AF-SEG: AN ANNOTATION-FREE APPROACH FOR IMAGE SEGMENTATION BY SELF-SUPERVISION AND GENERATIVE ADVERSARIAL NETWORK

Fei Yu[1*], *Hexin Dong*[1*], *Mo Zhang*[1], *Jie Zhao*[2], *Bin Dong*[3,1,2], *Quanzheng Li*[4,2], *Li Zhang*[1,2]

[1]Center for Data Science, Peking University, Beijing, China;

[2]Center for Data Science in Health and Medicine, Peking University, Beijing, China;
[3]Beijing International Center for Mathematical Research (BICMR), Peking University, Beijing, China;
[4]MGH/BWH Center for Clinical Data Science, Boston, MA 02115, USA.
[*] These authors contributed equally.

ABSTRACT

Traditional segmentation methods are annotation-free but usually produce unsatisfactory results. The latest leading deep learning methods improve the results but require expensive and time-consuming pixel-level manual annotations. In this work, we propose a novel method based on self-supervision and generative adversarial network (GAN), which has high performance and requires no manual annotations. First, we perform traditional segmentation methods to obtain coarse segmentation. Then, we use GAN to generate a synthetic image, on which the image foreground is pixel-to-pixel corresponding to the coarse segmentation. Finally, we train the segmentation model with the data pairs of synthetic images and coarse segmentations. We evaluate our method on two types of segmentation tasks, including red blood cell (RBC) segmentation on microscope images and vessel segmentation on digital subtraction angiographies (DSA). The results show that our annotation-free method provides a considerable improvement over the traditional methods and achieves comparable accuracies with fully supervised methods.

Index Terms— Image segmentation, Generative adversarial network, Annotation free, Deep learning

1. INTRODUCTION

With the increasing accumulation of digital medical images, automated segmentation has become one of the most important needs to quantitatively analyze the large-scale medical image datasets. Traditionally, researchers have proposed various automated segmentation models, including threshold segmentation, Level set [1], and Hessian analysis for vessel segmentation [2]. The advantage of these methods is the avoidance of massive manual interactions. However, such models are difficult to achieve accurate segmentation results. To obtain more satisfactory segmentation results, researchers often train the segmentation models in a supervised fashion.

In recent years, convolutional neural network (CNN) has made great progress in supervised semantic segmentation [3, 4, 5]. Such segmentation methods can produce highly reliable and accurate results after learning a large number of manual pixel-level annotations. However, pixel-level annotations are usually time-consuming, frustrating and even infeasible, especially in the field of medical image segmentation. To solve this problem, researchers have developed a variety of weakly supervised learning methods, expecting to use less complex annotations to train segmentation models. Dai *et al.* exploit bounding boxes to supervise convolutional networks for semantic segmentation [6]. Lin *et al.* propose to segment images with semantic scribbles [7]. In addition, Wei *et al.* demonstrate the validity of an adversarial erasing approach using only image-level annotations [8]. Recently, Yu *et al.* report an annotation-free segmentation method for coronary arteries on DSA images [9]. The method uses shape consistent generative adversarial network (SC-GAN) to generate the image foreground and the image background separately, which are later combined to train the segmentation model more effectively.

Although SC-GAN achieves an annotation-free coronary artery segmentation, it has two major limitations. First, SC-GAN needs an auxiliary labeled dataset (such as Fundus images in [9]) as the source domain data in the process of knowledge transfer. Second, to train the segmentation model, it requires synthesizing the data with additional images that have clean backgrounds which are often difficult to obtain. In this work, inspired by SC-GAN, we propose a more general framework of annotation-free segmentation, which does not require any labeled data or additional clean-background images. We call this new framework AF-SEG, and the details of the method will be introduced in the next section. We evaluate the performance of the proposed method on sickle cell disease (SCD) RBC dataset and DSA dataset by several metrics. Qualitative and quantitative results show that our

Corresponding author: Li Zhang, zhangli_pku@pku.edu.cn

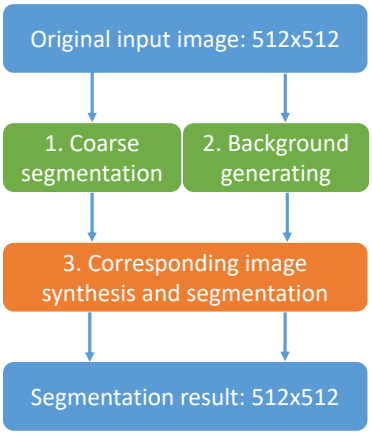

Fig. 1. The flowchart of the proposed method

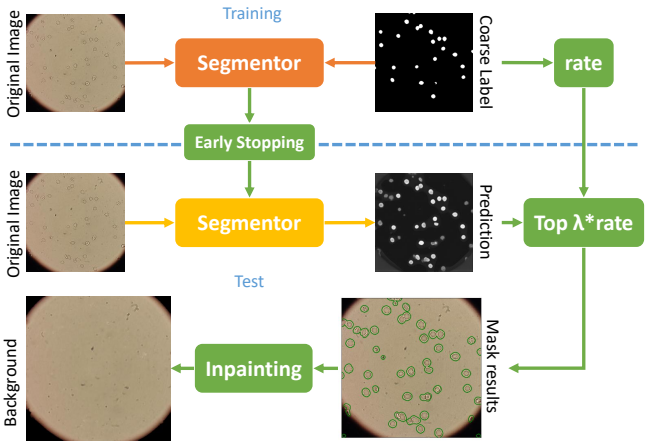

Fig. 2. The illustration of the background generating method. The mask (green circles) produced by our method are larger in order to mask-out entire false negatives.

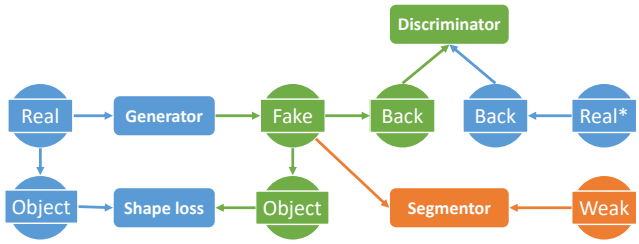

Fig. 3. The illustration of corresponding image synthesis and segmentation method.

method provides a considerable improvement over the traditional methods and achieves comparable accuracies with fully supervised methods.

2. METHOD

In this work, we first perform traditional annotation-free methods to obtain coarse segmentations. We then synthesize an image corresponding to the coarse segmentation using GAN. Finally, we use the coarse segmentation as a pixel-level annotation of the synthetic image, and supervisely train a high-quality segmentation model. Inspired by SC-GAN, the foreground and background of the synthetic image are generated separately, and then the two are fused, which ensures that the foreground of the image is consistent with the coarse segmentation. The flow chart of our method is shown in Figure 1, and each step will be described in detail below.

2.1. Coarse Segmentation

We first use the traditional annotation-free method to generate coarse segmentation results. For example, in cell segmentation, we adopt the classic Level set method. In vessel segmentation, Hessian analysis is also a good choice. Since the coarse segmentation obtained in this step will be used as the annotation of the later supervised learning, we apply a morphological erosion operation to reduce the false positive rate of the coarse segmentation.

2.2. Background Generating

In this step, we will remove the false negative response left by the previous step and generate a clean background image for final synthesis. Following the idea of Deep Image Prior (DIP), we report a new strategy called Deep Segmentation Prior (DSP) to enrich the coarse segmentation from Section 2.1 (to remove the false negatives). As shown in Figure 2, to perform DSP, we train a simple segmentation network, whose input image is the original image and the label is the corresponding coarse segmentation. DSP believes that the model will detect all label-like structures in the early stage of the training process. Therefore, we apply the early-stop strategy during training. The stop time of the training is controlled by a hyper-parameter λ, which represents the multiple of the size of the coarse segmentation that we want the final prediction to be. In this experiment, we simply set $\lambda = 2$. Using the results of DSP, we can effectively mask-out the false negatives in the background. Finally, we use image inpainting techniques, such as DIP, to generate clean (object-free) background images.

2.3. Corresponding Image synthesis and segmentation

We integrate image synthesis and segmentation into an end-to-end model. Figure 3 shows three major components of the model, including the generator, the discriminator and the seg-

menter. In the following text, we will explain the meanings of the terms in Figure 3, such as Real, Fake, Real*, and etc.

The generator (G) uses U-Net as its network backbone, whose input is an original image (Real) and output is a synthetic image (Fake). We introduce a shape consistent loss (L1 loss) to ensure that Real and Fake are consistent within the coarse segmentation (Weak).

$$\mathcal{L}_{shape}(G) = \mathbb{E}_{x\sim p_{data}(x)}[||label * (G(x) - x)||_1] \quad (1)$$

where $label$ represents the coarse segmentation (Weak).

The function of the discriminator (D) is to ensure that the synthetic image can have a clean background outside of the coarse segmentation (Weak). Firstly, we use the coarse segmentation to get the background region of the synthetic image (Fake) and the clean background image (Real*) generated in Section 2.2,

$$Fake_{bg} = \neg label * G(x), Real_{bg} = \neg label * Real^* \quad (2)$$

Using the adversarial training, the discriminator will remove the possible objects in the background of Fake, and the adversarial loss can be expressed as,

$$\begin{aligned}\mathcal{L}_{GAN}(G, D) = &\mathbb{E}_{x\sim p_{data}(x)}[logD(Real_{bg})] \\ &+ \mathbb{E}_{z\sim p_{data}(z)}[log(1 - D(Fake_{bg}))]\end{aligned} \quad (3)$$

The network architecture of the segmentor (S) uses the classical U-Net, whose objective function is,

$$\mathcal{L}_{seg}(S) = -[ylog\hat{y} + (1 - y)log(1 - \hat{y})] \quad (4)$$

where $\hat{y}$ is the prediction and y the Weak label.

Finally, the overall objective function is,

$$\mathcal{L}(G, D, S) = \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{seg}(S) + \mu\mathcal{L}_{shape}(G) \quad (5)$$

where μ is a hyper-parameter and we set $\mu = 50$ in our experiments.

3. EXPERIMENTS AND RESULTS

3.1. Data and experiment details

In this section, to evaluate the generalization and effectiveness of our method, we perform experiments on two datasets: 1) The SCD RBC dataset. Referring to [10], we use 308 raw microscope images of 5 different SCD patients. The raw image resolution is 1920×1080. Following [10], we preprocess them by removing two-side margins and resizing them into 512×512. 2) The DSA dataset. Referring to [9], we use 1092 coronary angiographies as our experimental data. Following [9], we preprocess them by median-filtering and contrast-limited adaptive histogram equalization [11]. In addition, both two datasets are randomly split into training set (50%), validation set (20%) and test set (30%). Finally, we perform grayscale transform and randomly choose 256×256 patches as the inputs of our model.

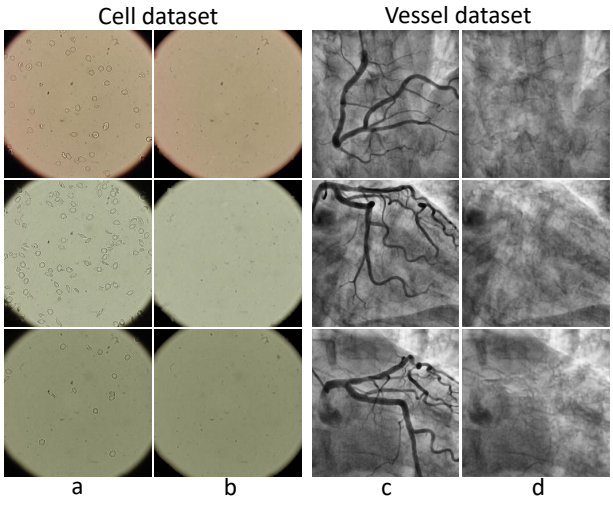

Fig. 4. The results of background synthesis. (a) Original cell images, (b) Synthetic cell background, (c) Original vessel images, (d) Synthetic vessel background

In all experiments, we use the Adam optimizer [12] with a learning rate of 2e-4. The learning rate remains constant for the first 50 epochs and then linearly decreases till 0 for another 50 epochs. In addition, the network defined in Figure 3 uses instance normalization [13] instead of batch normalization [14], LeakyReLU instead of ReLU.

3.2. Background and Image Synthesis Results

Figure 4 shows some examples of the background synthesis. We can find that the synthetic background is very realistic and almost all the objects are eliminated, which just proves the effectiveness of our background generating method.

Figure 5 shows some examples of the corresponding image synthesis. We can see that the synthetic images generated not only have a realistic background but also preserve the main object structures corresponding to the coarse labels.

3.3. Image Segmentation Results

Table 4 compares the quantitative results of different methods on two datasets. The traditional methods (Level set and Hessian Analysis) have Dice scores of 0.839±0.070 and 0.636±0.046 respectively, which are the baseline results. The proposed AF-SEG has higher Dice scores (0.915±0.029 and 0.820±0.028), which demonstrates its effectiveness. The Supervised U-Net has Dice scores of 0.967±0.011 and 0.913±0.017 respectively, which is trained with manual annotations. Figure 6 shows some results of image segmenta-

Table 1. The quantitative results of image segmentation on two datasets. The Level set [1], and Hessian Analysis [2] present the traditional annotation-free methods. The Supervised U-Net is trained with manual annotations, which should be the upper bound of our method.

Datasets	Cell dataset			Vessel dataset		
Annotation-free	✓		✓	✓		✓
Methods	Level Set	U-Net	**Proposed**	Hessian Analysis	U-Net	**Proposed**
Accuracy	0.982±0.016	0.996±0.003	**0.990±0.007**	0.927±0.014	0.976±0.007	**0.953±0.010**
Precision	0.894±0.072	0.961±0.018	**0.906±0.029**	0.943±0.057	0.892±0.025	**0.833±0.036**
Recall	0.806±0.124	0.974±0.012	**0.926±0.055**	0.481±0.048	0.937±0.021	**0.810±0.047**
Dice	0.839±0.070	0.967±0.011	**0.915±0.029**	0.636±0.046	0.913±0.017	**0.820±0.028**

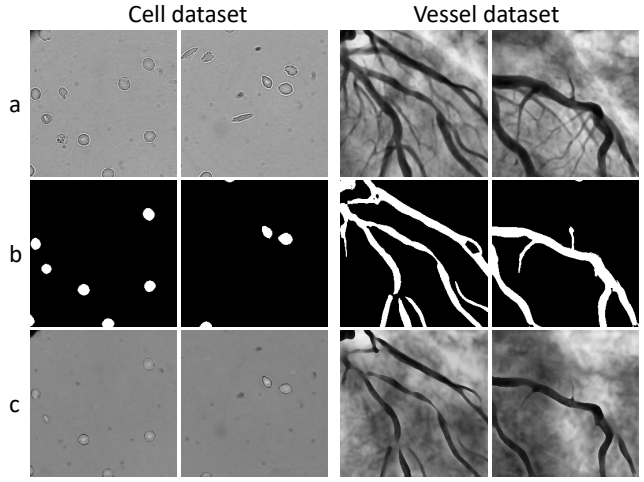

Fig. 5. The results of corresponding image synthesis. (a) Original images, (b) Weak labels, (c) Synthetic images

tion. Compared to traditional methods, the proposed method shows more accurate segmentation.

4. CONCLUSION

In this paper, we propose an annotation-free segmentation method that improves the segmentation accuracy of traditional methods. Using adversarial training, we effectively synthesize a target image corresponding to the traditional coarse segmentation. The synthetic images and the coarse segmentation allow us to better train segmentation models. The experimental results in both cell and vessel segmentation tasks show that our method could obtain a significant improvement compared to traditional methods, demonstrating the good generalization and effectiveness of our method. Of course, our method has limitations, despite the fact that it requires no manual labeling and meets the requirements in a wide range of medical quantitative analysis. For example, our method is slightly inferior to the fully supervised methods in some applications that require additionally high accuracy

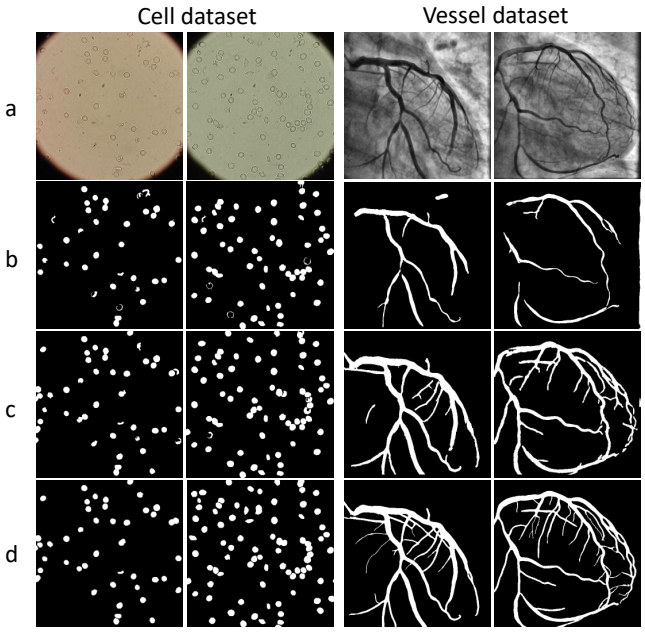

Fig. 6. The results of image segmentation. (a) Original images, (b) Traditional methods, (c) Proposed methods, (d) Groundtruth.

of image details. In the future, improvement of the method along this dimension could be fruitful.

5. ACKNOWLEDGMENTS.

This work was supported by Natural Science Foundation of China (NSFC) under Grants 81801778, 71704024, 11831002; National Key R&D Program of China (No. 2018YFC0910700); Beijing Natural Science Foundation (Z180001).

6. REFERENCES

[1] Chunming Li, Rui Huang, Zhaohua Ding, J Chris Gatenby, Dimitris N Metaxas, and John C Gore, "A

level set method for image segmentation in the presence of intensity inhomogeneities with application to mri," *IEEE transactions on image processing*, vol. 20, no. 7, pp. 2007–2016, 2011.

[2] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*. Springer, 1998, pp. 130–137.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[6] Jifeng Dai, Kaiming He, and Jian Sun, "Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1635–1643.

[7] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3159–3167.

[8] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang, "Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7268–7277.

[9] Fei Yu, Jie Zhao, Yanjun Gong, Zhi Wang, Yuxi Li, Fan Yang, Bin Dong, Quanzheng Li, and Li Zhang, "Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images," *arXiv preprint arXiv:1907.11483*, 2019.

[10] Mo Zhang, Xiang Li, Mengjia Xu, and Quanzheng Li, "Rbc semantic segmentation for sickle cell disease based on deformable u-net," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 695–702.

[11] Karel Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.

[12] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[13] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[14] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.