

Skin Lesion Image Segmentation Based on Adversarial Networks

Ning Wang¹, Yanjun Peng^{1,2,*}, Yuanhong Wang¹ and Meiling Wang¹
¹College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, China, 266590
[e-mail: skdwnlyjf@gmail.com]
²Shandong Province Key Laboratory of Wisdom Mining Information Technology, Shandong University of Science and Technology, Qingdao, China, 266590
[e-mail: pengyanjuncn@163.com]
*Corresponding author: Yanjun Peng

Received September 18, 2017; revised December 16, 2017; accepted January 18, 2018; published June 30, 2018

Abstract

Traditional methods based active contours or region merging are powerless in processing images with blurring border or hair occlusion. In this paper, a structure based convolutional neural networks is proposed to solve segmentation of skin lesion image. The structure mainly consists of two networks which are segmentation net and discrimination net. The segmentation net is designed based U-net that used to generate the mask of lesion, while the discrimination net is designed with only convolutional layers that used to determine whether input image is from ground truth labels or generated images. Images were obtained from “Skin Lesion Analysis Toward Melanoma Detection” challenge which was hosted by ISBI 2016 conference. We achieved segmentation average accuracy of 0.97, dice coefficient of 0.94 and Jaccard index of 0.89 which outperform the other existed state-of-the-art segmentation networks, including winner of ISBI 2016 challenge for skin melanoma segmentation.

Keywords: Skin lesion, image segmentation, convolutional neural networks, adversarial network.

1. Introduction

Skin cancer is a common type of cancer in our daily lives. The skin cancer can be classified into melanoma and non-melanoma and melanoma is the most dangerous type of skin cancer [1]. Early diagnosis is particularly important since melanoma can be cured with excision if detected early [2]. Melanoma is enable to early detection by expert visual inspection because lesions occurring on the surface of the skin and commonly different from around skin color. Skin lesion segmentation is a crucial point in processing skin image which is used to get rid of irrelevant areas of clinical and dermoscopic images. Skin lesion segmentation has many difficulties, such as low contrast, oil marks interference, tiny area, hair occlusion. See Fig. 1. These difficulties could affect the accuracy of the segmentation result. For these difficulties, some researchers have tried to solve it [3]. Traditional methods could not achieve an ideal performance due to the diversity of shape, colors, size of skin lesion. With the development of deep learning, some architectures have shown immense potential in lesion detection tasks, such as U-net [4], SegNet [5]. Applying deep learning to medical image segmentation has become a research hotspot.

In this paper, we design a segmentation structure based on generate and adversarial network and some existed skills, such as batch normalization and adding noise. A new loss function is developed to help skin lesion segmentation. Through a mass of experiments, we get a set of well-performed parameters and a better segmentation result compared with existed methods.

The rest of the paper is organized as follows. Section 2 describes the related work. Convolutional neural network model is proposed in Section 3. Section 4 gives out the experiment results. The conclusion and future work are described in Section 5.

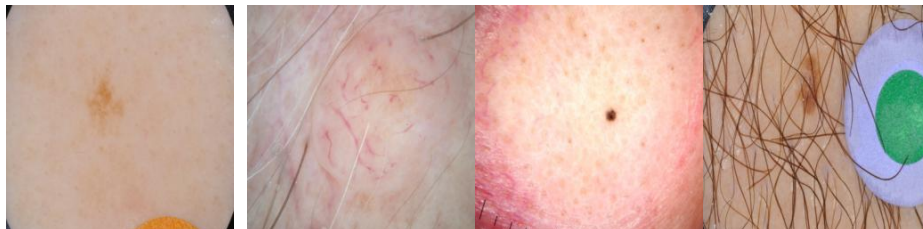


Fig. 1. Examples of skin lesion images.

From left to right: blurry border, low contrast, tiny area, hair occlusion

2. Related work

Nowadays, many image processing technologies have been used to assist doctor to diagnose lesion area. In paper [6], the fuzzy c-means (FCM) thresholding technique is used to segment and localize the lesion. In paper [7], a mean shift based gradient vector flow (GVF) snake algorithm that drives the internal and external energies towards the correct direction which resulted in a good segmentation. In paper [8], a fast and unsupervised approach to border detection in dermoscopic images of pigmented skin lesions based on the statistical region merging algorithm was proposed. Besides these direct processing algorithm, some researchers try to improve skin image quality, such as color and contrast enhancement [9]. These traditional methods can be categorized in three groups of thresholding, active contours, and region merging methods. As we mentioned in chapter introduction, to get a good result

is not easy for certain skin images. Traditional segmentation methods could be powerless in processing all kinds of lesion images. Traditional segmentation methods based image intensity could achieve high accuracy, but for low contrast images, it is very hard to get a better result. A pre-processing or post-processing is needed to correct the result. Also, for methods based edge and contour, a better accuracy is hard to get due to the lack of clear lesion borders. For region merging methods, tiny areas are difficult to get a good result. In a word, not a single traditional method could almost handle all kinds of skin images well.

Deep learning brings brand new ideas in image segmentation. As we can see, convolutional neural network has made a great success in many fields. In image segmentation task, FCN [10] is an important beginning on semantic image segmentation using convolutional neural network. FCN changes fully connected layer to convolution layer and uses deconvolution to restore to the original image resolution. FCN has some shortcomings: its result is not fine in border due to the lack of spatial consistency. Starting from FCN, some architectures based fully convolutional network were proposed. Among them, U-net and SegNet are more popular in segmentation. U-net architecture is popular convolutional network for biomedical segmentation and use deconvolution to upsample, while SegNet uses uppooling to upsample. There are also some architectures combined with non-parametric method, like Matching-CNN Meets KNN[11] combining with K nearest neighbor method to solve human parsing task. In some segmentation network[12, 13], fusion operation is also a effective skill.

Generate and adversarial network (GAN) [14] has achieved great success in image generation and image style transfer from its birth to the present time[15, 16]. Phillip Isola *et al* proposed a method using conditional adversarial networks to implement image to image operation [17]. Recently, semantic segmentation using adversarial training makes some progress, Pauline Luc *et al* used adversarial networks [18] to get a semantic segmentation which motivated us to make an improvement in biomedical segmentation.

3. Skin lesion segmentation with adversarial network

3.1 Training with adversarial network

Actually, we can describe generate and adversarial network as a method to learn distribution of training data. In papers[19, 20], more details can be found about learning distribution of data, like zero-shot learning. Theoretically, any network that can output an image can be a generate network. Therefore, many segmentation networks can be described as a generator. Discrimination network is essentially a two-category network so the whole process is generation network tries to learn the distribution of training data and discrimination network tries to tell apart whether a mask if from ground truth labels or generated images. We detail segmentation network and discrimination network separately in next subsection.

Fig. 2 shows the overall framework in incorporating adversarial process to the skin lesion images segmentation. The top left corner is the images and corresponding ground truth labels to input discrimination network. The top right corner is segmentation network which gets skin images of three channels input and output skin lesion masks of one channel. The lower right corner is discrimination network which gets ground truth labels and generated images by segmentation network input and output probability of ground truth label.

In training process, the segmentation network S and discrimination network D are trained by back propagation from the L2 loss. First, fix S and train D for one step using gradients computed from the loss function. After D is trained well, fix D and active S . Then train S for

one step using gradients computed from the same loss function passed to S from D . The discrimination network alternate segmentation network until reach to maximum number of epoch. After trained done, we get a powerful segmentation network to segment skin lesion image and a powerful binary classification network. This training process shows min-max game iconically.

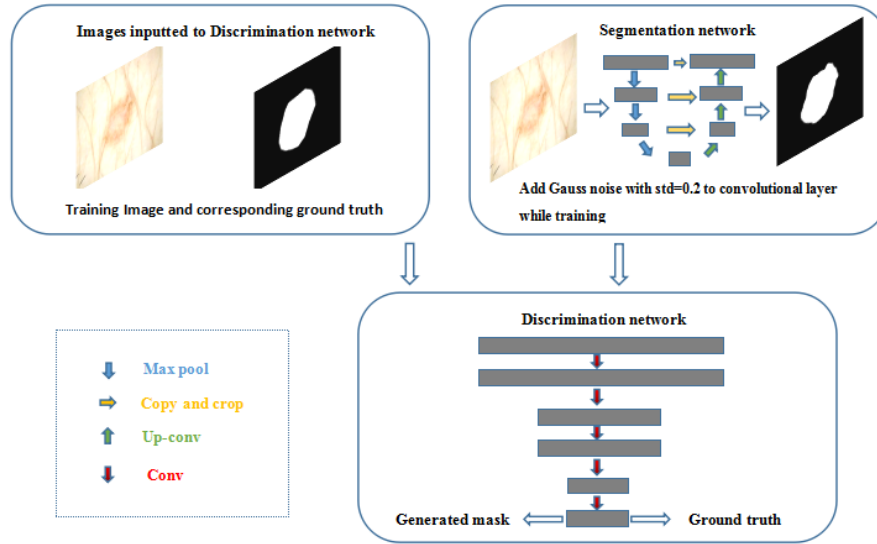


Fig. 2. Whole framework of skin lesion segmentation.

3.2 Object function

Let S, D represent segmentation network and discrimination network, respectively. Let x represents image and y represents corresponding ground truth label. So we get $S: \{x \rightarrow y\}$, mapping x to y . $\{x, y\} \rightarrow \{0, 1\}$ represents the mapping $\{x, y\}$ to $\{0, 1\}$ which 0 denotes image is from generated masks, 1 denotes image is from ground truth labels. From this mapping, we can also see that, we put both image and ground truth label to the discrimination network.

Loss function of GAN is cross entropy which don't give us a good segmentation result on skin lesion segmentation task. In paper [21], authors consider that using cross entropy could let segmentation network stop optimizing the image that discrimination network inferred as ground truth label, even the generated mask is not "realistic" compared with ground truth label. Then, L2 loss was used to be loss function which is LSGANs. Inspired by LSGANs, we use L2 loss as our gan loss function which is mean squared error mse . See formula (1).

$$mse = (d_pre - d_true)^2 \tag{1}$$

Where d_pre represents gan's output and d_true represents the right classification which is 0 or 1, respectively.

For discrimination network's loss function, we still use binary cross entropy.

So, we can get formula(2).

$$\min_D L_{SAN}(S, D) = -E_{x, y \sim p_{data}(x, y)}[\log D(x, y)] - E_{x \sim p_{data}(x)}[\log(1 - D(S(x)))]$$

$$\min_S L_{SAN}(S, D) = \frac{1}{2} E_{x \sim p_x(x)}[D(S(x))^2] \quad (2)$$

Which $S(x)$ represents mask generated by segmentation network. The Segmentation network's object is to generate masks that are indiscernible to the real data. The discrimination network's object is to reduce the probability of judgment failure. That is what we called adversarial networks.

Previous research has demonstrated that it is effective to combine GAN objective with a traditional loss, such as L2 loss [22]. The discrimination network's task is unchanged. The segmentation network is not only need to fool the discrimination network, but also to be near corresponding ground truth label. We test L2 loss, binary cross entropy, L1 loss and found L2 loss and binary cross entropy get a blurring border. So, we adopt L1 loss.

$$L_{SAN}^* = L_{SAN}(S, D) + \lambda L_S(S) \quad (3)$$

At last, we get the next objective function:

$$L_S(S) = E_{x, y \sim p_{data}(x, y)} mae(S(x), y) \quad (4)$$

$$mae = |y_{pre} - y_{true}| \quad (5)$$

Where y_{pre} and y_{true} represent segmentation network's output and label. Its size is $256 \times 256 \times 1$. λ balances segmentation network and L1 loss.

3.3 Segmentation network

Our segmentation network is a U-net network based architecture proposed by Olaf Ronneberger *et al.* U-net essentially is a fully convolutional network (FCN) and make great success in segmentation of biomedical images. Given that the shape of U-net, the left part capture context and the right part get the precise localization. The left part was called contracting path and the right part was called expanding path. The contracting path follows the typical architecture of a convolutional network and tries to capture context to propagation. The important part is expanding path that is upsampling in this part. High resolution features from the contracting path are combined with the upsampled output so that pixels could be located precisely. In the upsampling part there are a large number of feature channels, which allow the network to propagate context information to higher resolution layers.

One modification in our architecture is that we add batch normalization layer in the network. Batch normalization allows us to use much higher learning rates and be less careful about initialization, and in some cases eliminates the need for Dropout[23]. In training process, parameters have been changing and the distribution of every layer except input layer in the network have been changing. Update of former layer's parameters result in the change of distribution of latter layer and that is called "Internal Covariate Shift". In a word, batch normalization solved vanishing gradient problem and exploding gradient problem in back propagate phase.

The other modification in our architecture is that we inject gauss noise in contracting path. About this modification, we have two considerations. First, researches [24, 25] have

demonstrated that appropriate noise can be useful in training. Injection of gauss noise can be seen as an increment of training data randomly and it is effective to overcome overfitting problem. Based our experiments on skin lesion segmentation tasks, gauss noise with $std(\text{variance})=0.2$ can improve average accuracy. Accuracy with gauss noise is 1% higher than accuracy without noise on average. Second, original GAN is hard to train well and we also saw gradient problem while tested GAN experiments. The crucial reason is Kullback–Leibler divergence and Jensen-Shannon divergence. When the discrimination network is trained well enough, optimize generator network's loss function is equivalent to optimize JS divergence of real data's distribution and generated mask's distribution. Because these two distributions almost have no overlap, no matter how far between two distributions, their JS divergence is always $\log 2$. At last, gradient of generator is 0 and vanish. About this problem, Jose Caballero *et al* proposed a scheme which is add noise to generated image and ground truth label to enforce them to produce overlap [23]. Although we adopt L2 loss as objective function, adding noise should be still effective. Experiments has proved that is effective.

Fig. 3 shows the segmentation network architecture. We use less feature maps from 32 to 512 compared with original U-net's features from 64 to 1024 and that is why we abandon dropout layer. We can't predict where the best location to add noise is and we just add noise in second layer. The input shape has three channels and output shape has one channel.

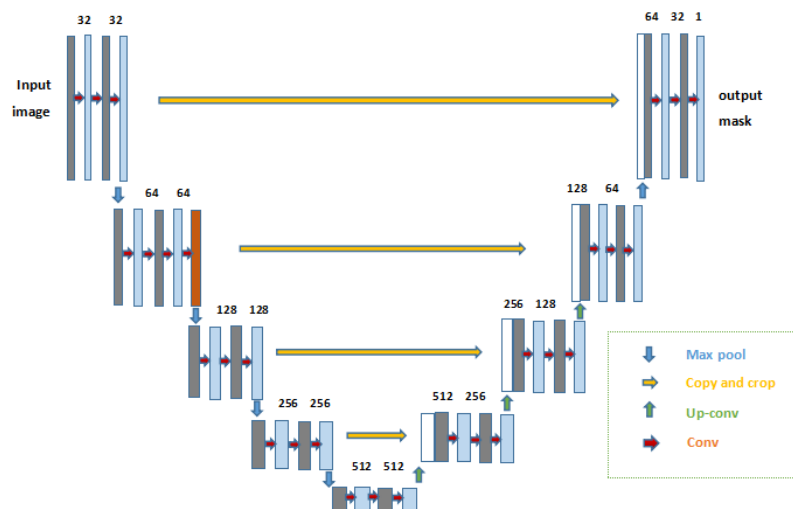


Fig. 3. Segmentation architecture (based U-net). Gray boxes represent convolutional layer, white boxes represent copied feature maps, blue boxes represent batch normalization layers, orange box represents gauss noise layer. The arrows denote the different operations.

3.4 Discrimination network

The discrimination network basically is a binary classification network. Our discrimination network architecture is shown in **Fig. 4**. Although in paper [17], PatchGAN is proposed and get a good performance, we don't get what we wanted in our experiments. Inspired by DCGAN [27], the discrimination network is designed with only convolutional layers and a fully connected layer. Using stride convolutions to replace max pooling layers, for this can allow the discrimination network to learn its own spatial downsampling. We explored the network with or without the max-pooling layers and found stride convolutions really do

better than the max-pooling. For discrimination network, we put image and corresponding ground truth label into the network, so we get four channels to input. The output is just a number represents the mask is from generated images or ground truth labels. Like segmentation network, we don't set many feature maps due to limit of CPU memory.

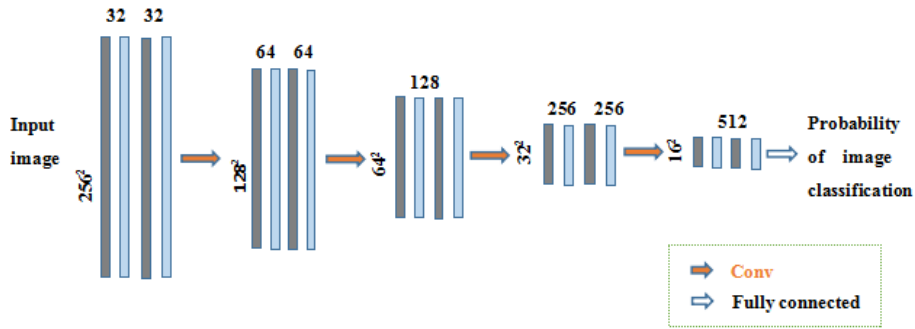


Fig. 4. Discrimination architecture. Gray boxes represent convolutional layer, blue boxes represent batch normalization layers.

4. Experiments

4.1 Dataset and processing

We use dataset from first task of ISBI 2016 challenge “Skin Lesion Analysis Toward Melanoma Detection” which consists of 900 train images and corresponding ground truth labels [28]. Also 379 test images with ground truth labels was provided.

Considering that size of both training images and test images various a lot and limit of CPU, we resize all images and corresponding ground truth labels to 256×256 . This operation not only preserve sufficient information, but also save training time.

Due to the high variation in image contrast or colors, we perform per image normalization.

Given that an image x we normalize it with $x^* = \frac{x - \bar{x}}{std(x)}$, where \bar{x} and $std(x)$ represent

mean and variance of image x , respectively. Considering the color, shape and size of true lesion, we are conceived that data augmentation could not be appropriate and that is why we abandon data augmentation.

4.2 Training setting

The segmentation net encodes input images using 10 convolutional layers with 4 max-pooling layers. Symmetrically, 4 upsample layer with 10 convolutional layers are used in decode phase. As for discrimination net, it only consists of 10 convolutional layers and 5 batch normalization layers.

We adopt Adam as optimizer and set learning rate as 0.0002, β_1 as 0.5. We trained the whole network for 20 epoches. We set mini-batch as 16. About gauss noise in segmentation, we test a wide range and set gauss noise with $std(\text{variance})=0.2$. When training involves discrimination network, for each mini-batch, we train 4 steps on segmentation network for each step on discrimination network. We train the whole network with 8G memory and CPU. For each epoch, it cost approximately 80 minutes. Total number of parameters is 1257k. The summary is as follows in [Table 1](#).

Table 1. Summary of parameters settings

lr	2e-4	mini-batch	16
beta_1	0.5	std	0.2
epoch	20	λ	0.125

4.3 Evaluation metric

For quantitative evaluation of our method, three commonly used metrics for classification problems are used. These metrics are sensitivity, specificity and accuracy. The segmentation problem can be stated as a classification problem, where each pixel is classified into one of the two classes of lesion (positive) and normal skin (negative). As a result, by comparing the output mask with the ground-truth, the classification metrics can be calculated as:

$$sensitive = \frac{TP}{TP + FN} \quad (6)$$

$$specificity = \frac{TN}{FP + TN} \quad (7)$$

$$acc = \frac{TP}{TP + FP} \quad (8)$$

Where TP , TN , FP and FN denote the number of true positive, true negative, false positive and false negative classified pixels respectively.

For segmentation task, we also use two common metrics to evaluate the methods. These metrics are Jaccard index and dice coefficient. These metrics can be calculated as:

$$dice = \frac{2n_t}{n_x + n_y} \quad (9)$$

$$jaccard = \frac{n_t}{n_x + n_y - n_t} \quad (10)$$

Where n_x and n_y denote the number of x image pixels, y image pixels and intersecting pixels within x, y .

In summary, the performance of the proposed method is compared to other methods using same metrics on the pixel-level, as defined in the challenge: Jaccard index, accuracy, sensitivity, specificity and dice coefficient.

4.4 Results and comparison

All comparisons are performed on 379 images and corresponding ground truth labels. We show the respective contribution of gauss noise and adversarial net in [Fig. 5](#) where AC, DI, JA, SE, SP represent accuracy, dice coefficient, Jaccard index, sensitivity and specificity. From this chart, it is clear that adversarial net make more contribution to segmentation result and the final result is not just a linear combination of modifications.

Batch normalization's contribution is not easy to show from chart, but we actually feel its power in experiments. Without batch normalization, it is more prone to be unstable while training, especially choosing of parameters.

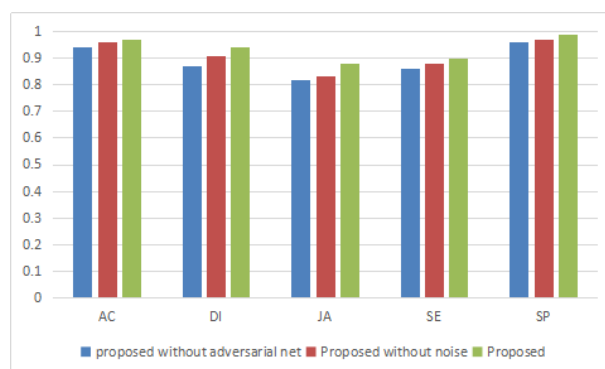


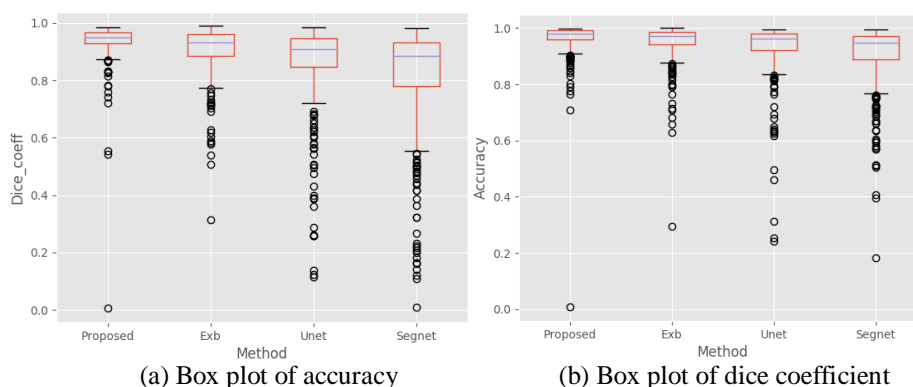
Fig. 5. Comparison metrics of modifications

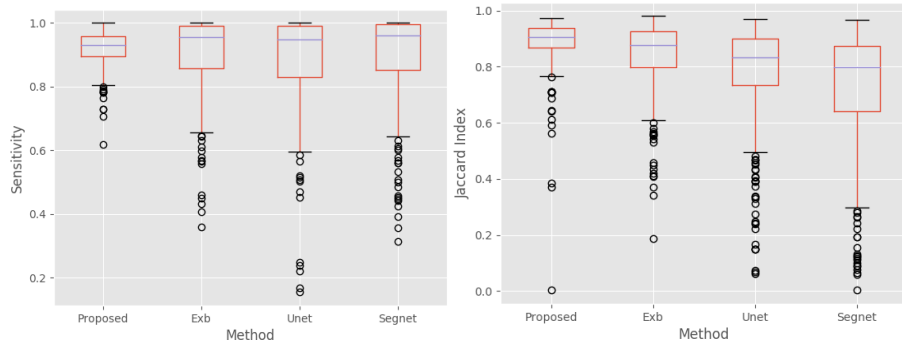
We compared proposed method with U-net, SegNet and ExB which is the winner of the challenge. Comparison result are shown in Table 2. From Table 2, we can see clearly our method outperform the other methods in a mass except sensitivity.

Table 2. Comparison of metrics

	AC	DI	JA	SE	SP
SegNet	0.90	0.84	0.73	0.86	0.92
U-net	0.93	0.88	0.78	0.85	0.96
ExB	0.95	0.91	0.84	0.91	0.97
Proposed	0.97	0.94	0.88	0.90	0.99

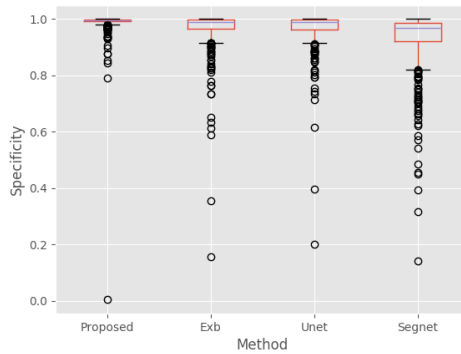
Box plot can show the distribution of data, so we use box plot to curve the distribution of five metrics. On each metric, our method has highest median and mean value while maintaining the lowest variance. From Fig. 6, we can see that our method has a highest value with robust performance.





(c) Box plot of Jaccard index

(d) Box plot of sensitivity



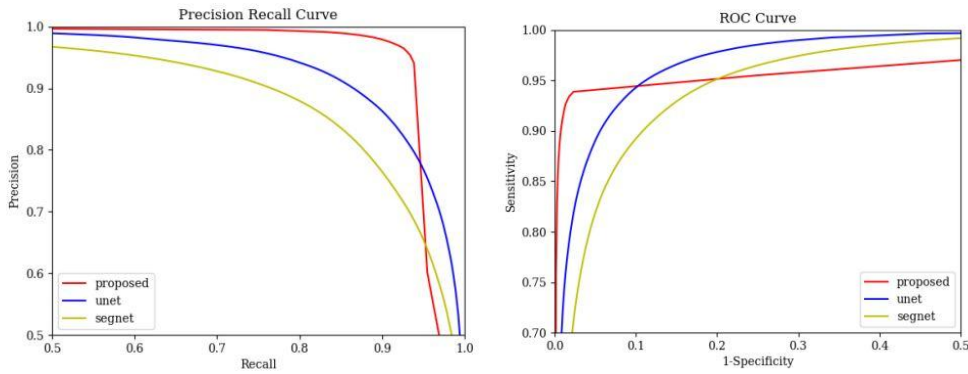
(e) Box plot of specificity

Fig. 6. Box plot of each metric

An image segmentation is also a classification task, for us segmentation network is a binary classifier which classify pixels into foreground and background. Usually, ROC(receiver operating characteristic)curve, PR(precision recall)curve and AUC(Area Under Curve) are used to estimate a classifier. We also use these curves to estimate our method and compared with other method. The result are as follows.

Table 3. AUC value of methods

	Proposed method	U-net	SegNet
AUC	0.968	0.978	0.960



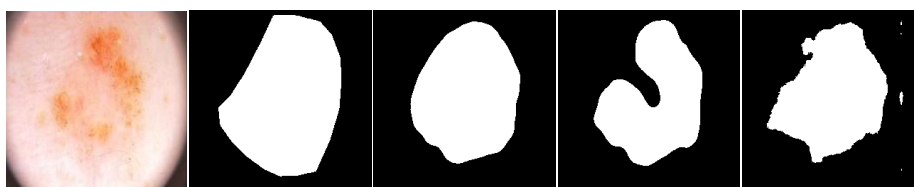
(a) ROC curve

(b) PR curve

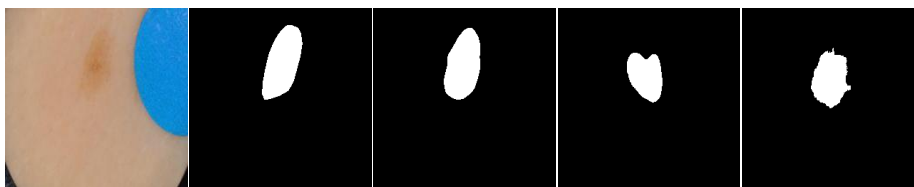
Fig. 7. Receiver Operating Characteristic (ROC) curve and Precision and Recall (PR) curve for various methods.

So, we have three metrics to analysis which are AUC, ROC curve, PR curve. AUC used to estimate a classifier. From [Table 1](#) we can see that our method outperforms the other methods but it is unexpected that our method is not the highest AUC value as shown in [Fig. 7](#). Also, in ROC curve, it is hard to say that our method is the best. We consult some research and find an explanation about why this happens. ROC curve is used to characterize the relationship between *TPR* and *FPR*. PR curve is used to characterize the relationship between precision and recall. Particularly, if true negative is not much valuable to the problem, or negative examples are abundant. Then, PR curve is typically more appropriate [\[29\]](#). In a word, number of positive samples is far less than number of negative samples. For more evidence in detail, we can see that from dataset of ISBI 2016 challenge.

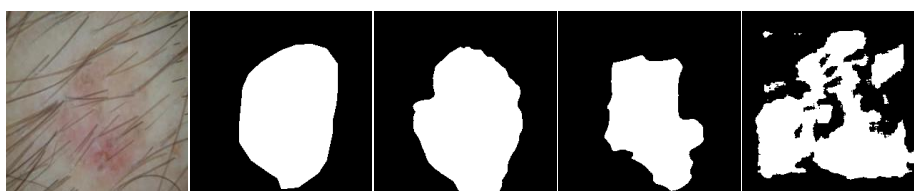
[Fig. 8](#) shows segmentation results of different methods. We choose some representative images which means it is hard to get a better segmentation results. We can see that our method gets a better segmentation results on these representative images from [Fig. 8](#), particularly on images 8-c, 8-d, 8-f which U-net and SegNet are a mass while our methods still get a stable output.



(a) Segmentation result of a blurred edges



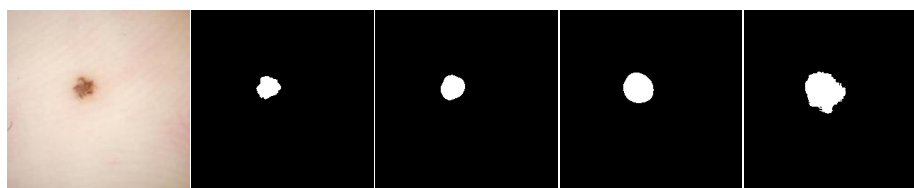
(b) Segmentation result of a low contrast lesion



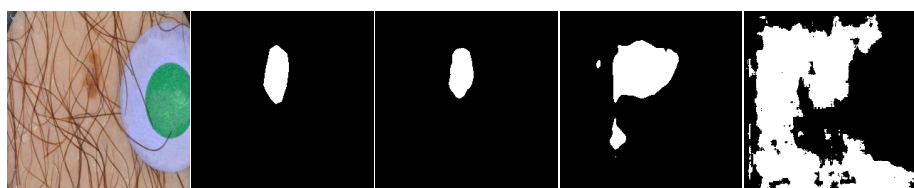
(c) Segmentation result of a hair-occluded lesion



(d) Segmentation result of a very low contrast and hair-occluded lesion



(e) Segmentation result of a small lesion



(f) Segmentation result of a hair-occluded and small lesion

Fig. 8. Examples of lesions segmentation, from left to right: image, ground truth label, proposed, U-net, SegNet

5 Conclusion and Future Work

In this paper, we have presented a new method to segmentation of skin lesion, which adopts adversarial network to assist segmentation to train. It helps to improve average accuracy of segmentation result compared with the best result to the best of our knowledge. Even for some tough images, we still can get stable output and that shows our method has a good robustness. Also through the experiments we find a troublesome problem which is there are more than one suspicious lesion spots in image. The problem is not all these lesions are needed to process which means some lesion is not a melanoma. For these images, our method is also powerless. In paper[30], an architecture with two discrimination networks gives us a inspiration. So, in the future work, we try to solve this kind of problem and optimize the whole network to run faster and more stable.

References

- [1] R. J. Mermelstein and L. A. Riesenber, "Changing knowledge and attitudes about skin cancer risk factors in adolescents," *Health Psychology*, vol. 11, no. 6, pp. 371, 1992.
[Article \(CrossRef Link\)](#)
- [2] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362-373, September, 2007.
[Article \(CrossRef Link\)](#)
- [3] Q. Abbas, M. E. Celebi and I. F. García, "Hair removal methods: a comparative study for dermoscopy images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 395-404, October, 2011. [Article \(CrossRef Link\)](#)
- [4] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234-241, October, 2015. [Article \(CrossRef Link\)](#)

- [5] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, December, 2017. [Article \(CrossRef Link\)](#)
- [6] S. Sookpotharom, "Border detection of skin lesion images based on fuzzy C-means thresholding," in *Proc. of Genetic and Evolutionary Computing, 2009. WGECC'09. 3rd International Conference on. IEEE*, pp. 777-780, October 14 - 17, 2009. [Article \(CrossRef Link\)](#)
- [7] H. Zhou, G. Schaefer, M. E. Celebi, H. Iyatomi, K. A. Norton, T. Liu and F. Lin, "Skin lesion segmentation using an improved snake model," in *Proc. of Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 1974-1977, November 11, 2010. [Article \(CrossRef Link\)](#)
- [8] C. M. Emre, H. A. Kingravi, H. Iyatomi, A. Y. Alp, W. V. Stoecker, R. H. Moss and S. W. Menzies, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347-353, 2008. [Article \(CrossRef Link\)](#)
- [9] G. Schaefer, M. I. Rajab, M. E. Celebi and H. Iyatomi, "Colour and contrast enhancement for improved skin lesion segmentation," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 99-104, March, 2011. [Article \(CrossRef Link\)](#)
- [10] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, June 7-12, 2015. [Article \(CrossRef Link\)](#)
- [11] S. Liu, X. Liang, L. Liu, X. Shen, J. Yang, C. Xu and S. Yan, "Matching-CNN meets KNN: Quasi-parametric human parsing," *IEEE Computer Vision and Pattern Recognition*, pp. 1419-1427, April, 2015. [Article \(CrossRef Link\)](#)
- [12] S. Liu, C. Wang, R. Qian, H. Yu, R. Bao and Y. Sun, "Surveillance Video Parsing with Single Frame Supervision," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, July, 2017. [Article \(CrossRef Link\)](#)
- [13] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong and S. Yan, "Deep Human Parsing with Active Template Regression," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 12, pp. 2402-2414, December, 2015. [Article \(CrossRef Link\)](#)
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, pp. 2672-2680, December 08 - 13, 2014.
- [15] W. Liu, X. Liu, H. Ma and P. Cheng, "Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Prior GAN," *ACM Multimedia*, October, 2017. [Article \(CrossRef Link\)](#)
- [16] Z. Li and J. Tang, "Weakly supervised deep matrix factorization for social image understanding," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 276-288, 2017. [Article \(CrossRef Link\)](#)
- [17] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967-5976, July, 2017. [Article \(CrossRef Link\)](#)
- [18] P. Luc, C. Couprie, S. Chintala and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint, arXiv:1611.08408*, 2016.
- [19] W. Liu, T. Mei, Y. Zhang, C. Che and J. Luo, "Multi-task deep visual-semantic embedding for video thumbnail selection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3707-3715, June, 2015. [Article \(CrossRef Link\)](#)
- [20] Z. Li, J. Liu, J. Tang and H. Lu, "Robust structured subspace learning for data representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 2085-2098, February, 2015. [Article \(CrossRef Link\)](#)
- [21] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. of IEEE International Conference on Computer Vision*, pp. 2813-2821, 2017. [Article \(CrossRef Link\)](#)

- [22] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536-2544, June 27-30, 2016. [Article \(CrossRef Link\)](#)
- [23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. of International Conference on Machine Learning*, pp. 448-456, 2015.
- [24] P. Vincent, H. Larochelle, Y. Bengio and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. of the 25th international conference on Machine learning. ACM*, pp. 1096-1103, July 05 - 09, 2008. [Article \(CrossRef Link\)](#)
- [25] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, January, 2014.
- [26] C. K. Sønderby, J. Caballero, L. Theis, W. Shi and F. Huszár, "Amortised map inference for image super-resolution," *arXiv preprint arXiv:1610.04490*, 2016.
- [27] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *Computer Science*, 2015.
- [28] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra and A. Halpern, "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)," *eprint arXiv:1605.01397*. 2016.
- [29] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. of the 23rd international conference on Machine learning*, pp. 233-240, June 25-29, 2006. [Article \(CrossRef Link\)](#)
- [30] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu and S. Yan, "Face Aging with Contextual Generative Adversarial Nets," in *Proc. of the 2017 ACM on Multimedia Conference*, pp. 82-90, October, 2017. [Article \(CrossRef Link\)](#)



Ning Wang is currently a master candidate in Science, Shandong University of science and technology, China. His main research interest is image processing.



Yanjun Peng received his Dr. degree in march, 2004. He joined the Department of Computer Science, Shandong University of science and technology, Qingdao, China, in 1996, where he was promoted to professor in 2010. His main research interests include medicine visualization, virtual reality, and image processing.



Yuanhong Wang received the BS degree from the computer science at the Shandong University of Science and Technology of China in 1993. Her main research interest is image processing.



Meiling Wang is currently a master candidate in Science, Shandong University of science and technology, China. Her main research interest is image processing.