IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier xx.xxxx/ACCESS.2019.DOI

Coronary Arteries Segmentation based on 3D FCN with Attention Gate and Level Set Function

YE SHEN[1], ZHIJUN FANG[1],(Senior Member, IEEE), YONGBIN GAO[1], NAIXUE XIONG[2], (Senior Member, IEEE), CENGSI ZHONG[1],AND XIANHUA TANG[1],

[1]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science,Shanghai,201620,China
[2]Department of Math-ematics and Computer Science, Northestern State University,Tahlequah,OK,U.S

Ye Shen and Yongbin Gao contributed equally to this work
Corresponding author: Zhijun Fang (zjfang@sues.edu.cn) and Naixue Xiong (xiongnaixue@gmail.com).

This paper is supported by the Shanghai Science and Technology Commission(No.18411952800).

ABSTRACT Coronary heart disease is one of the most serious health problems in the world nowadays. By segmenting and examining the coronary arteries in medical images, we can find artery stenosis and plaque, which are the main causes of this certain disease. Segmenting coronary arteries manually is time-consuming and subjective, and the traditional segmentation method requires a good initial point, which is also difficult to apply for 3D coronary computed tomography angiography (CTA) data. In this paper, we propose a joint framework for coronary CTA segmentation based on deep learning and traditional level set method. A 3D fully convolutional network (FCN) is used to learn the 3D semantic features of coronary arteries, which provides an excellent initial point for traditional level set. Moreover, an attention gate is added to the entire network, aiming to enhance the vessels and suppress irrelevant regions. The output of 3D FCN with attention gate is optimized by level set to smooth the boundary to better fit the ground truth segmentation. The proposed algorithm is evaluated by Jaccard index (JI) and Dice similarity coefficient (DSC) scores, experimental results show that the proposed framework provides significant better segmentation results than the vanilla 3D FCN intuitively and quantitively.

INDEX TERMS Coronary CTA, image segmentation, 3D FCN, attention gate, level set function.

I. INTRODUCTION

Cardiovascular disease is one of the most dangerous fatal disease in the world. With the improvement of our living standards, cardiovascular diseases continue to spread, and the age of onset is becoming younger, which is a serious threat to human health. Coronary artery heart disease (CHD), also known as coronary heart disease, is relatively common. The source of the disease is the deposits on the coronary artery wall which leads to the stenosis of the blood vessel, resulting in myocardial ischemia and thus cardiac organic dysfunction. So it is also called ischemic heart disease (IHD) [1].The main cause of coronary heart disease is atherosclerosis of the coronary arteries, so it is customary to call coronary heart disease as coronary atherosclerotic heart disease. Clinicians often face certain type of coronary heart disease patients, whose coronary lesions are at a critical state (with a 50% to 70% stenosis). Whether a patient with a critical lesion needs to be implanted with a stent has always been a controversial topic. The severity of such lesion should be assessed not only from the coronary but also from the damage of the patient's myocardial function. Different doctors may have different therapies for the same lesion with the sole knowledge of coronary angiography results. In fact, some patients have more than 50% coronary artery stenosis, but their myocardium has no obvious ischemia. For others however, although the degree of coronary stenosis is less than 50%, their myocardium has obvious ischemia. Evidently, using coronary angiography to guide the interventional treatment is very subjective, and cannot reflect the true impact of stenosis on myocardial blood supply.

Segmenting coronary arteries with computer-aided diagnosis technology is clinically significant for early detection and treatment of coronary heart disease. Segmentation is a highly correlated task in medical image analysis. Image segmentation is the process of dividing an image into smaller partitions based on some characteristics of the pixels in the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

image. In medical imaging, the division of images can be the area of certain tissue types, organs or other related structures [2]. Segmentation tasks are mostly used for quantitative analysis and diagnosis. The gold standard for medical image segmentation is manually completed by clinical experts. This is a time-consuming task because modern medical imaging modalities (such as CT and MRI) are capable of generating large amount of data in the form of 3D image volumes. There are also deviations and errors in manual segmentation. Some semi-automated methods have already been used in clinical diagnosis to speed up segmentation, but clinical experts are still needed to initialize or direct the process. Fully automated segmentation method is becoming more and more important with the increase of patients' data.

Segmentation algorithms based on level sets have been widely used and become the preferred algorithm for medical image segmentation in past decades [3], [4]. The level set method performs segmentation based on energy minimization by integrating different types of regularization (smooth terms) and priors [5]. They can provide segmentation functions with a tendency to change topological properties, with the disadvantage that they require proper contour initialization to obtain valid segmentation results and lack of semantic features. Recently, a deep learning method based convolutional neural network (CNN) has been successfully applied to medical imaging, especially for segmentation and detection tasks [6], [7]. The difference from the levelset-based approach is that deep learning methods can automatically learn the appearance model from large training data, extract features from complex structures and patterns, and use these semantic features for prediction.

However, medical image segmentation based on deep learning is a more difficult task than natural image segmentation. First, patients' data is extremely diverse. In other words, the pattern of the same pathology varies from patient to patient. Second, small and incomplete medical data sets make CNN training more prone to overfitting [8]. Third, medical images such as MRI (magnetic resonance imaging) or CT (computed tomography) scans are usually in the form of 3D volume, while existing CNNs are mostly 2D. These 2D CNNs are applied by layers, so that the spatial information in the third dimension is ignored [9], while 3D CNN is discouraged due to computational complexity and memory requirements. In view of the these problems, recently a three-dimensional fully convolutional network (3D FCN) has been proposed for segmentation and detection from MRI or CT images, in which all of the volume data is used as an input to obtain a 3D output in the form of volume. This process is finished directly in a single forward propagation, and the output is regarded as a 3D predictive score. In this way the computational complexity is significantly decreased [10]. Unlike other 2D CNN-based methods, they use 3D convolution kernels to share spatial information in all three dimensions. Although 3D FCN does well in handling 3D medical data, some practices are overly dependent on multi-level cascade CNN, especially when the target organs between different patients show large differences in shape and size. The cascading framework first extracts a region of interest (ROI) and intensively predicts that particular ROI. It is mainly used in the heart [11]and abdomen image segmentation [12] and lung nodule detection [13]. However, this approach leads to complex parameters and a waste of computing resources. For example, low-level features similar to each other are repeatedly extracted by all models in the cascade. In order to solve this problem, a simple and effective solution, attention gate (AG), has recently been proposed [14]. The CNN model with AG does not change its standard way of training, while the AG can automatically learn features that focus on the target structure without additional supervision. During the testing process, these gates dynamically and implicitly generate soft area proposals and highlight salient features that are useful for a particular task. In addition, they do not incur significant computational costs or learn a large number of model parameters like a multi-model framework does. The advantage is that the proposed AG can improve the model sensitivity and accuracy of dense label prediction by suppressing feature activation in uncorrelated regions. In this way, the necessity of using an external organ positioning model is eliminated while high prediction accuracy is maintained.

The main contributions are summarized as following three-folds:

(1) 3D FCN network is adopted to three-dimensional coronary CTA images, which is the first attempt to use 3D deep learning for coronary arteries segmentation. Compared to traditional level set segmentation, 3D FCN is capable of learning 3D shape of the coronary arteries as prior knowledge.

(2) Attention gate mechanism is incorporated into the 3D FCN network structure to improve the accuracy, which highlights the vessels for each CTA slice.

(3) A level set function is integrated with 3D FCN to optimize the network prediction results, aiming to smooth the boundary of the coronary arteries segmentation.

The remaining of this paper is organized as follows: Section 2 mainly introduces the methods and principles adopted in this paper, including the structure of the network and the flow chart of the algorithm. Section 3 is the conclusion analysis and visualization of the experiment. Section 4, Summary of Work.

II. METHODOLOGY

In this study, a joint training framework is proposed to segment the coronary arteries as shown in Figure 1. In the training phase, the weights are trained through the 3D FCN network integrating attention gate. For the test phase, the trained weights are used to give coarse predictions, which are then optimized using the level set function to obtain a refined smooth segmentation. Compared with traditional segmentation algorithm, 3D FCN learns the semantic features and 3D shape of the coronary arteries. The attention gate highlights the vessels for each CTA slice, and the additional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

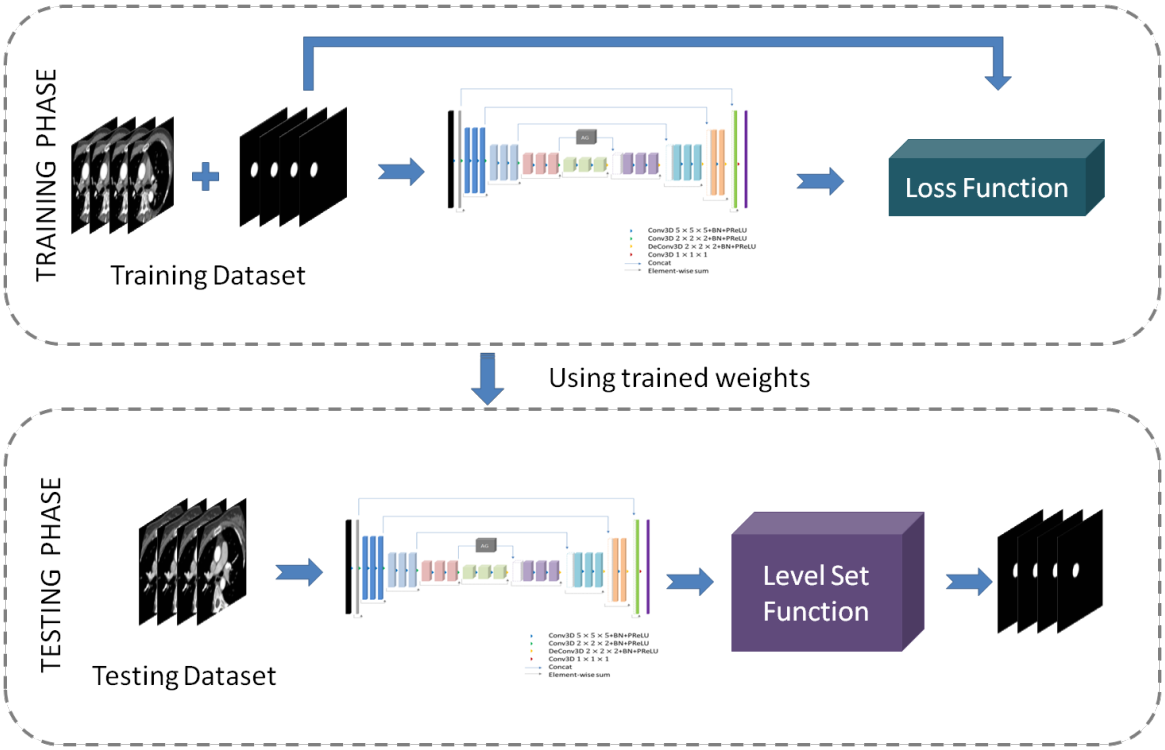

FIGURE 1. Overall flow chart.

level set function smooths the boundary of the segmentation of coronary arteries.

A. 3D FULLY CONVOLUTIONAL NETWORK FOR COARSE SEGMENTATION

Traditional algorithms segment each slice by analyzing their low-level features, which is difficult to extract the 3D correlation between slices since no prior shape knowledge is involved during segmentation. In this paper, we proposed to use a 3D fully convolutional network for coarse segmentation of coronary arteries as shown in Figure 2. This network is similar to the mainstream U-NET [16], [17] and V-NET [18] in structure. We perform 3D convolution to extract features from 3D CTA image data, the 3D convolution learns the prior shape and 3D temporal correlation of coronary arteries and extracts high-level semantic features for segmentation. The left part of the network is referred as encoder, which is used to extract high-level semantic features, while the decoder on the right side restores the semantic features to the coarse segmentation.

Assuming that F_k^{l-1} is the k^{th} feature volume of the $l-1$ layer, the m^{th} feature volume of the l layer is given as:

$$F_m^l(x,y,z) = \sigma(\sum_m (\sum_{m,n,t} F_k^{l-1}(\alpha,\beta,\gamma) \otimes W_{ki}^l) + b_l^k) \quad (1)$$

where $\alpha = x - m, \beta = y - n, \gamma = z - t$;$W_{ki}^l$ represents a 3D filter with a kernel size of $(m \times n \times t)$, which is used to convolve on the feature volume F_k^{l-1} of each previous layer;$\otimes$ represents a 3D convolution operation;b_l^k denotes the deviation and $\sigma(\cdot)$ is a nonlinear activation function.A nonlinear activation function, the parametric rectification linear unit (PReLU) [19], is introduced here to replace the rectifying linear unit. The PReLU function is given as:

$$\sigma(F_i) = max(0, F_i) + \alpha \cdot min(0, F_i) \quad (2)$$

where F_i represents the input,$\sigma(F_i)$ is the output, and α_i, almost zero in ReLU, is the training parameter controlling the learning process of the negative part of F_i.Therefore, PReLU can adjust the rectifier according to the input, thus improving the accuracy of the network. This hardly increases the computational cost and additionally reduces the risk of overfitting. The left part of the network is divided into different segments, running at different resolutions. In a segment there are one to three convolutional layers, each performs a nonlinear operation on the input while their output sum up with that of the last, enabling the learning of the residual function [20]. The advantage of incorporating such residual function into the network is that the network can be brought to a convergent state in a short time. In each segment, the convolution kernel size is $5 \times 5 \times 5$ and the stride is 1. As the data advances along the compression path to different segments, its resolution gradually decreases. This is done by convolution with a kernel size of $2 \times 2 \times 2$ and a stride of 2. Since the second operation extracts features only through non-overlapping $2 \times 2 \times 2$ convolution kernels, the size of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

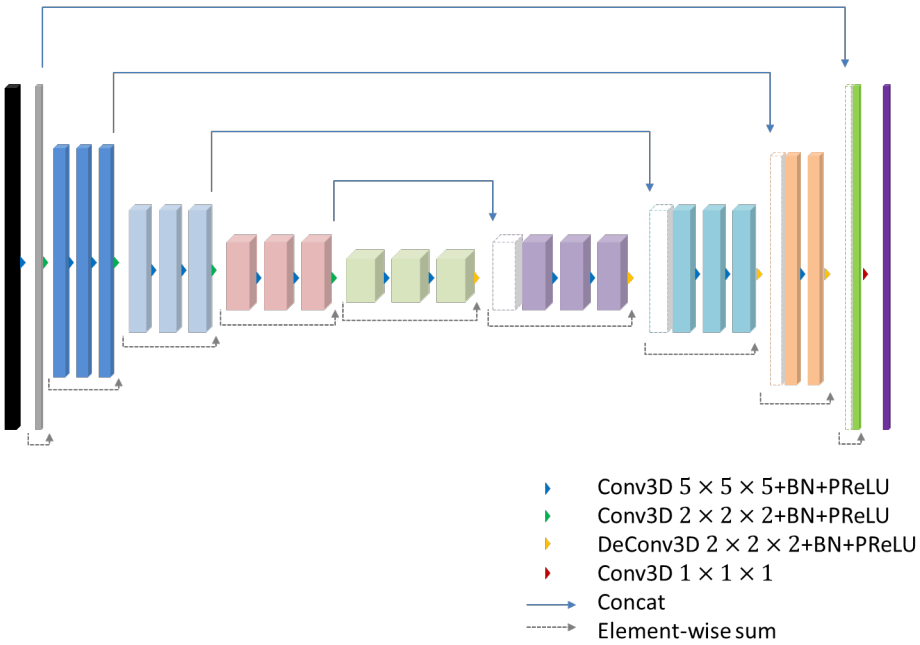

FIGURE 2. 3D FCN network structure.

the feature map is halved. This method of halving the feature map by convolution also replaces the pooling operations commonly used in previous CNNs [21]. Furthermore, since the number of feature channels is doubled in each segment of the encoder path in the network, and that the model is formed as a residual network, we use these convolution operations to double the number of feature maps while reducing their resolution. The PReLU nonlinear activation function is applied throughout the network. And before each PReLU function we also adopt Batch Normalization [22] to forcibly pull back each layer's input to a standard normal distribution. This can make the gradient larger, avoid the disappearance of the gradient, and accelerate the convergence of the network. Encoder through convolution also helps the network take up less memory during training. The downsampling part of the network's encoder path allows us to reduce the size of the input and extend the receive domain of features in subsequent network layers. The decoder mainly extracts features and extends lower resolution feature maps to better collect and combine the necessary information, and each layer calculates twice as many features as the previous one. The last convolutional layer uses a $1 \times 1 \times 1$ convolution kernel to produce an output of the same size as the input volume. The output is then converted to the probability value in the foreground and background regions by applying the sigmoid activation function. We use deconvolution to increase the size of the input after each stage of the decoder of the network. Similar to the encoder of the network, we also learn the residual function during the convolution phase to accelerate the convergence of the network model.

In this work, we use an objective loss function based on the tweezer coefficient with a value between 0 and 1. Our goal is to maximize the coefficient. The tweezer coefficient D between two binary volumes can be written as:

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3}$$

where the sum runs on N voxels, with the predicted binary partition volume $p_i \in P$ and the ground truth binary volume $g_i \in G$.

B. VESSEL ENHANCEMENT WITH ATTENTION GATE

The vessel shape and size of CTA varies from slice to slice for coronary arteries, vessel enhancement is important to filter out the noisy and confusing area in CTA slice. An attention gate is incorporated into the 3D FCN to highlight the vessel-like structures. In the standard CNN network model, we generally take a gradual downsampling of the feature graph grid to capture a large enough perceptual field to better capture semantic context information. However, it is still difficult to reduce false positive predictions for small objects that exhibit large shape variability by only downsampling. To improve accuracy, current segmentation frameworks rely on additional prior object location models to simplify tasks into separate positioning and subsequent segmentation [23]. In this task, we can achieve the same goal by incorporating attention gate (AG) into the standard CNN model. Contrary to the positioning model in pyramid CNN, it does not need to train multiple models. Additionally, AG can gradually

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

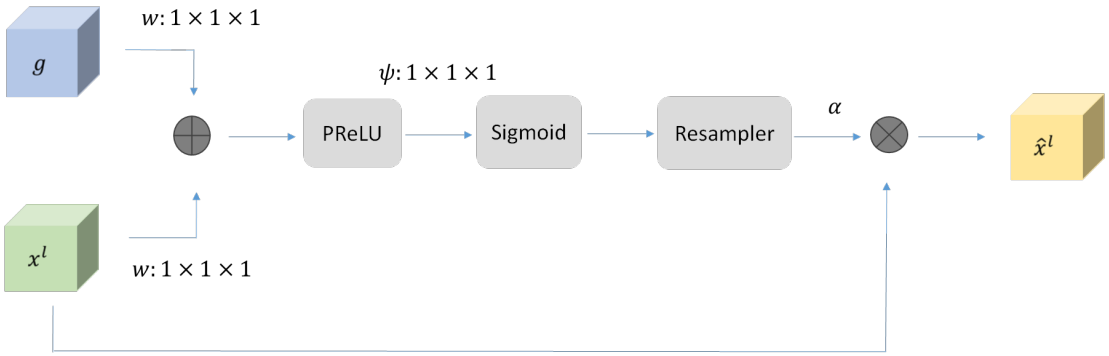

FIGURE 3. The workflow of Attention Gate.

suppress the characteristic response in unrelated background regions without passing multiple levels.

The output of attention gate is an element-by-element multiplication of the input feature map and the attention coefficient. The formula is as follows:

$$\widehat{x}_{i,c}^{l} = x_{i,c}^{l} \cdot \alpha_i^l \tag{4}$$

The attention coefficients $\alpha_i \in [0, 1]$ is included in the formula which is to identify salient image regions and prune feature responses to retain activation related to specific tasks. Generally, a single scalar attention value is calculated for each pixel vector $x_i^l \in \mathbb{R}^{F_l}$, where F_l corresponds to the number of feature maps in the layer l.

We obtained multi-dimensional attention coefficient to deal with multiple semantic conditions. Therefore, every AG focuses on the subset of the target structure and it contains gating vectors $g_i \in \mathbb{R}^{F_g}$ for each pixel i which is used to determine the focus area. The gating vector is used to prune the lower feature responses with context information [24]. By comparing the performance of multiplicative attention [25] and additive attention [26], we emperically chose additive attention to obtain the gating coefficient. The additional attention formula is as follows:

$$q_{att}^{l} = \psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\psi \tag{5}$$

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l, g_i; \Theta_{att})) \tag{6}$$

And the features of AG are characterized by a set of parameters including: linear transformation $W_x \in \mathbb{R}^{F_l \times F_{int}}$,$W_g \in \mathbb{R}^{F_g \times F_{int}}$,$\psi \in \mathbb{R}^{F_{int} \times 1}$ and bias term $b_\psi \in \mathbb{R}$,$b_g \in \mathbb{R}^{F_{int}}$. For input tensors, we chose a vector Cascade-based attention method with $1 \times 1 \times 1$ convolution kernels to compute linear transformations, in which cascade features x^l and gate are linearly mapped to $\mathbb{R}^{F_{int}}$ dimensions. The attention coefficient (σ_1) uses PReLU as non-linear activation function, because PReLU activation function can adjust the rectifier according to the input situation to improve accuracy compared with the widely used ReLU activation function; The software Max function will produce sparse activation at the output if used sequentially, thus attention coefficient (σ_2) employs sigmoid activation function so that the AG parameters can converge better in the training process. The overall process of AG is shown in the following figure:

We highlight the vessels of coronary arteries through skip connections by merging AG into the network architecture of 3D FCN. The structure is shown in Figure 3. The roughly extracted information is used for the gates to eliminate the irrelevance and noise caused by the transition. Additionally, AG filtered neuron activation during forward and back propagation, the gradients from background regions are weighted downward during back propagation. This allows model parameters in shallower layers to be updated mainly based on the spatial region associated with a given task. In each subAG, additional information is extracted and fused to define the output of the skip connection. We used $1 \times 1 \times 1$ convolution kernel in linear transformation to reduce the number of trainable parameters and the computational complexity of AG, the resolution of the input feature map being downsampled to the gating signal is similar to that of a non-local block [27]. The corresponding linear transformation decouples the feature map and maps them to the lower dimensional space for gating operations. The low-level feature maps are discarded in gating because they do not represent input data in high-dimensional space. Thus we add AG directly to the skip connection of the last layer of network structure to enhance the learning of relevant features of the whole network.

C. OPTIMIZATION WITH LEVEL SET

The 3D FCN with attention gate learns the semantic shapes of coronary arteries. Because the output of deep learning is typically non-smooth and blurry, the boundary of the segmentation is also not perfectly fit with the real edges of coronary arteries. In this study, the traditional level set algorithm is added to refine and smooth the boundary of the segmentation results. The basic idea of level set [28] is that the plane closed curve is implicitly expressed as the level set of 2D surface function, that is, the point set with the positive/negative value corresponding to the foreground/background. Although this transformation makes the problem more complex in form, it

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

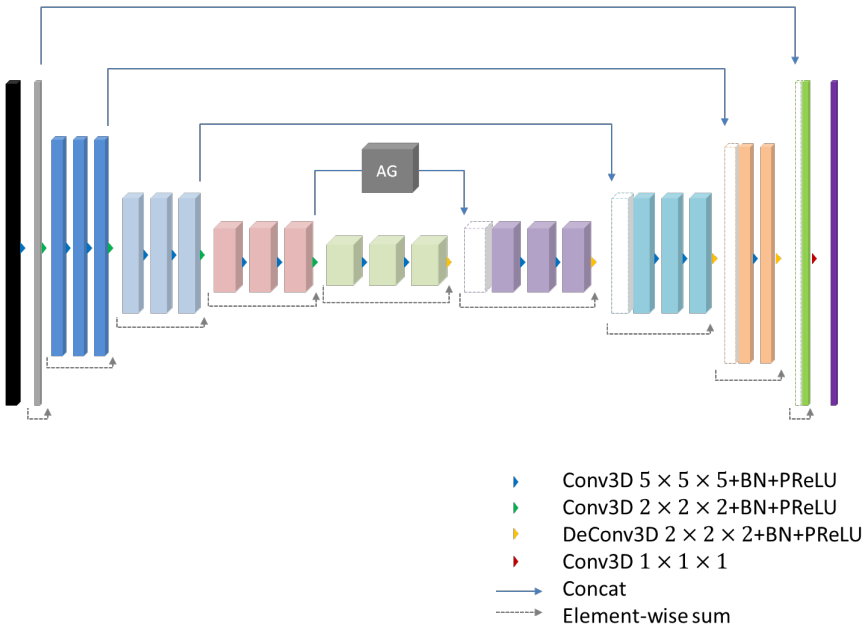

FIGURE 4. 3D FCN Integrating Attention Gate.

brings many advantages in calculating. Its greatest advantage is that the topological change of the curve can be handled naturally and the only weak solution satisfying the entropy condition can be obtained. The evolution of level set function satisfies the following basic equations:

$$\phi_t + F|\nabla\phi| = 0 \tag{7}$$

where ϕ is the level set function whose zero level set represents the target contour curve $\Gamma(t) = x|\phi(x,t) = 0$, which is the gradient norm of the level set function, the motion of the curve is controlled by the velocity function F in the normal direction of the surface. Generally, it includes items related to images (such as gradient information) and geometric shapes of curves (such as curvature of curves).

III. EXPERIMENTAL RESULTS

The coronary CTA image data used in our experiment consisted of 70 groups of patients with the number of slices ranged from 250 to 350 in each group, these patients are randomly selected from Shanghai Zhongshan Hospital. The coronary CTA data contains irrelevant slices at the beginning and the end, where there is no aorta or vascular existed. As shown in Figure 5, Figure 5(a) is the slice that contains the aorta and the coronary artery (as shown by the arrow pointing). Figure 5(b) contains only the aorta and the coronary artery has not appeared yet. Figure 5(c) is the slice where the blood vessels have disappeared completely. The valid slices that contains the aorta or the coronary artery in each group of patients is around 150 slices through statistical analysis, we select the first ten slices of coronary artery as the starting frame by manual screening and 160 slices were selected as experimental data for each group of patients, resulting in a dataset consisting of 11200 CTA images totally. The size of the CTA data is $512\times512\times160$ for each group of patients, 50 groups of patients are used as training set, and the remaining 20 groups of patients are used as test set.

We use Keras library to implement the model [29]. The Adam algorithm is used to optimize the network model. The learning rate is initially set to 0.00001 and train 500 epochs on a single NVIDIA GPU (Nvidia GTX 1080Ti). The training process takes about 10 hours. Due to the limitation of running memory, the input size of our data is set to $128 \times 128 \times 160$.

The visualization of the segmentation results is shown in Figure 6, each row in this figure corresponds to the different stages of coronary CTA data, (a) is the original CT image, (b) - (d) are the segmentation results of various methods, the green mark in the figure is the segmented coronary arteries, (e) is the GT, and the purple mark in the figure is the human labelled coronary arteries. First row shows that the aorta segmentation of vanilla 3D FCN is not good, the boundary is not fitted with the ground truth. Second row shows the false positive case where the none-vessels is classified as the coronary artery. Third row shows the false negative case in which some coronary artery is missing. Overall, the proposed joint framework incorporating 3D FCN, attention gate, and level set function receives the best visualization performance.

Quantitively, the performance of coronary arteries segmentation is evaluated by the Jaccard index (JI) and Dice similarity coefficient (DSC) scores. The range of the two values are between 0 and 1, and the higher value corresponds to the better segmentation accuracy. The calculation formulas

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

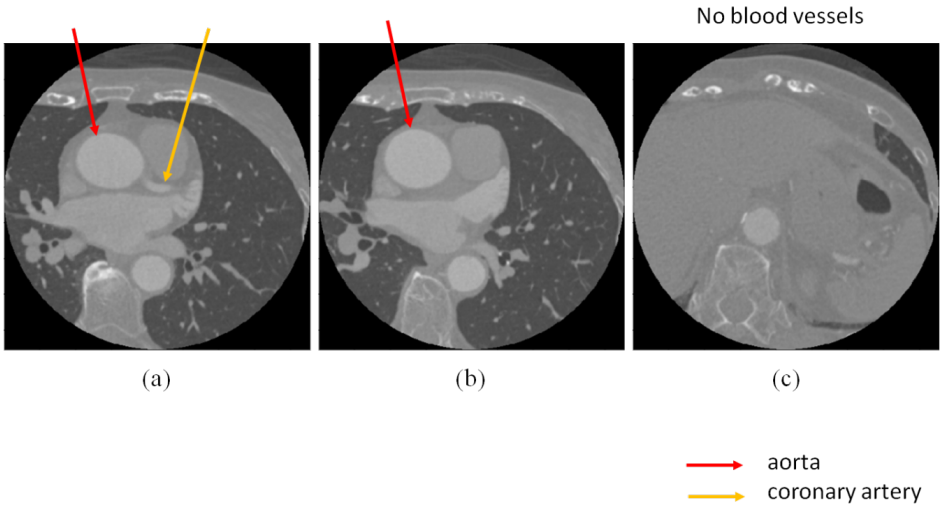

FIGURE 5. Data sample of various slices.

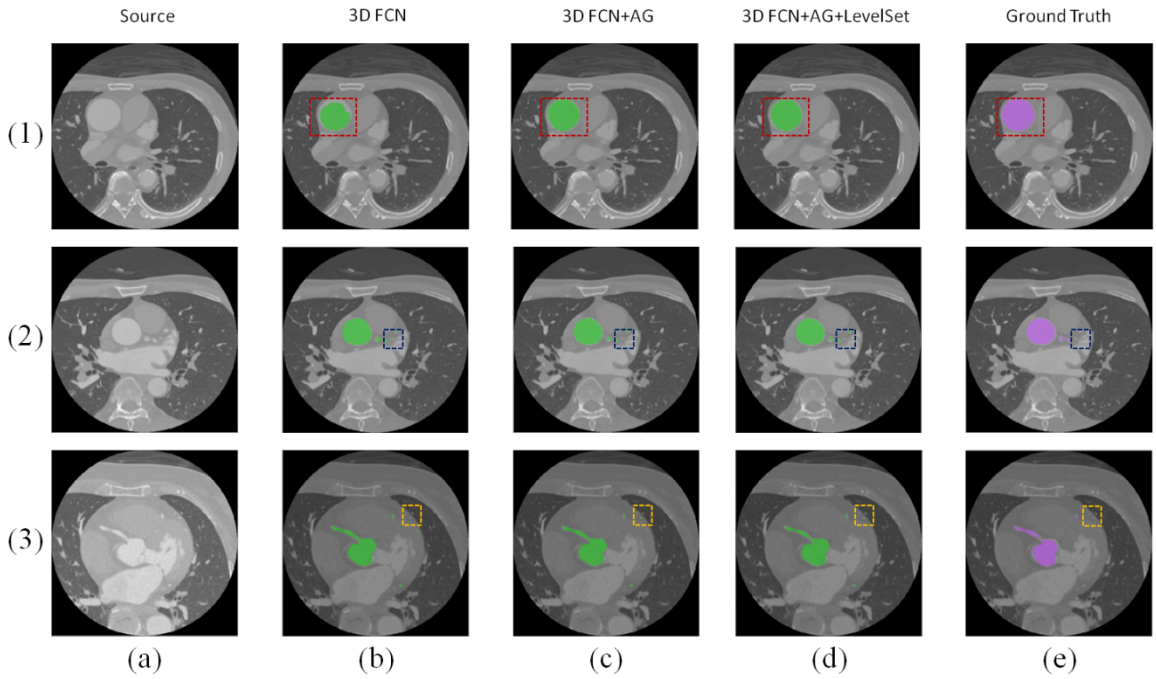

FIGURE 6. Segmentation results of different methods for three stages of coronary CTA data.

of JI and DSC can be defined as:

$$JI = \frac{|Y_+ \bigcap \widehat{Y}_+|}{|Y_+| \bigcup |\widehat{Y}_+|} \tag{8}$$

$$DSC = \frac{2|Y_+ \bigcap \widehat{Y}_+|}{|Y_+| + |\widehat{Y}_+|} \tag{9}$$

Where Y_+ represents Ground Truth (GT) and $\widehat{Y}_+$ represents predicted value.

The performance of the vanilla 3D FCN is first evaluated on the coronary artery for 20 groups of test patient data, the Mean JI and Mean DSC are 0.7961 and 0.8842, respectively. For the vanilla 3D FCN, the aorta can be segmented accurately while some small vessels of the coronary artery are missing. Then, We incorporate the 3D FCN with attention gate and find that Mean JI and Mean DSC values exceed the vanilla network significantly, which are 0.8155 and 0.8967, respectively. The improvement is more significant for the cases that the brightness is not particularly obvious in the coronary artery. The deep learning based method provides a good initial skeleton of coronary arteries, while the boundary

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2908039, IEEE Access

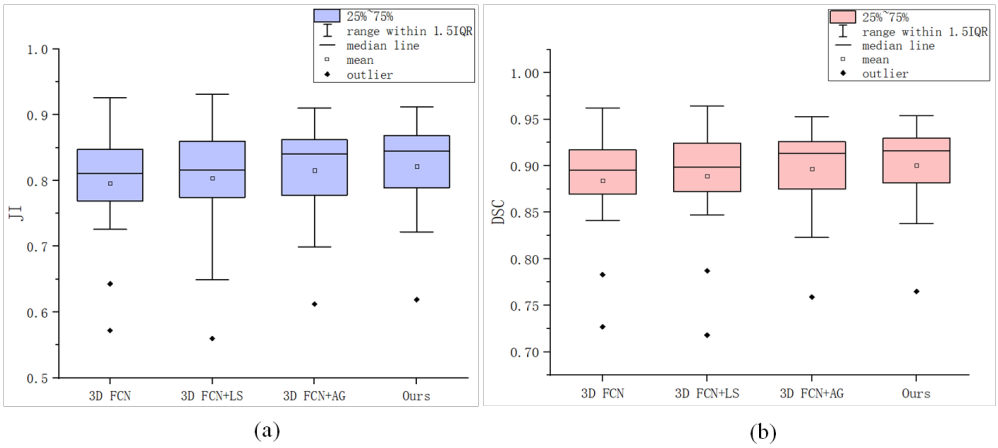

FIGURE 7. Boxplots of experimental data results.

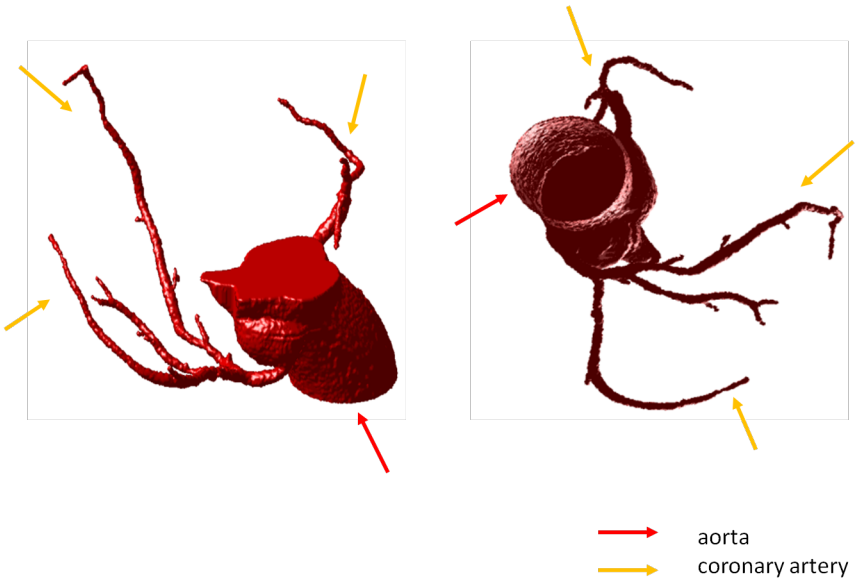

FIGURE 8. 3D visualization of Coronary segmentation.

TABLE 1. Comparison of Segmentation Accuracy for various algorithms.

Method	Mean JI	Mean DSC
3D FCN	0.7961	0.8842
3D FCN+LS	0.8035	0.8886
3D FCN+AG	0.8155	0.8967
Ours	**0.8217**	**0.9005**

is not smooth. Thus, we use a level set method to refine and optimize the results, and the final Mean JI and Mean DSC of our joint framework are 0.8217 and 0.9005, respectively. The performance in terms of Mean JI and Mean DSC for different algorithms are shown in table 1. At the same time, JI and DSC values of 20 groups of test patient data are shown in the form of boxplots in Figure 7.

Lastly, we reconstruct the final segmentation results into three dimensions by marching cube algorithm to intuitively check the segmentation results and its consistency. The visualization of the 3D reconstruction is shown in Figure 8, the 3D reconstruction of coronary arteries can facilitate the doctor to intuitively and effectively localize artery stenosis and plaque, which is critical to diagnose the coronary heart disease.

IV. CONCLUSION

In this paper, we proposed a method based on deep learning and level set for segmentation of coronary artery in CTA images. In order to process 3D volume data, we trained

an end-to-end 3D FCN network to predict the location of coronary artery, attention Gate mechanism is also incorporated into the 3D FCN network structure to suppress feature activation in irrelevant regions and improve the sensitivity and accuracy of the model for dense label prediction. At the same time, the network prediction results are fed into level set function to optimize the edge contour iteratively to obtain the final segmentation results. Compared with the segmentation results of vanilla 3D FCN network, our method provides better segmentation accuracy in terms of Mean JI and Mean DSC, and the visualization results in both 2D and 3D demonstrate the effectiveness of proposed framework, the segmentation and 3D reconstruction can facilitate the doctor to find artery stenosis and plaque more intuitively and effectively.

REFERENCES

[1] E. G. Nabel and E. Braunwald, "A tale of coronary artery disease and myocardial infarction, " New England Journal of Medicine, vol. 366, no. 1, pp. 54–63, Mar. 2012.

[2] M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios and I. Kokkinos, "Sub-cortical brain structure segmentation using F-CNN's ," International Symposium on Biomedical Imaging(ISBI), pp. 269–272, Feb. 2016.

[3] M. H. Soomro, G. Giunta, A. Laghi, D. Caruso, M. Ciolina, C. De Marchis, S. Conforto and M. Schmid, "Segmenting MR Images by Level-Set Algorithms for Perspective Colorectal Cancer Diagnosis," VipIMAGE 2017, ECCOMAS 2017, Lecture Notes in Computational Vision and Biomechanics, vol. 27, pp. 396–406, Oct. 2017.

[4] Y. -T. Chen, "A novel approach to segmentation and measurement of medical image using level set methods, " Magnetic Resonance Imaging, vol. 39, pp. 175–193, June. 2017.

[5] D. Cremers, M. Rousson and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," International Journal of Computer Vision, pp. 195-215, vol. 72, Apr. 2007.

[6] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis, vol. 36, pp. 61–78, Feb. 2017.

[7] J. Dolz, C. Desrosiers and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," NeuroImage, vol. 170, pp. 456–470, Apr. 2018.

[8] B. H. Menze et al, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[9] H. R. Roth et al, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," Medical Computing and Computer Assisted Interventions(MICCAI),Lecture Notes in Computer Science, vol. 9349, pp. 556–564, Nov. 2015.

[10] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi and P. -A. Heng, "Automatic detection of cerebral microbleeds from MR Images via 3D convolutional neural networks," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1182–1195, May. 2016.

[11] M. Khened, V. A. Kollerathu and G. Krishnamurthi, "Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," Medical Image Analysis, vol. 51, pp. 21–45, Jan. 2019.

[12] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," Medical Image Analysis, vol. 45, pp. 94–107, Apr. 2018.

[13] F. Liao, M. Liang, Z. Li, X. Hu and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3D deep leaky noisy-or network," arXiv preprint arXiv:1711.08324, Nov. 2017.

[14] J. Schlemper, O. Oktay, L. Chen, J. Matthew, C. Knight, B. Kainz, B. Glocker and D. Rueckert, "Attention-Gated Networks for Improving Ultrasound Scan Plane Detection," Medical Imaging with Deep Learning(MIDL), arXiv preprint arXiv:1804.05338, Apr. 2018.

[15] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," Medical Imaging with Deep Learning(MIDL), arXiv preprint arXiv:1804.03999, Apr. 2018.

[16] O. Ronneberger, P. Fischer and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," Medical Computing and Computer Assisted Interventions(MICCAI), Lecture Notes in Computer Science, vol. 9351, pp. 234–241, Nov. 2015.

[17] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," Medical Computing and Computer Assisted Interventions(MICCAI), Lecture Notes in Computer Science, vol. 9901, pp. 424–432, Oct. 2016.

[18] F. Milletari, N. Navab and S. A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571, Oct. 2016.

[19] K. He, X. Zhang, S. Ren and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," the IEEE International Conference on Computer Vision(ICCV), pp. 1026–1034, Feb. 2015.

[20] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 770–778, Dec. 2015.

[21] J. T. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller, "Striving for simplicity: The all convolutional net," The International Conference on Learning Representations(ICLR), arXiv preprint arXiv:1412.6806 , Dec. 2014.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167 , Feb. 2015.

[23] H. R. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, K. Misawa and K. Mori, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," arXiv preprint arXiv:1704.06382 , Apr. 2017.

[24] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang and X. Tang, "Residual attention network for image classiiňĄcation," the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 3156–3164, Apr. 2017.

[25] M. T. Luong, H. Pham and C. D. Manning, "Effective approaches to attention-based neural machine translation," the Conference on Empirical Methods in Natural Language Processing(EMNLP), arXiv preprint arXiv:1508.04025, Aug. 2015.

[26] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate," The International Conference on Learning Representations(ICLR), arXiv preprint arXiv:1409.0473, Sep. 2014.

[27] D. Bahdanau, K. Cho and Y. Bengio, "Non-local neural networks," the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 7794–7803, Jun. 2018.

[28] S. Osher and J. A Sethian, "Fronts propagation with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," Journal of Computation physics, vol. 79, no. 1, pp. 12–49, Nov. 1988.

[29] F. Chollet et al, "Keras,", 2015, [Online]. Available: https://github.com/fchollet/keras