathcal{W}(p_v) \cdot \mathbf{B}(p_v + \gamma \Delta p) \tag{8}$$

Similarly, $\mathcal{W}(p_v)$ represents the SDM differences between p_v and its 6-neighbour points, which also has two forms as Eq.5 and Eq.6. $\mathcal{R}$ is the offset set represented in Eq.3, and function $D(x)$ is SDM shown as Equation 2. Function **B** represents bi-linear interpolation.

The most difference between function $F_2(p_v)$ and $F_1(p_v)$ is that we add an extral variable γ to adjust the steps of the adaptive aggregated features. This allows each point to find neighbour points that are more similar to its feature expression, thereby increasing the similarity of inter-class features.

C. Training Objective

To obtain the final segmentation map *out_map*, we processed the SDM [1] branch in the network as follows:

$$out\_map = Sigmoid(\boldsymbol{\mu} \cdot out\_dis) \tag{9}$$

where $\boldsymbol{\mu}$ is a hyperparameter set to -1500 in this experiment, and *out_dis* is learned by SDM of VNet.

The proposed method involved loss function on three parts:

- L^i_{Dice}: the initial pixel-wise segmentation loss
- L_1: the regression SDM output loss, which is given by $L^s_1 = |out\_dis - T|$
- L^s_{Dice}: the final segmentation map coming out from the *out_dis*, so we calculate between *out_map* and groundtruth T

The overall loss function is defined as follows, where α is set to 0.5:

$$L_A = \alpha(L^s_1 + L^s_{Dice}) + (1 - \alpha)L^i_{Dice} \tag{10}$$

III. Experiment

A. Datasets and Implementation Details

We conducted experiments on two datasets:

- ***: the corresponding manual annotation comes from ***. This dataset contains cardiac CT of 87 patients ranging from 5-day-old infants to 68-year-old adults. We further divide the 87 training images into 85% training, 5% validation and 10% test.

- MICCAI ASOCA2020: The organization provides both image and groundtruth which are totally of number 40. We divide the datasaet into 55% training, 25% validation and 20% test.

a) Evaluation metrics: We adopt the widely used 3D Dice coefficient [9], 3D Jaccard coefficient [10] to measure our method.

b) Implementation Details: The ultimate optimization goal of the network is the proposed loss function in Eq. 10 using ADAM optimizer [11] with the learning rate set to 0.01 for 80 epochs. Data augmentation is applied to prevent over-fitting including random rotation with the random angle between -25 and 25 (degree measure), as well as adding random Gaussian noise. We train the network on GeForce GTX TITAN X, and due to the limitations of GPU, the batch size of each GPU is 1 with resized $160 \times 160 \times 160$ inputs.

B. Results

We evaluated our proposed method on our-collected dataset and the MICCAI ASOCA2020 dataset. Firstly, in order to verify that the SDM module is effective, we use SDM as the only output, which is the SOTA work proposed by Ma. et al [12]. Another VNe_MultiHead architecture is conducted where one head is for distance output and the other for segmentation output conducted by [13]. Secondly, in the fixed octahedron-shaped convolution module, we try to add features of adjacent points termed as $\mathbf{FOSC/6-Neighbor}$. Finally, we let the network learn an offset map as AOSC module named $\mathbf{AOSC/6-Offset}$. Table.I shows the comparison of those prior SOTA networks and our methods.

TABLE I
COMPARISON OF OUR METHODS WITH SOTA METHODS

Method	DICE	JACCARD
our-collected dataset		
3D U-Net(baseline) [14]	0.6681(±0.1323)	0.5113(±0.0092)
VNet_SDM [1]	0.7309(±0.1543)	0.5990(±0.1910)
VNet_MultiHead [15]	0.7909(±0.0656)	0.0669(±0.1002)
Ours(FOSC)	0.8301(±0.0959)	0.7221(±0.1370)
Ours(AOSC)	**0.8493(±0.0971)**	**0.7499(±0.1410)**
MICCAI ASOCA2020		
VNet_SDM [1]	0.7122(±0.1992)	0.5633(±0.0112)
VNet_MultiHead [15]	0.7039(±0.0758)	0.5483(±0.0889)
Ours(FOSC)	0.7399(±0.1928)	0.5901(±0.0318)
Ours(AOSC)	**0.7434(±0.0385)**	0.5930(±0.0473)

For ablation study, we respectively verified that aggregating immediate neighbour 6 feature points into a new feature (named $\mathbf{Fixed/Direct\ 6-neighbours}$) and aggregating only several features with same symbols of SDM (named $\mathbf{Fixed/Signed\ 6-neighbours}$) in FOSC module. While in AOSC module, we explored whether different offsets in three axial directions affect the segmentation results. So we designed $\mathbf{Adaptive/Same-steps}$ to regular the same offset steps in x,y and z axis, and $\mathbf{Adaptive/Diff-steps}$) to learn different offset steps in three axis respectively. Table.II illustrates the results of ablation experiments.

TABLE II
COMPARISON OF OUR METHODS WITH SOTA METHODS

Method		DICE	JACCARD
Fixed	Direct 6-neighbours	0.8310(±0.0959)	0.7221(±0.1370)
	Signed 6-neighbours	0.8321(±0.0990)	0.7238(±0.1348)
Adaptive	Same-steps	**0.8493(±0.0971)**	**0.7499(±0.1410)**
	Diff-steps	0.8339(±0.1010)	0.7278(±0.1464)

We visualized the aorta segmentation results from our-collected dataset as Fig.5 and Fig.6. We compared the results of VNet, VNet_MultiHead, FOSC and AOSC, then we visualized the segmentation results from MICCAI ASOCA2020 dataset as Fig.7. Experiment results demonstrated that the segmentation of 3D U-Net is obviously discontinuous at the aortic branches while significant improvements have been achieved in the proposed FOSC and AOSC, because our proposed network aggregated feature vectors and made better use of contextual information.

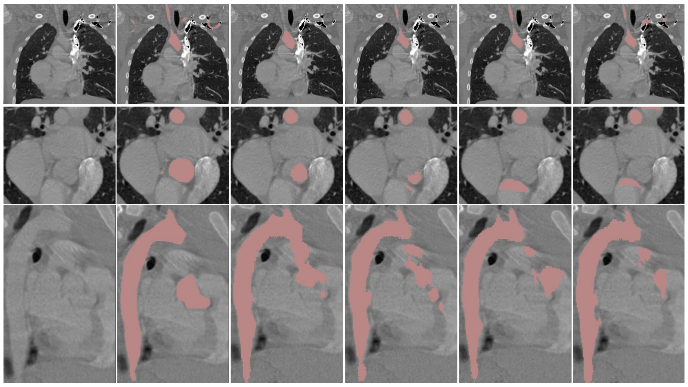

Fig. 5. Three-view slices of a patient: from the first to the third row, we show the axial, sagittal and coronal view of segmentation results respectively. From left to right, slices are image, groundtruth, VNet, VNet_MultiHead, FOSC/Direct 6-neighbours, and AOSC/Same-steps. For the second row, (a)-(e) represent image, groundtruth, VNet_SDM, VNet_MultiHead, AOSC/Diff-steps.

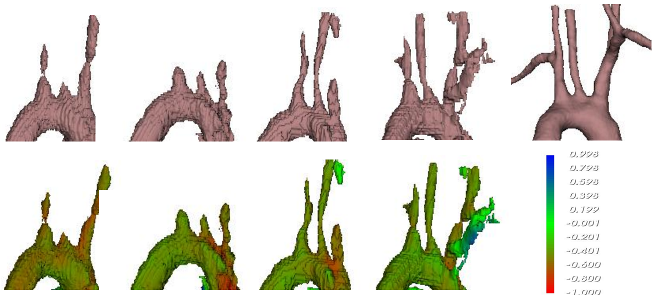

Fig. 6. 3D reconstruction results of the Signed Hausdorff Distance: from left to right, they are VNet, VNet_MultiHead, FOSC/Direct 6-neighbours, AOSC/Same-steps and groundtruth.

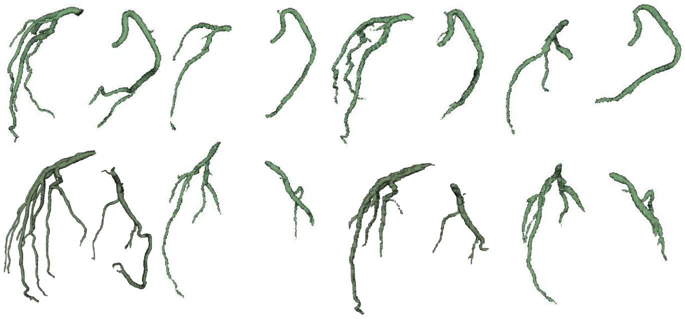

Fig. 7. The segmentation results two patients with coronary artery, from left to right, they are groundtruth, VNet, AOSC/Same-steps and AOSC_Diff-steps. Reconstruction result

The results demonstrated that, compared with VNet, the adaptive octahedron-shaped convolution can improve the accuracy of segmentation. When the branch bifurcation points are predicted accurately, the small diameter branches are segmented more finely, and the predicted length of the branches are also increased. The reason why AOSC/Diff-steps is better than AOSC/Same-steps is that, in 3D space, different offsets of the x, y, and z axes can find more diverse similar features. Although the relative positions of these features are difficult to determine, more offsets make the coordinates more accurate.

IV. Conclusion

In this paper, we presented an effective method making octahedron-shaped convolution for CT images. In order to explore the contextual semantic relationship between points, the proposed method aggregated feature vectors of sampling points, especially at the narrow tubular branches. Besides, we introduced signed distance map to constrain the shape of segmentation organs, and this map can measure the similarity of points. The experiment results demonstrated that our method achieved the state-of-the-art results.

References

[1] Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang, "Shape-aware organ segmentation by predicting signed distance maps," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 565–12 572.

[2] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 552–561.

[3] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 165–174.

[4] S. Dangi, C. A. Linte, and Z. Yaniv, "A distance map regularized cnn for cardiac cine mr image segmentation," *Medical physics*, vol. 46, no. 12, pp. 5637–5651, 2019.

[5] N. Audebert, A. Boulch, B. Le Saux, and S. Lefèvre, "Distance transform regression for spatially-aware deep semantic segmentation," *Computer Vision and Image Understanding*, vol. 189, p. 102809, 2019.

[6] N. Kyurkchiev and S. Markov, "On the hausdorff distance between the heaviside step function and verhulst logistic function," *Journal of Mathematical Chemistry*, vol. 54, no. 1, pp. 109–119, 2016.

[7] F. Cheng, C. Chen, Y. Wang, H. Shi, Y. Cao, D. Tu, C. Zhang, and Y. Xu, "Learning directional feature maps for cardiac mri segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 108–117.

[8] J. Chu, Y. Chen, W. Zhou, H. Shi, Y. Cao, D. Tu, R. Jin, and Y. Xu, "Pay more attention to discontinuity for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 166–175.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[10] J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 92–100.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] J. Ma, Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, G. Chen, J. Liu, C. Peng, L. Wang, Y. Wang, and J. Chen, "How distance transform maps boost segmentation cnns: An empirical study," in *Medical Imaging with Deep Learning*, ser. Proceedings of Machine Learning Research, T. Arbel, I. B. Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, and C. Pal, Eds., vol. 121. PMLR, 06–08 Jul 2020, pp. 479–492. [Online]. Available: http://proceedings.mlr.press/v121/ma20b.html

[13] F. Navarro, S. Shit, I. Ezhov, J. Paetzold, A. Gafita, J. C. Peeken, S. E. Combs, and B. H. Menze, "Shape-aware complementary-task learning for multi-organ segmentation," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 620–627.

[14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[15] Y. Wang, X. Wei, F. Liu, J. Chen, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Deep distance transform for tubular structure segmentation in ct scans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3833–3842.

1

[1] Xi Xiang, Tel:17863137573, Email:19S103133@stu.hit.edu.cn