ISBN 978-981-14-4787-7
Proceedings of 2020 the 10th International Workshop on Computer Science and Engineering (WCSE 2020)
Shanghai, 19-21 June, 2020, pp. 185-194
doi: 10.18178/wcse.2020.06.030

AM-AN: Adversarial Network Based on Attention Mechanism for Medical Image Segmentation

Lingjing Qin, Wenxin Hu [+] and Jun Zheng

School of Data Science & Engineering, East China Normal University, China

Abstract. Image segmentation is crucial to a series of medical applications. One of the current problems is that imbalanced pixel classes have a negative effect on the results of medical image segmentation. However, the method of employing weighted loss cannot address this problem well. Therefore, we are motivated to propose a novel method based on Generative Adversarial Networks (GANs) and attention mechanisms to train our segmentation model with effective loss functions. Firstly, the proposed model consists of the segmentor network and the critic network, which are trained by adversarial learning. Moreover, the attention module we used in each residual block takes not only channel attention but also spatial attention into consideration and makes the segmentor perform better. The overhead of computation and parameters costed by our attention module is negligible. In addition, our work extends other methods by means of using pixel-wise loss functions, which include multi-scale loss and binary cross entropy loss. Finally, our work presents the results of comparing three different methods on medical image segmentation and our method yields a higher performance than theirs on MICCAI PROMISE12, DSB2018 and the lung nodule datasets.

Keywords: Medical image segmentation, Attention mechanism, Adversarial network, Pixel-wise Loss.

1. Introduction

So far, semantic segmentation [1-3] has been commonly used in the field of medical imaging. This technology aims at figuring out the precise location and shape of structures in the human body. It makes medical disorders and their treatment being assessed properly. Methods based on convolutional neural networks (CNN) [4-8] currently perform well. However, there are some challenges in the field of medical image segmentation. Firstly, one of the challenges is concerned with the imbalanced classes of pixels. More specifically, although a balanced number of patches from each class can be sampled by patch training, the numbers of pixels, which correspond with their belonged classes, are imbalanced in methods of training whole images. Secondly, a new problem is that weighted loss function proposed for the first issue is task-specific and difficult to be optimized. As a result, their methods cannot resolve the problem effectively. Thirdly, another limitation is that they also cannot learn multi-scale spatial constraints directly.

To overcome these challenges, we propose a novel end-to-end network for medical image segmentation (AM-AN). Our model consists of two parts: the segmentor network and the critic network. To construct this model, we take the concept of adversarial learning as a reference. In addition, we employ the attention mechanism to improve performance of the segmentor. Our attention module is also made up of two parts which correspond to channel attention and spatial attention, respectively. Channel attention module is used to figure out inter-channel relationship of features while spatial attention module is used to decide where to focus on. We eventually place the attention module at the end of each residual block. This attention module can improve model's ability of extracting features. Besides, the overhead of calculation and the total number of parameters do not increase. In terms of loss function, we use the critic network to extract hierarchical features and then concatenate these features to compute a pixel-wise loss. The multi-scale loss has a good effect on optimizing

[+] Corresponding author. Tel.: + (021)62233797; fax: + (021)62576192.
E-mail address: wxhu@cc.ecnu.edu.cn.

our network. Furthermore, we also use a pixel-wise binary cross entropy loss function to help the segmentor perform better. In the experiment, we use MICAAI PROMISE12, Data Science Bowl 2018 (DSB2018) and lung nodule datasets to verify our method work better than other's. The experiment results demonstrate that our method increases the accuracy of medical image segmentation with effect.

Summarizing, the main contributions of our work are the following:

- We propose a novel end-to-end adversarial network, namely AM-AN, which is based on attention mechanism for medical image segmentation. We demonstrate that our network outperforms the networks constructed by other methods when segmenting medical images.
- We propose an effective attention module in each residual block of our model consisting of both spatial attention submodule and channel attention submodule. The experiment demonstrates the better results of using this attention module.
- We propose the pixel-wise loss functions capable of providing more precise segmented images, including multi-scale L1 loss with a different activation function and binary cross entropy loss.

The rest of this paper is organized as follows. In section 2, there is a brief overview of related work. In section 3, we display our network architecture including attention module and illustrate the loss functions. In section 4, there are results and analysis. Finally, we conclude with a summary of our main contributions and results.

2. Related Work

2.1. Medical Image Segmentation

Because of the high complexity possessed by medical image and lack of simple linear feature, medical image segmentation technology is always a complex and critical step in the field of medical image processing and analysis. In traditional segmentation methods, Kedir et al. [9] studied the impact of smoothing filters when applied to waveform image pre-processing for segmentation. Yao-Tien Chen [10] proposed a level set method based on the Bayesian risk for textured segmentation. Traditional methods are also used for medical image segmentation [11]. Javed et al. [12] proposed a system to enhance the quality of the human brain magnetic resonance image (MRI). In this system, median filter is used for image enhancement of brain MRI and fuzzy c-means for segmentation purpose. Pun et al. [13] proposed a skin colour segmentation approach by texture feature extraction and k-mean clustering. This approach improved the traditional skin classification by combining both colour and texture features for skin segmentation.

So far methods based on deep learning have been popular in this field and have achieved remarkable breakthroughs. For example, Korez et al. proposed a 3D fully convolutional networks (FCN) [14] which improves the segmentation accuracy of spinal MRI. After the FCN-based U-net [5] has been proposed, Brosch et al. [15] used U-net when segmenting white matter lesions in brain MRI. With less training data, this method still had a positive effect on the experiment results. Noh et al. [16] used an encoder-decoder structure to obtain more details of images. However, the imbalance of pixel classes is inevitable in their networks and cannot further increase the accuracy.

2.2. Generative Adversarial Network

Goodfellow et al. proposed Generative Adversarial Networks (GANs) [17] which trains generator G and discriminator D with a minmax game. Throughout this adversarial game, G is responsible for generating samples similar to real data while D is responsible for discriminating between samples and real data.

So far GANs have attracted plenty of attention within the computer vision field. Numerous works such as [18-21] have further extended and improved the original GAN [17]. Radford et al. [18] proposed a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints. Salimans et al. [19] presented a variety of new architectural features and training procedures that they apply to the GANs framework. Arjovsky et al. [20] introduced a new algorithm named Wasserstein GAN, an alternative to traditional GAN training. Gulrajani et al. [21] proposed an alternative to clipping weights: penalize the norm of gradient of the critic with respect to its input.

2.3. Attention Mechanism

There is no doubt that attention mechanism has been widely used because of its benefits proved by a series of tasks [22-24]. According to the concept of focusing on mask to form the attention mechanism, attention can be divided into hard attention and soft attention. From the perspective of attention domain, Hu et al. [25] propose a Squeeze-and-Excitation module that computes channel-wise attention belonging to soft attention. Wang et al. [26] propose Residual Attention Network which uses an encoder-decoder attention module. However, these methods just extract suboptimal features that aim to infer channel attention and are lacking in spatial attention.

3. Method

3.1. Network Architecture

The proposed network, namely AM-AN, is a novel end-to-end adversarial network based on attention mechanism for medical image segmentation. As illustrated in Fig. 1, components making up our AM-AN are the segmentor subnetwork and the critic subnetwork. Given an input image, a predicted mask is the output of the segmentor. We use attention mechanism in the segmentor to preserve the significant information of input images. Then, the predicted masked image can be achieved by pixel-wise multiplication of the predicted mask and the input image as shown. Similarly, the target masked image can be achieved by pixel-wise multiplication of the ground truth mask and the input image. The masked image is the input of the critic. We employ the output of the critic to calculate the proposed pixel-wise loss functions, which makes more detailed information propagated backward.

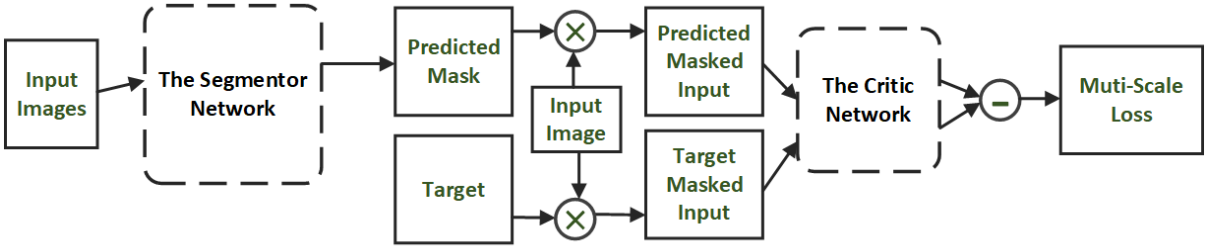

Fig. 1: The architecture of AM-AN.

a. Segmentor Subnetwork

Our method of constructing the segmentor can extract features of input images better. Furthermore, it also expands receptive field so that more detailed information can be achieved. More specifically, the framework of the segmentor is based on the U-net [5]. However, we employ global convolution instead normal convolution when the kernel size is larger than 7 as shown in Fig. 2. This choice has two advantages. From the aspect of classification, tight connections can be built between feature maps and pixel-wise classifiers. In addition, local information can be preserved. Besides, the stride of convolution layers in encoder is 2 while stride in decoder is 1. All downsampling blocks take leaky ReLU as their activation layer. In contrast, all upsampling blocks after executing batch normalization take ReLU as their activation layer.

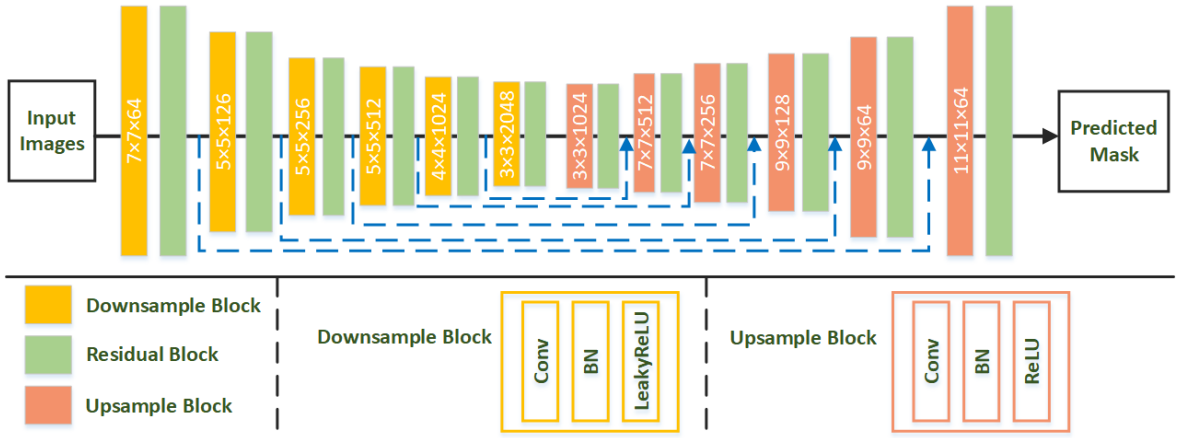

Fig. 2: The architecture of the segmentor.

b. Critic Subnetwork

Apart from the segmentor, the critic is also a crucial part of our module. The structure used to construct the critic network is easier to achieve the larger receptive fields and reduce the number of parameters. We employ fully convolutional layers in the critic. Similarly, all downsampling blocks in the critic take leaky ReLU as their activation layer. Note that there are no residual blocks in the critic as shown in Fig. 3.

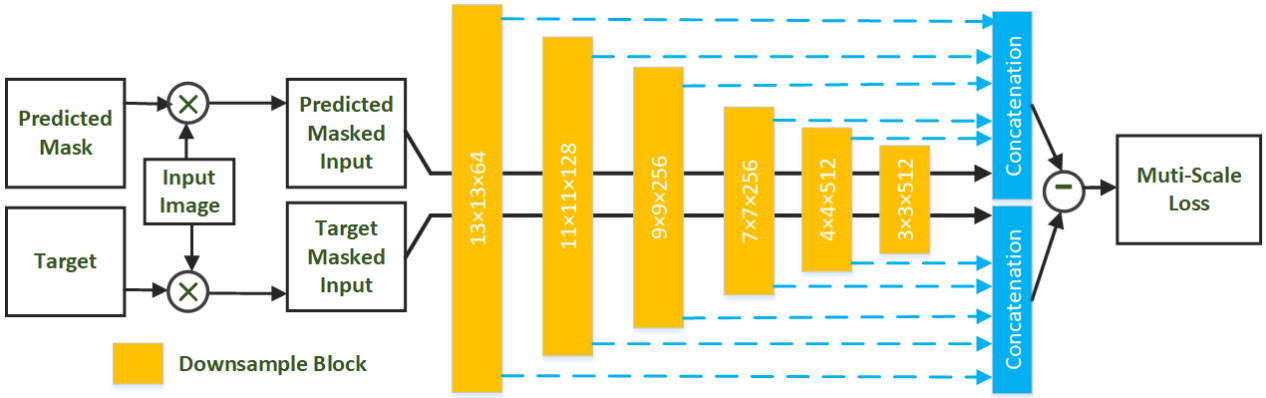

Fig. 3: The architecture of the critic.

3.2. Attention Module

The goal of using both channel attention and spatial attention in our attention module is to make the proposed adversarial network perform better in medical image segmentation. Valid information of the input medical images can be well preserved. Besides, it is designed to be light-weight and will not lead to the increase of the overhead of calculation and parameters. As shown in Fig. 4, attention module used in residual blocks of the segmentor is composed of two parts corresponding to channel attention and spatial attention, respectively. This attention module is placed at the end of each residual block in the segmentor. Given a feature map $\mathsf{F} \in \mathbb{R}^{C\times H\times W}$ as input, a channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$ and a spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$ will be achieved. Procedures of this process are implemented as follows:

$$\mathsf{F}' = M_c(\mathsf{F}) \otimes \mathsf{F} \tag{1}$$

$$\mathsf{F}'' = M_s(\mathsf{F}') \otimes \mathsf{F}' \tag{2}$$

where $\otimes$ refers to element-wise multiplication.

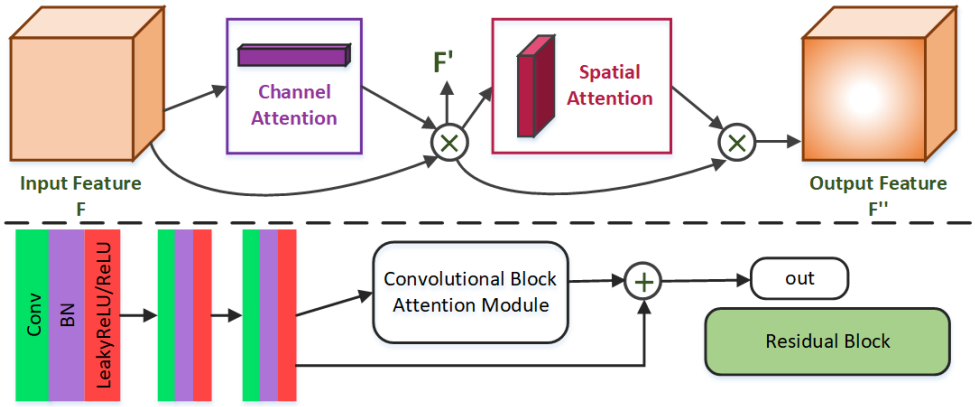

Fig. 4: Convolutional Block Attention Module.

During the process of producing an attention map, the emphasis is to figure out the significant information in an input image. As illustrated in Fig. 5, part of the channel attention module is similar to SE-Net proposed by Hu et al. [20]. However, we also take max pooling into consideration besides average pooling. Only the place with the greatest response in the feature map has gradient feedback during the calculation of gradient backward propagation. Therefore, the step of max pooling can be a complement to the step of average pooling. Then the two achieved 1D features are the input of shared network made up of multi-layer perceptron (MLP). Finally, it is essential to use sigmoid function. The process is presented as follows:

$$M_c(F) = \sigma(MLP(AvgPooling(F)) + MLP(MaxPooling(F))) \tag{3}$$

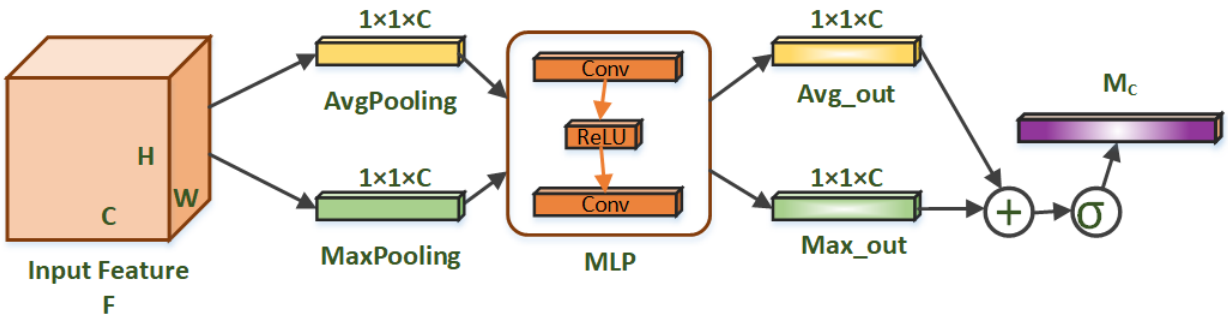

Fig. 5: Channel Attention Module.

It is attention on the spatial dimension that makes the proposed module different. Spatial attention makes it clear which parts of the feature map have the greater response from the respect of spatial dimension. Average pooling and max pooling are used the same as in channel attention module to compress the input feature map firstly. Whereas this compression is performed on the channel dimension instead of the spatial dimension as shown in Fig. 6. The two 2D achieved features are concatenated to get a feature map with two channels. Then the feature map is convolved by a hidden layer with a single convolution kernel. Besides, the filter size is 7×7. In the final step, sigmoid function is still needed. The process is presented as follows:

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPooling(F); MaxPooling(F)])\right) \tag{4}$$

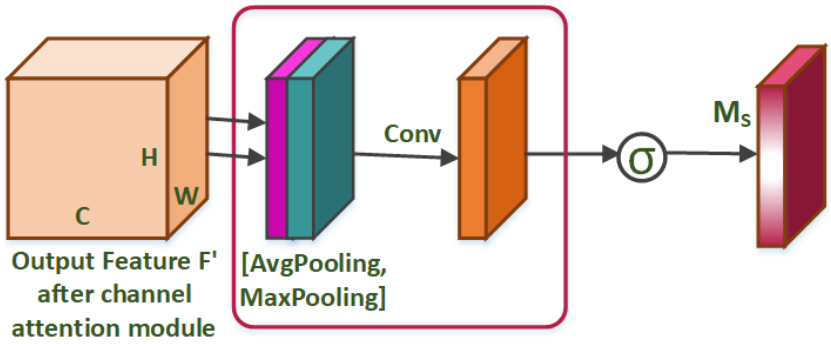

Fig. 6: Spatial Attention Module.

3.3. Loss Functions for Training

The traditional loss functions of generator and discriminator in GANs [12] are separated. They cannot back propagate more information of gradients and make the adversarial training process not stable enough. However, the loss function of our segmentor and critic network are trained by the same multi-scale loss function. The goal of our network is to figure out the map between an input image and its corresponding segmentation mask. Note that the segmentation mask here refers to pixel-wise label maps. Therefore, the significant step of implementing our multi-scale loss is to compute the Mean Absolute Error (MAE) value of two feature maps obtained by each convolution layer in the critic. The multi-scale loss makes the adversarial training process stable. It is defined as:

$$\min_{\theta_s}\max_{\theta_c} \mathcal{L}(\theta_s, \theta_c) = \frac{1}{N}\sum_{n=1}^{N} \ell_{mae}\left(f_c(x_n \circ S(x_n)), f_c(x_n \circ y_n)\right) \tag{5}$$

where x_n refers to N training images and y_n refers to N ground truth label maps. Besides, ℓ_{mae} refers to the MAE or L_1 distance. In addition, $x_n \circ S(x_n)$ represents the input image masked by the predicted result from the segmentor while $x_n \circ y_n$ represents the input image masked by the corresponding ground truth label map. More specifically, the masked image is obtained by pixel-wise multiplication of the input image and label map. $f_c(x)$ denotes that the critic network extracts hierarchical features from the image x. Furthermore, the ℓ_{mae} function is defined as:

$$\ell_{mae}(f_c(x), f_c(x')) = \frac{1}{L}\sum_{i=1}^{L}\left\|f_c^i(x) - f_c^i(x')\right\|_1 \tag{6}$$

where $f_c^i(x)$ denotes the ith layer of the critic extracts feature maps from the image x. Besides, L represents the total number of layers of the critic network.

Since it is not easy to distinguish between the predicted result and the corresponding ground truth label map, we take an adaptive logistic activation function as the last layer of the segmentor instead of sigmoid function. The new activation function is defined as:

$$f(z) = \frac{1}{1+e^{-z/k}} \tag{7}$$

where k is used to control the steepness of the curve. This function will become into sigmoid function when the value of k is 1. Note that the curve of this function will be steeper with the decrease of k's value.

Apart from the multi-scale L_1 loss function for training the whole network, we still use binary cross entropy (BCE) loss function to train the segmentor network. The predicted result generated by the segmentor can be more similar to ground truth label map because of the different use of BCE loss. More specifically, BCE loss function is employed in the form of pixel-wise loss. In other words, there are as many binary decisions as pixels in the image. Such choice of loss function can provide more details so that gradients with more information can be back propagated to optimize our segmentor.

4. Experiment

4.1. Datasets

In this work, we use MICAAI PROMISE12, Data Science Bowl 2018 (DSB2018) and lung nodule datasets for verifying the performance of our model and compare it with other models'.

The MICAAI PROMISE12 dataset includes 50 MRI images of prostates. 261 sets of data which are slices of prostate region can be achieved. The DSB2018 dataset includes 670 sets of training data. One original image together with multiple labels compose one set of data. Each label corresponds to a single nucleus. Both are public medical image datasets. In addition, there are 19 patient data in the lung nodule dataset.

We divide each dataset into three parts with a ratio of 7:2:1, which corresponds to training data, validation data and testing data, respectively.

4.2. Evaluation Metrics

One of our evaluation metrics is the *Dice score* which can normalize the number of true positives to the average size of the two segmented regions. It is computed as:

$$\text{Dice} = \frac{2|A_1 \cap A_2|}{|A_1|+|A_2|} \tag{8}$$

where A_1 and A_2 refer to the ground truth label region and the corresponding predicted region respectively. Dice values range between 0 and 1, with 1 corresponding to a perfect overlap. The larger the Dice score is, the better the method performs.

The other evaluation metric is the mean Intersection-over-Union (mIoU), which is also known as Jaccard index. It is calculated as:

$$\text{mIoU} = \frac{1}{n}\sum_n \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} \tag{9}$$

where n represents the number of input images. Similarly, the larger the mIoU score is, the better the method performs.

4.3. Implementation Details

All the experiments have been implemented in a single NVIDIA Tesla P100 GPU, and the CUDA edition is 9.0. In addition, the implement ed framework is based on Pytorch 1.1.0 and torchvision 0.3.0 The initial learning rate is set to 0.0002. Besides, batch size of training and testing is set to 36. All the experiments are trained in 200 epochs.

We center-cropped images of datasets to size 128×128 because of the limitation of hardware memory. the input image uses bilinear interpolation while the ground truth label uses nearest interpolation. In addition, we normalize the resized images by the value of mean and standard deviation. Then we randomly flip images and labels vertically and horizontally. This step is only operated for training, not including validation and testing.

In each training epoch, we fix the segmentor S and train the critic C firstly, then fix C and train S. Each step uses gradients obtained by the loss function mentioned above. Robustness of the whole network can be

enhanced during the training process. As the process moves on, the results predicted by the segmentor become more and more similar to the corresponding ground truth label map.

4.4. Results and Analysis

4.4.1. Comparison with Baselines

Each comparison based on different datasets includes three models:

- U-Net [5] is an encoder-decoder network based on FCN. It concatenates features on the channel dimension.
- U-Net++ [27] is a deeply-supervised encoder-decoder network where the subnetworks are connected through a series of nested, dense skip pathways.
- SegAN [8] is constructed with the structure resembling U-Net. It also uses the concept of adversarial learning to segment medical images.
- Our AM-AN is an end-to-end adversarial network based on attention mechanism and pixel-wise loss functions. Our attention module includes both channel attention and spatial attention.

In Table 1, we compare our model to other models on two public datasets, PROMISE12 and DSB2018. The third column represents the Dice-score of different methods. The forth column represents the mIoU-score of different methods. The Dice-score and the mIoU-score of our model is larger than those of other three models. Compared to the result of SegAN on PROMISE12 dataset, Dice-score of our model has an increase of 1.9% while mIoU-score has an increase of 1.8%. Compared to the result of SegAN on DSB2018 dataset, Dice-score of our model has an increase of 1.2% while mIoU-score has an increase of 0.9%. Our network has a better performance than networks using other methods. The metrics of our module are larger than those of other modules, representing the predicted segmented result is more similar to the ground truth label. Besides, significant information of the prostate image and the cell nuclear region has been preserved well because of the proposed attention module, and pixels have been precisely classified due to the pixel-wise loss functions.

In Table 2, the comparison is based on the lung nodule dataset. The form of Table 2 is the same as Table 1. In addition, the comparison is also made on the four models. In Table 2, the Dice-score and mIoU-score of our network are higher than those of other models as well. Compared to the result of SegAN, Dice-score of our model has an increase of 3.0% while mIoU-score also has an increase of 3.0%. Therefore, our network also performs better on the lung nodule dataset. Besides, Dice-score and mIoU-score of results on the lung nodule dataset are lower than those results on PROMISE12 and DSB2018. Because number of pixels belonging to the background are much larger than that belonging to the segmented region in the lung nodule dataset. It makes the segmentation task of lung nodule more difficult.

Table 1: Comparison between different models on two public datasets

Dataset	Model	Dice	mIoU
PROMISE12	U-Net	0.865	0.864
	U-Net++	0.884	0.887
	SegAN	0.882	0.879
	Our model	**0.899**	**0.895**
DSB2018	U-Net	0.909	0.904
	U-Net++	0.916	0.912
	SegAN	0.913	0.913
	Our model	**0.924**	**0.921**

Table 2: Comparison between different models on the lung nodule dataset

Dataset	Model	Dice	mIoU
Lung nodule	U-Net	0.713	0.664
	U-Net++	0.729	0.698
	SegAN	0.734	0.701
	Our model	**0.756**	**0.722**

4.4.2. Ablation Study

In ablation study, we present a further experiment that support our proposed network structure and pixel-wise training loss functions. This experiment is based on the three datasets as well, including two public datasets and the lung nodule dataset. The results of comparison based on two public datasets are displayed in Table 3. The comparison contains three cases. The first one is that only the proposed attention module is added to the adversarial network. The second case is that only the pixel-wise loss functions, which include multi-scale loss and binary cross entropy loss, are employed to the adversarial network.

The table demonstrates that not only the attention module but also the loss functions have an increase of Dice-score and mIoU-score. Compared to the baseline on PROMISE12 dataset, Dice-score of the module with proposed attention module has an increase of 1.2% while mIoU-score has an increase of 1.1%. Compared to the baseline on PROMISE12 dataset, Dice-score of the module with pixel-wise loss functions has an increase of 0.9% while mIoU-score has an increase of 1.3%. Compared to the baseline on DSB2018 dataset, Dice-score of the module with proposed attention module has an increase of 0.7% while mIoU-score has an increase of 0.4%. Compared to the baseline on DSB2018 dataset, Dice-score of the module with pixel-wise loss functions has an increase of 0.4% while mIoU-score has an increase of 0.4%.

The increase brought by attention module is usually larger than that brought by the proposed loss functions. According to the results, the proposed attention module is crucial for the segmentor to segment images more precisely. Besides, the proposed loss functions help the network extract hierarchical features and make the classification of each pixel more precise. A conclusion can be drawn that both attention module and loss functions have effectively improved the performance of our model.

In addition, the results of comparison based on the lung nodule dataset are displayed in Table 4. Compared to the baseline, Dice-score of the module with proposed attention module has an increase of 1.9% while mIoU-score has an increase of 1.5%. Compared to the baseline, Dice-score of the module with pixel-wise loss functions has an increase of 1.7% while mIoU-score has an increase of 2.3%. The table also proves that both attention module and pixel-wise loss have a good effect on the task of segmenting medical images.

Table 3: Ablation study on two public datasets

Dataset	Method	Dice	mIoU
PROMISE12	SegAN	0.882	0.879
	+AM	0.893	0.889
	+ pixel-wise loss	0.890	0.891
	+AM+ pixel-wise loss (our model)	**0.899**	**0.895**
DSB2018	SegAN	0.913	0.913
	+AM	0.919	0.917
	+ pixel-wise loss	0.917	0.916
	+AM+ pixel-wise loss (our model)	**0.924**	**0.921**

Table 4: Ablation study on the lung nodule dataset

Dataset	Method	Dice	mIoU
Lung nodule	SegAN	0.734	0.701
	+AM	0.748	0.713
	+ pixel-wise loss	0.745	0.717
	+AM+ pixel-wise loss (our model)	**0.756**	**0.722**

Besides, we verify whether the proposed attention module is light-weight. As shown in Table 5, the number of parameters in our model is almost the same as that of other methods. Apart from the parameters, GFLOPs of our model also has a negligible increase. These metrics are barely changed when our attention module is added. Therefore, a conclusion can be drawn that the proposed attention module is quite light weight and do not increase the overhead of computation and parameters in our network.

Table 5: Comparison of the overhead of different methods

Method	Parameters	GFLOPs
SegAN	156.27M	18.90
+channel	156.42M	18.91
+channel +spatial (our model)	156.42M	18.91

5. Conclusion

We present a novel method for segmenting medical images, which performs better comparable to other methods. We demonstrate how this method employs the adversarial learning and attention mechanisms to generate predicted results that resemble the corresponding ground truth labels. The employed attention module can enhance the ability of extracting features of input images. More than that, it does not increase the cost of computation and the number of network parameters because it is light-weight. Moreover, results on datasets prove that our pixel-wise loss functions have a positive effect on segmenting.

The experiment in our work shows that our method possesses a higher accuracy than others on three different medical datasets. In our work, the limitation is that the quantity of data in datasets are not large enough. In the future, our method can be extended to other 3D multi-modal segmentation tasks with data augmentation using learned transforms.

6. Acknowledgements

This work is funded by Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China. The experiment is completed with the support of ECNU Multifunctional Platform for Innovation (001).

7. References

[1] A. Zhao, G. Balakrishnan, Frédo Durand, John V. Guttag, Adrian V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. *CVPR 2019*. 2019.

[2] C. Chen, Q. Dou, H. Chen, J. Qin, P. Heng. Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation. *AAAI 2019*. 2019.

[3] X. He, S. Yang, G. Li, H. Li, H. Chang, Y. Yu. Non-Local Context Encoder: Robust Biomedical Image Segmentation against Adversarial Attacks. *AAAI 2019*. 2019.

[4] J. Long, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431-3440.

[5] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp.

234–241.

[6] V. Badrinarayanan, A. Kendall, R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, 2015.

[7] S. Pereira, A. Pinto, V. Alves, C. Silva. A. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging.* 2016, 35(5): 1240–1251.

[8] Y. Xue, T. Xu, H. Zhang, L. Rodney Long, X. Huang. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. *Neuroinform.* 2018, 16: 383–392.

[9] Kedir Kamu Sirur, Ye Peng, and Zhang Qinchuan. Smoothing Filters for Waveform Image Segmentation. *International Journal of Machine Learning and Computing*. 2017, vol. 7, no.5, pp. 139-143.

[10] Yao-Tien Chen. A Level Set Method Based on Bayesian Risk for Textured Image Segmentation. *International Journal of Machine Learning and Computing*. 2017, vol. 7, no.4, pp. 89-93.

[11] Sreedhar Kollem, Katta Rama Linga Reddy, and Duggirala Srinivasa Rao. A Review of Image Denoising and Segmentation Methods Based on Medical Images. *International Journal of Machine Learning and Computing*. 2019, vol. 9, no. 3, pp. 288-295.

[12] Arshad Javed, Wang Yin Chai, Abdulhameed Rakan Alenezi, and Narayan Kulathuramaiyer. Enhancement of Magnetic Resonance Image Using Soft Computing Based Segmentation. *International Journal of Machine Learning and Computing*. 2014, vol. 4, no. 1, pp. 73-78.

[13] Chi-Man Pun and Pan Ng. Skin Segmentation Using GMM Classifier and Texture Feature Extraction. *International Journal of Machine Learning and Computing*. 2014, vol. 4, no. 1, pp. 57-62.

[14] R. Korez, B. Likar, F. Pernus, T. Vrtovec: Model-Based Segmentation of Vertebral Bodies from MR Images with 3D CNNs. *MICCAI* (2). 2016, pp. 433-441.

[15] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee and R. Tam. Deep 3D Convolutional Encoder Networks with Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. *IEEE Transactions on Medical Imaging*. 2016, pp. 1229-1239.

[16] H. Noh, S. Hong, B. Han. Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1520–1528.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[18] A. Radford, L. Metz, S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *ICLR 2016*. 2016.

[19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen. Improved techniques for training gans. *Advances in Neural Information Processing Systems*. 2016, pp. 2234–2242.

[20] M. Arjovsky, S. Chintala, L. Bottou. Wasserstein gan. arXiv preprint arXiv:1701.07875, 2017.

[21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C Courville. Improved training of wasserstein gans. *Advances in Neural Information Processing Systems*. 2017, pp. 5767–5777.

[22] L. Itti, C. Koch, E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* TPAMI, 1998.

[23] R.A. Rensink. The dynamic representation of scenes. *Visual cognition 7.1-3.* 2000.

[24] M. Corbetta, G.L. Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience 3.3.* 2002.

[25] J. Hu, L. Shen, G. Sun. Squeeze-and-excitation networks. *CVPR 2018*. TPAMI, 2018.

[26] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang. Residual attention network for image classification. *CVPR 2017*. 2017.

[27] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2018, pp. 3-11.