UENet: A Novel Generative Adversarial Network for Angiography Image Segmentation

Xiaotong Shi[1], Tianming Du[1], Shuang Chen[1], Honggang Zhang[1], Changdong Guan[2] and Bo Xu[2]

***Abstract*— Convolutional neural networks (CNNs) have been widely used in medical image segmentation. Vessel segmentation in coronary angiography remains a challenging task. It is a great challenge to extract fine features of coronary artery for segmentation due to the poor opacification, numerous overlap of different artery segments and high similarity between artery segments and soft tissues in an angiography image, which results in a sub-optimal segmentation performance. In this paper, we propose an adapted generative adversarial networks (GANs) to complete the conversion from coronary angiography image to semantic segmentation image. We implemented an adapted U-net as the generator, and a novel 3-layer pyramid structure as the discriminator. During the training period, multi-scale inputs were fed into the discriminator to optimize the objective functions, producing high-definition segmentation results. Due to the generative adversarial mechanism, both generator and discriminator can extract fine feature of coronary artery. Our method effectively solves the problems of segmentation discontinuity and intra-class inconsistencies. Experiment shows that our method improves the segmentation accuracy effectively comparing to other vessel segmentation methods.**

I. INTRODUCTION

In the diagnosis procedure of coronary artery disease (CAD), coronary angiography imaging is utilized to visualize the lumen status of coronary artery and becomes an important tool for CAD diagnosis and treatment guidance. During the processing pipeline for coronary angiography imaging, coronary artery segmentation is always the first and most important step. However, precise coronary arteries segmentation in angiography still remains a challenge due to the following characteristics of coronary angiography imaging [1, 2]: (1) plenty of overlap between different artery segments and their indistinguishable boundary; (2) very high similarity between coronary arteries and soft tissues in the background, as shown in Fig. 1. Thus, precise coronary arteries segmentation still remains a challenge.

In recent years, many image segmentation methods appeared. Among them, Fully Convolutional Network (FCN) [3] is the first CNN based methods for semantic segmentation. SegNet [4] achieves a better segmentation result with the help of information sharing between encoder and decoder. PSPnet [5] uses spatial pyramid pooling to obtain a set of feature maps with different receptive fields. After concatenating these feature maps, it produces multi-level semantic feature. U-net [6] is first proposed for cell segmentation and achieves good results for biomedical images. V-net [7] segments the volume of prostate magnetic resonance image, completing 3D segmentation. However, these methods are not suitable for coronary artery segmentation. In addition, utilizing CNN based segmentation methods is difficult to extract the fine feature to distinguish different arteries and background.

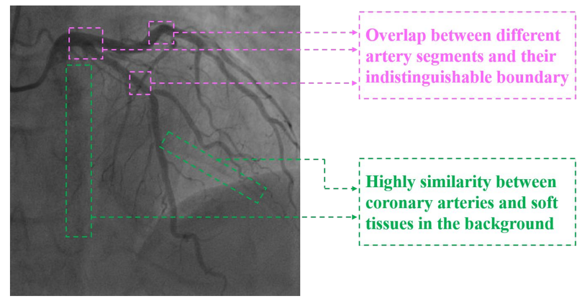

Fig. 1. Difficulties for angiography image analysis: (1) overlap between different artery segments; (2) highly similarity between arteries and irrelevant soft tissues.

In this paper, we propose a modified conditional generative adversarial networks (cGANs) [8, 9] to achieve the coronary artery segmentation in angiography imaging. Two reference conditional constraints are added to the loss function to ensure better segmentation accuracy. We build an image pyramid based on real images and synthetic images referring to the method of pix2pixHD [10], and train a discriminator for each layer. They have the same convolutional structures but different input size. The whole discriminator looks like an 'E' as shown in Fig. 2, thus we call it an E-shape discriminator. In experiment, we tackle two main problems of other segmentation methods, such as FCN and U-net. The first is blood vessels discontinuity, which means in the segmentation images, some segmentation results of coronary segments are discontinuous. The another one is intra-class inconsistencies, which means that pixels that do not belong to an artery segment appear in that artery segment. The experiment shows our model tackle these two problems effectively.

Our model has two contributions: (1) We propose an adapted cGAN network for vessel segmentation; (2) We improve segmentation accuracy compared with other segmentation methods.

[1]Xiaotong Shi, Tianming Du, Shuang Chen, and Honggang Zhang are with Faculty of Information and communication Engineering, Beijing University of Posts and Telecommunications, CHINA
Email:{sxt2015, dtm, shuang, zhhg}@bupt.edu.cn
[2]Changdong Guan and Bo Xu are with Fu Wai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, CHINA
Email:{cdguan, bxu}@citmd.com

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

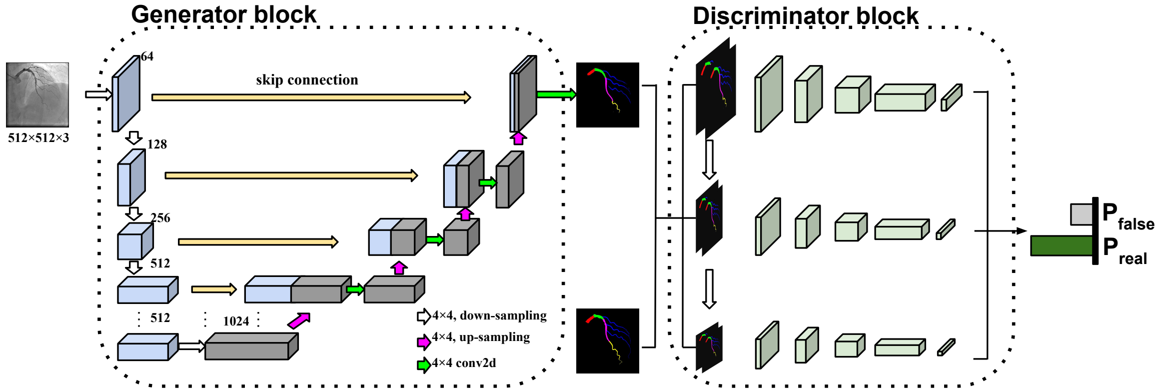

Fig. 2. The input image is 512×512 with 3 channels. In the generator part, we modified the U-net with 8 down-sampling and 8 up-sampling. In the discriminator part, we built 3-layer image pyramid and trained discriminator for each layer, which looks like an 'E', then used a 1D-tensor standing for real or false image. Our model implement an adapted U-net as the generator and E-shape discriminator, which is named UENet.

II. METHODOLOGY

A. GENERATIVE ADVERSARIAL NETWORK

For image segmentation tasks, networks usually predict the category of each pixel at the pixel-wise. The segmentation results may lack continuity and obviously differ from the ground-truth, because deep networks usually ignore the correlation between pixels. We explore to use generative adversarial mechanism [8] to solve the above problems. GANs can be treated as a competitive procedure between the generator and the discriminator until the entire model reaches Nash equilibrium and the accuracy of the discriminator is equal to 50%. However, it may become uncontrollable owing to large size of training images. To tackle above issues, [9] adds conditions to both the generator and discriminator to guide the training of the model. A condition constrain can be any supplementary information, such as label images. The process can be described as Eq (1).

$$\begin{aligned}\min_G \max_D F(G,D) =& \mathbb{E}_{(y,z)}[\log D(y|z)] \\ &+ \mathbb{E}_{(x,z)}[\log(1-D(G(x|z)))].\end{aligned} \tag{1}$$

B. Application of GANs

In our task, two problems will be solved: blood vessel discontinuity and intra-class inconsistency. To extract better feature of fine coronary arteries in angiography imaging, we modify the cGAN to achieve the artery segmentation. We only add condition constraint to the generator to simplify the network structure.

The training dataset is given as a set of pairs of corresponding images $\{x_i, y_i\}$, where x_i is a coronary angiogram and y_i is a fine annotation image. z_i is a conditional information. Our optimization process is shown as Eq (2).

$$\begin{aligned}\min_G \max_D F(G,D) =& \mathbb{E}_{(y)}[\log D(y)] \\ &+ \mathbb{E}_{(x,z)}[\log(1-D(G(x|z)))].\end{aligned} \tag{2}$$

C. UENet

As the baseline for most medical image semantic segmentation tasks, U-net is modified to adapt the angiography image segmentation.

Model structure. We apply the U-net structure with 8 down-samplings and 8 up-samplings as the generator part, which is shown in Fig. 2. Inspired by [10], the discriminator part uses a multi-scale discriminator to discriminate on three different scales and average the results. The three distinguishing scales are: the original image, 1/2 down-sampling of the original image, and 1/4 down-sampling of the original image. These three layers build an image pyramid, and train a discriminator for each layer. In addition, the residual module is used for feature extraction.

Loss functions.

(1) Binary Cross Entropy Loss: the cGAN outputs a vector of real or fake, which represents the evaluation of the entire image. We adopt a novel PatchGAN [10], which outputs an $N \times N$ matrix. Each element of this matrix, X_{ij}, has only two options, true or false. The mean value of all X_{ij} is used as the final output of the discriminator. The label is also an $N \times N$ matrix. Binary cross entropy loss is utilized in our model to avoid generating smooth transition and insufficient details. Furthermore, a sigmoid layer is added to the loss layer to improve numerical stability, that is, BCEWithLogits. In addition, we add weight parameters for labels, which avails to tackle the class imbalance issue. The objective function is as follows Eq (3):

$$\begin{aligned}\mathscr{L}_{(GAN)} =& x(1-y) \\ &+ w_1\left[\log\left(exp(-y)+exp(-x-m)\right)\right]+m,\end{aligned} \tag{3}$$

where $m = \max\{-x, 0\}$, $w_1 = 1 + x(w_2 - 1)$, w_1 is the weight parameter for label, w_2 is initialized with uniform distribution and has the same shape as x.

(2) Feature Matching Loss: we feed the generated y' and the ground truth y to the discriminator to extract features, and then calculate the pixel-wise loss of the feature. The calculation process is shown as Eq (4) :

$$\mathscr{L}_{(FM)} = W_{FM}\left\|y' - y\right\|_1, \tag{4}$$

where W_{FM} is a weight parameter.

(3) Content Loss: we feed the generated y' and the ground truth y to the VGG19 [11] to extract features, and then

calculate pixel-wise loss on the features. The process is shown as Eq (5) :

$$\mathscr{L}_{(VGG)} = W_{VGG} \left\| y^{'} - y \right\|_1, \tag{5}$$

where W_{VGG} is the proportion of each layers and is different according to different layers.

The total objective function is the sum of above three parts, which is as follows Eq (6):

$$\mathscr{L} = \mathscr{L}_{(GAN)} + \mathscr{L}_{(FM)} + \mathscr{L}_{(VGG)}. \tag{6}$$

TABLE I
DIFFERENT LOSS FUNCTIONS

PatchGAN	CRA	RIGHT
kMSE +(1-k)BCEWithLogits		
k=0.0	0.921	1.039
k=0.2	0.929	1.108
k=0.5	0.925	1.071
k=0.7	0.951	1.083
k=1.0	0.953	1.078
BCEWithLogits with weights	**0.909**	**1.006**

III. EXPERIMENT ANALYSIS

A. Datasets

The FuWai coronary angiography (CA) dataset was collected in FuWai Hospital. The dataset is divided by different angiographic views such as CRA, LAO, RIGHT, etc. In our experiments, we use the most common angiographic view, CRA (754 images) including 6 different vascular segments, and the RIGHT (3119 images) including 4 different vascular segments. For each view, we randomly split the images into training set 90% and testing set 10%.

B. Label processing

The generating processing of label data can be divided into the following steps, shown as Fig. 3:

- Extracting images randomly from the coronary angiography video as raw images.
- Labeling manually with accurate pixel-wise labels for background, catheters and coronary artery segments.
- Removing unnecessary information from the precise label images, only keeping the trunk and first-level branches, to which doctors mainly pay attention.
- Utilizing one-hot encoding for precise label image obtained in the previous step, mapping RGB value to an integer value for each pixel.

C. Training Process

We separated the training process into two serial steps for our UENet.

In first step, we converted the raw images to binary segmentation images. Binary segmentation images, which can ensure the integrity and connectivity of blood vessels, stipulate the areas of interested labeled with white pixels and others with black pixels. The obtained binary images will be used as the conditions for the second step. The model is used as the pre-training model, and its parameters are used to initialize the network.

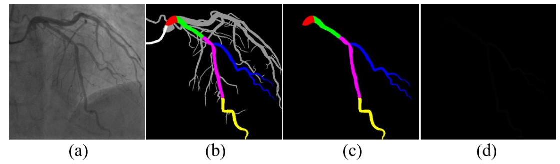

Fig. 3. (a) is the raw image, (b) is the fine-labeled image, (c) is the image remained trunk and (d) is the first-level branches and one-hot encoding image.

In second step, both binary segmentation images and one-hot encoding images, which provide intra-class label to avoid inconsistency, are regarded as the conditional variables. They superimpose with raw images as the input together. The training pipeline is shown as Fig. 4. The two pieces of supervised information can solve the corresponding two problems. In this step, we completed the conversion of raw images to semantic segmentation images.

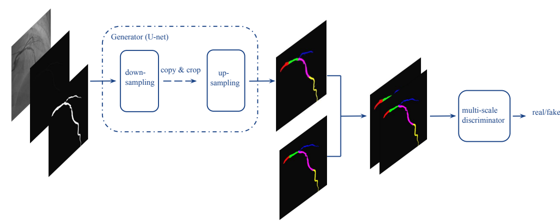

Fig. 4. The generator's input includes the blood vessel binary image, the one-hot encoding image and the raw image. The discriminator's input includes the generated image and the ground truth.

D. Implementation Detail

For training, the initial learning rate is 2×10^{-5}, which gradually dropped to 10^{-6} during training. For w_{FM}, the weight parameter, we set the value as 0.5 according to the number of discriminators. For w_{VGG}, there are 5 layers in VGG-model, so we set them as $1/32$, $1/16$, $1/8$, $1/4$ and 1 respectively. The number of training epochs is 400. Graphics processing units used in this research are NVIDIA GTX 1080Ti GPUs. Training time lasted about 5 days.

E. Quantitative Analysis

The experimental result of different angiographic views is shown in Fig. 5. Obviously, we solve the problems of vascular discontinuity and intra-class inconsistency. Compared with the results only use CNNs, the definition of the synthesis images are improved. For verifying the effectiveness of our loss function, we calculated the stable loss value in Table I. k stands for proportion of MSELoss in objective function. It did improve convergence by adding weight information to the interested blood vessel. In Fig. 6, we performed multiple sets of comparative experiments to validate our method. The column (b) is the result of using independent FCN. The image is directly transposed convolution in the last layer after down-sampling to restore the original size, which will lose a lot of detailed information due to its simple structure. Then we used the U-net with 4 layers. It is unstable and has large-scale pixels prediction errors. It is obviously seen that as the network deepening, 8-layer U-net would be better. We counted their accuracy in Table II.

We use mean pixel accuracy (MPA), shown as Eq (8), to quantify segmentation accuracy and stipulate the error pixel whose difference between the predicted pixel value and the label pixel value exceeds 5. Due to the large proportion of background pixels in the image, it will increase the average accuracy if we directly process the whole images. Therefore, in Eq (7), we divide a 512×512 picture into 64 patches with a size of 64×64, calculate the pixel accuracy on each patch, and finally take the average value as the pixel accuracy of the whole image, which is called PA_s.

TABLE II
EXPERIMENTAL RESULTS OF UTILIZING CNNS ALONE

CNNs	CRA (%)	RIGHT (%)
FCN	80.87	78.09
4-unet	81.13	79.31
8-unet	81.49	79.33
GAN+8-unet (Without one-hot image	82.90	79.53
GAN+8-unet (With one-hot image)	**83.53**	**80.47**

$$PA_s = \frac{1}{64}\sum_{i=1}^{64}\frac{p_i - e_i}{p_i}, \tag{7}$$

where p_i is the sum of pixels in each patch, e_i is the sum of error pixels in each patch.

$$MPA = \frac{1}{N}\sum PA_s. \tag{8}$$

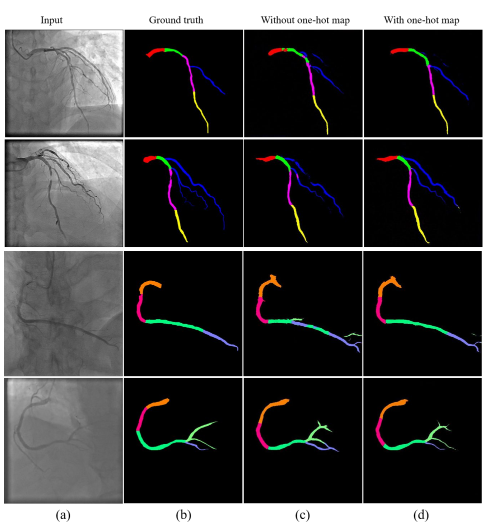

Fig. 5. Example results of our method on CRA and RIGHT. (a) is the input of network, (b) is the ground truth, (c) is the result of without one-hot map and (d) is the final result. Our method has the better performance.

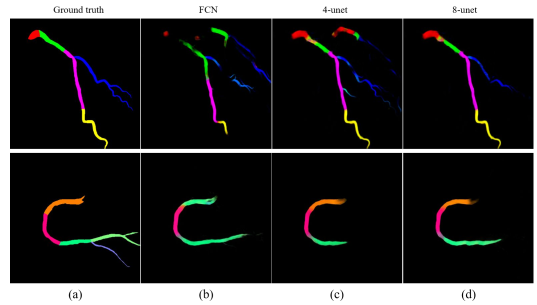

Fig. 6. Example results of utilizing CNNs alone. (a) is the ground truth, (b) is the result of FCN, (c) is the result of 4-layer U-net and (d) is the result of 8-layer U-net. FCN is not suitable for coronary angiography data due to its simple structure and with network deepening, U-net will perform better.

IV. CONCLUSIONS

In this task, we applied to adapted U-net and E-shape convolutional network and improved the angiography image segmentation result. It is important for UENet to add multiple prior information. Meanwhile, our segmentation strategy can provide a reliable reference for doctors in diagnosis, save labor and time costs.

ACKNOWLEDGMENT

The study was supported by research funding from the Chinese Academy of Medical Sciences Innovation Project (2018-12M-AI-007) and the Beijing Municipal Science and Technology Commission Project (Z181100001718113).

REFERENCES

[1] Zhai Mo, et al. Coronary Artery Vascular Segmentation on Limited Data via Pseudo-Precise Label. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.
[2] Du Tianming, et al. Morphology Reconstruction of Obstructed Coronary Artery in Angiographic Images 2019 IEEE International Conference on Visual Communications and Image Processing.
[3] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[4] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.
[5] Zhao Hengshuang, et al. Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[6] 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.
[7] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.
[8] Goodfellow, Ian, et al. Generative adversarial nets. Advances in neural information processing systems. 2014.
[9] Dai Bo, et al. Towards diverse and natural image descriptions via a conditional gan. Proceedings of the IEEE International Conference on Computer Vision. 2017.
[10] Wang Ting-Chun, et al. pix2pixHD: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs.
[11] Simonyan, Karen, and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).