TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 06/18 pp68–78
DOI: 10.26599/TST.2020.9010056
Volume 27, Number 1, February 2022

Combining Residual Attention Mechanisms and Generative Adversarial Networks for Hippocampus Segmentation

Hongxia Deng*, Yuefang Zhang, Ran Li, Chunxiang Hu, Zijian Feng, and Haifang Li

Abstract: This research discussed a deep learning method based on an improved generative adversarial network to segment the hippocampus. Different convolutional configurations were proposed to capture information obtained by a segmentation network. In addition, a generative adversarial network based on Pixel2Pixel was proposed. The generator was a codec structure combining a residual network and an attention mechanism to capture detailed information. The discriminator used a convolutional neural network to discriminate the segmentation results of the generated model and that of the expert. Through the continuously transmitted losses of the generator and discriminator, the generator reached the optimal state of hippocampus segmentation. T1-weighted magnetic resonance imaging scans and related hippocampus labels of 130 healthy subjects from the Alzheimer's disease Neuroimaging Initiative dataset were used as training and test data; similarity coefficient, sensitivity, and positive predictive value were used as evaluation indicators. Results showed that the network model could achieve an efficient automatic segmentation of the hippocampus and thus has practical relevance for the correct diagnosis of diseases, such as Alzheimer's disease.

Key words: magnetic resonance imaging; generative adversarial network; residual network; attention mechanism

1 Introduction

The hippocampus is located between the thalamus and the medial temporal lobe of the brain, and it is a part of the limbic system. It is mainly responsible for the storage, conversion, and orientation of long-term memory. The hippocampus is closely related to many neurological diseases, such as Alzheimer's disease, schizophrenia, and dementia[1]. However, the shape of the hippocampus is irregular, its volume is small, its edges have no clear boundaries, and individual differences are large. At present, manual segmentation results are still considered to be the gold standard for hippocampal morphological analysis. The study of the volume and shape of the hippocampus is a necessary condition for the diagnosis of these diseases. Therefore, automatic and accurate segmentation of the hippocampus using magnetic resonance imaging (MRI) images and its analysis and research are of great practical relevance for the correct diagnosis of these diseases.

At present, the following methods are used for brain MRI hippocampus segmentation: manual, semiautomatic, and automatic segmentation methods. The manual segmentation method is performed by experts to mark and segment the contours of the hippocampus on each slice. Although the result of manual segmentation is still considered to be the gold standard, the process is time consuming and subjective. Semiautomatic segmentation uses the threshold method and the image boundary tracking method to introduce initial image contour information and achieve hippocampus segmentation. This process requires precise control of the prior parameters,

• Hongxia Deng, Yuefang Zhang, Ran Li, Chunxiang Hu, Zijian Feng, and Haifang Li are from the School of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China. E-mail:denghongxia@tyut.edu.cn; 792331696@qq.com; 1319715230@qq.com; 1623481484@qq.com; 314503969@qq.com; lihaifang@tyut.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2020-10-20; accepted: 2020-11-17

and the parameter adjustment process is too time consuming. Automatic segmentation methods are divided into traditional automatic and deep learning-based segmentation methods. Traditional segmentation methods include graph-[2–4] and deformation-based[5, 6] methods. These methods often rely excessively on auxiliary technologies, such as classifiers and optimizers, which are difficult to rely on when using simple registration methods. An accurate segmentation of different hippocampi with large differences is difficult to achieve through simple registration methods.

In recent years, deep learning methods have been widely used in computer vision[7], and convolutional neural networks have made some progress in medical image processing. Alaoui et al.[8] proposed a method based on machine learning to classify tumors. Shelhamer et al.[9] used the fully convolutional neural network (FCNN) for the first time to perform pixel-level segmentation of natural images. The fully connected layers are converted into convolution operations, and the global information and the local information are taken into account by combination of shallow and deep features. It also supports arbitrary size image training and segmentation, and has made a breakthrough in semantic segmentation. Subsequently, the UNet model[10] was proposed for wide use in medical segmentation. This model can use few images for end-to-end training. At each stage, the decoder can learn the relevant characteristics lost during encoder pooling. Chen et al.[11] proposed U-Seg-Net to segment 2D slices of MRI images and then reconstructed the 3D hippocampus segmentation results through 2D slice image sequence reconstruction. Chen et al.[12] proposed a method of combining FCNNs and recurrent neural networks to achieve automatic hippocampus segmentation. Cao et al.[13] used an improved 3D-UNet[14] network to perform three-dimensional segmentation of nuclear magnetic images; however, this method requires a large training dataset, and training three-dimensional networks requires huge resources. The above models perform excellently in medical image segmentation. However, when performing segmentation tasks, the relationship between pixels can be easily ignored, resulting in the problem of semantic gap.

Numerous methods based on deep learning can be used at present; however, the segmentation of details is crucial for medical images. Moreover, ordinary network models tend to ignore the correlation between image pixels. To solve this problem, Goodfellow et al.[15] first proposed the generative adversarial net work (GAN) model through a cyclic iteration between the generative and adversarial models; thus, the generative model can generate fake and real pictures. However, GAN is prone to problems, such as gradient explosion and structural instability during the training process. To solve the problem in which the model is difficult to train, improvements were made on the basis of the model[16–21]. Luc et al.[22] applied GAN to natural image segmentation for the first time and obtained high segmentation accuracy by training VGG-based generative adversarial networks. The results obtained through adding the adversarial model are more spatially consistent than those obtained through only the generative model. Xue et al.[23] proposed a new end-to-end adversarial network architecture called SegAN. The generative model uses an FCNN to segment samples and proposes a new discriminant network with multiscale L1 loss to force the generative and adversarial models to learn simultaneously the global and local features that capture the long- and short-distance spatial relationships between pixels. To modify the standard GAN architecture, DCGAN[24] was proposed; it uses convolution generator and discriminator, uses batch normalization, replaces all pooling layers with convolution, and forces the generator to create a piecework mask in addition to image generation. To overcome the instability of the GAN training process, the slow convergence rate, and the possibility of mode collapse, Mondal et al.[25] proposed a semisupervized segmentation method using GAN for 3D multimodal images with few training samples available.

Recently, deep learning methods have been introduced into medical image segmentation[26]. Murugesan et al.[27] proposed a novel context-based CE loss function for UNet and a novel architecture Seg-GLGAN to address the problem of the foreground-background class imbalance in medical images. Izadi et al.[28] proposed a novel approach for skin lesion segmentation by leveraging generative adversarial networks. Li et al.[29] proposed an adversarial training approach to train the CNN network for brain tumor segmentation. The combination of the generative and adversarial models enhances the spatial continuity of the segmentation results and improves the accuracy of segmentation. Shi et al.[30] proposed a method based on GAN to achieve the hippocampus segmentation. This method uses the UNet network as a generative model and performs interactive training with the adversarial model to achieve

pixel-level classification of brain MRI images. Hui et al.[31] proposed a super resolution generative adversarial network (SRGAN) migration learning model on the Lung Nodule Analysis 2016 Dataset. The emergence of GAN makes the segmentation results of medical images smoother.

This study proposes a generative adversarial network method (Res_SEblock_GAN) that combines residual blocks and attention mechanisms to achieve end-to-end automatic hippocampus segmentation. The main contributions of this study are the following three points: First, to obtain semantic information, ResNet34 is used as the coding structure of the generative model to extract the shallow features of the image. Second, the attention mechanism SEblock module is added to the decoding structure of the generative model. Third, the proposed method is used to segment hippocampus images on the Alzheimer's disease neuroimaging initiative (ADNI) dataset and compare it with recent segmentation models. Experimental results show that the proposed method can well complete the task of automatic and accurate hippocampus segmentation.

2 Method

In this section, the proposed model for hippocampus segmentation is described. To improve the accuracy of hippocampus segmentation, this study proposes a generative adversarial network model based on Pixel2Pixel[32] as the basic architecture. The proposed model (Res-SEblock-GAN) is mainly composed of the generative and adversarial models. The generative model is a codec structure combining a residual network[33] and an attention mechanism. The adversarial network uses a convolutional neural network to discriminate the segmentation results generated by the generative model and expert segmentation results. Through the generative and adversarial models, the loss is continuously transmitted; thus, the generative model reaches the optimal state of hippocampus segmentation. When GAN is applied in image segmentation, the generative model is a segmentation network; the adversarial network judges the trueness of the segmentation of the generative network and then transmits the loss back to the generative model; then, the generative model is segmented again. By iterating between the generative model and the adversarial model, the segmentation accuracy of the segmentation network is continuously improved. Figure 1 shows the proposed generative adversarial network model. When the discriminator cannot distinguish between the generator's segmentation result and the expert segmentation result, at this time, the network has reached the optimal state.

2.1 Generative model

The generative model is designed following the principle of semantic segmentation, as shown in Fig. 2a. This design is achieved by combining the codec structure of the residual network and the attention mechanism. The

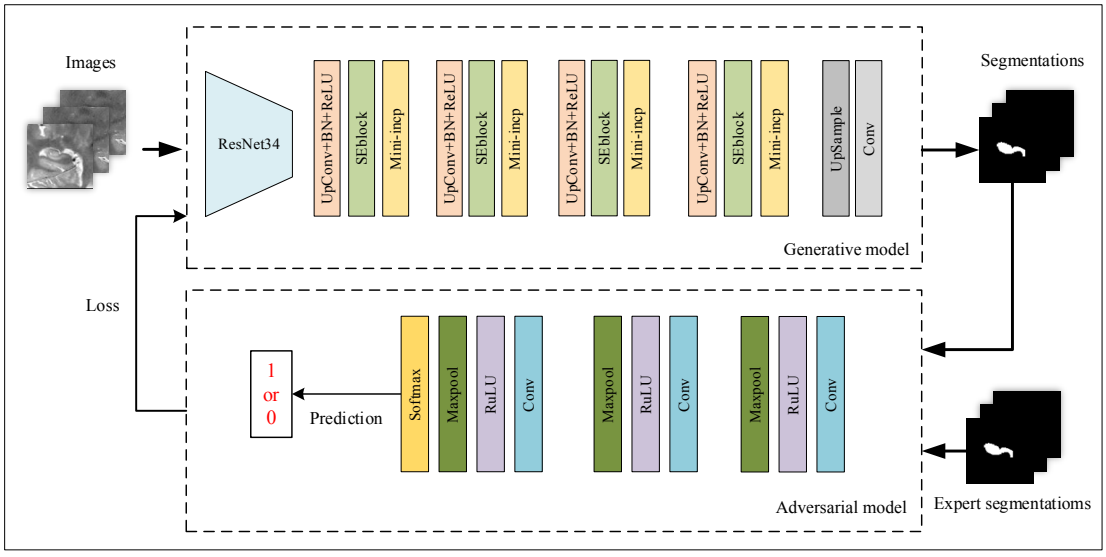


Fig. 1 Proposed generative adversarial network model: First, the images are input into the generative model. Then, the segmented hippocampus label map is output. Afterward, the result of the segmentation of the generative model and the result of the expert's segmentation of the hippocampus are input into the adversarial model. Finally, a predicted value is obtained through the SoftMax function.

encoding part uses the ResNet34 residual network, and the decoding part is integrated into the SEblock structure and Inception layer. SENet[34] can strengthen the characteristics of important channels and weaken the characteristics of nonimportant channels. It also models the relationship in the feature map channel in an efficient computing manner and is designed to enhance the network module's expressive ability in the network. The network model can improve its learning ability by deepening the number of network layers to extract deep features. However, this method also has certain deficiencies. First, an increase in deep network parameters likely results in overfitting. Transmitting gradient updates to the entire network can be difficult, thus causing gradient dispersion. Second, simply stacking large convolutional layers consumes computer resources. To make the network lightweight, an Inception[35] module is proposed. By embedding a multiscale feature extraction method in the network model, the network model can obtain considerable accuracy. A convolutional layer is used in front of the original SEblock structure to adjust the channel spliced by the previous jump connection. The SEblock structure strengthens the characteristics of important channels and weakens the characteristics of nonimportant channels. Through learning, the importance of each feature channel is automatically obtained; then, in accordance with this importance, the useful features are enhanced, and the features that are not useful for the current task are suppressed[34]. Multiscale features are extracted through the Inception structure to improve segmentation accuracy.

Figure 2b shows the SEblock structure. The core idea of SEblock is to learn feature weights in accordance with loss through the network so that the effective feature map weight is large, and the invalid or small effect feature map weight is small.

First, we perform the global average pooling (GAP) operation on the feature map extracted by the convolutional layer. Then we convert the input of $H \times W \times C$ into $1 \times 1 \times C$ output, and extract global information. H is the height, and W is the width. The calculation process is shown in Eq. (1).

$$Z_C = F_{sq}(U_C) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i, j) \quad (1)$$

where function F_{sq} represents squeeze operation. The feature U extracted from each convolutional layer is denoted as $U = U_1, U_1, U_1, \ldots, U_c$, where U_i represents the i-th characteristic feature map, and c represents the total number of feature maps. Next, as in Eq. (2), the excitation operation is performed. The

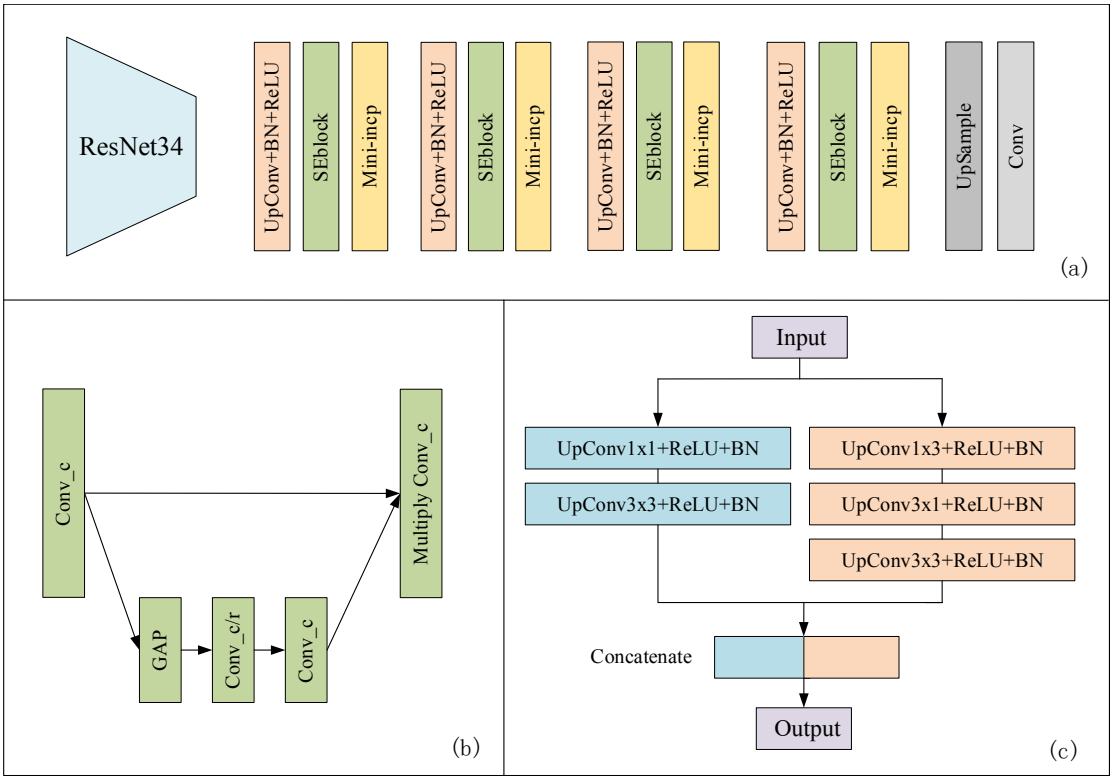


Fig. 2 Proposed generative model. (a) The generated model structure, (b) the SEblock structure, and (c) the Inception structure.

result obtained by GAP is z. Here, multiplying W_1 by z is a fully connected layer operation. The dimension of W_1 is $C/r \times C$, where r represents a scaling parameter and is set to 16 in the study; the purpose is to reduce the number of channels and thus reduce the amount of calculation. The dimension of z is $1 \times 1 \times C$, and the dimension of $W_1 z$ is $1 \times 1 \times C/r$. After the ReLU layer, the output dimension remains unchanged; then it is multiplied by W_2, Multiplying with W_2 is also a process of fully connected layer. The dimension of W_2 is $C \times C/r$; thus, the output dimension is $1 \times 1 \times C$.

$$S = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \tag{2}$$

where function F_{ex} represents excitation operation. δ represents the ReLU activation function and σ represents the sigmoid activation function, and s is obtained through the sigmoid function..

After obtaining s, a dot product operation is performed to complete the recalibration of the original input feature channel. The calculation process is performed as in Eq. (3), where function F_{scale} represents re-weighting operation. S_c represents weight of the output of the excitation operation, and it is regarded as the importance of each feature channel after feature selection. U_c is a two-dimensional matrix, and S_c is the weight; thus, ecalibration is equivalent to multiplying each value in the U_c matrix by S_c. This process is the Multiply operation in Fig. 2b.

$$\tilde{X}_c = F_{scale}(U_c, S_c) = S_c \cdot U_c \tag{3}$$

Figure 2c shows the Inception structure. With reference to the structure of InceptionV1, two different convolutional layers are used to extract multiscale information to capture additional features. To reduce the amount of calculation, an asymmetrical convolution kernel is used to solve the $N \times N$ two-dimensional convolution into two one-dimensional convolutions of $1 \times N$ and $N \times 1$, that is, the decomposition of 3×3 into 1×3 and 3×1.

2.2 Adversarial model

The adversarial model takes the hippocampus image and the corresponding label image as the input. The corresponding label diagram is the result of expert segmentation and the segmentation result of the generative model. They are input into the model through a series of convolution layers and maximum pooling layers, and a probability is generated to judge whether the label map is close to the expert segmentation standard. The closer it is to 1, the more realistic the segmentation result of the generative model is, and the closer it is to the expert segmentation result. The adversarial model then transfers the loss to the generative model; after many iterations of confrontation, when the discriminator cannot distinguish between the generator's segmentation result and the expert segmentation result, the network reaches the state of optimal segmentation effect.

3 Implementation Detail

3.1 Data

The experimental dataset comes from the ADNI database. A total of 130 sets of baseline T1-weighted whole brain MRIs from different subjects and their corresponding label images were downloaded, and data were all from the normal control group. Figure 3 is a visualization of the fusion of data of whole brain MRI images and corresponding hippocampal tags.

Considering that the hippocampus occupies a small volume in the entire brain and its position in the entire brain is relatively fixed, we roughly cropped the data and used it as input.

Figure 4 shows the result of preprocessing. Among them, Figs. 4b and 4d represent the left and right hippocampus, respectively, and the size is 64×64; this size can accommodate the hippocampus of each subject.

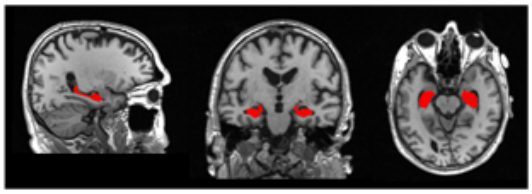


Fig. 3 Visualization of the fusion of data. The left picture is the sagittal plane, the middle picture is the coronal plane, and the right picture is the cross section.

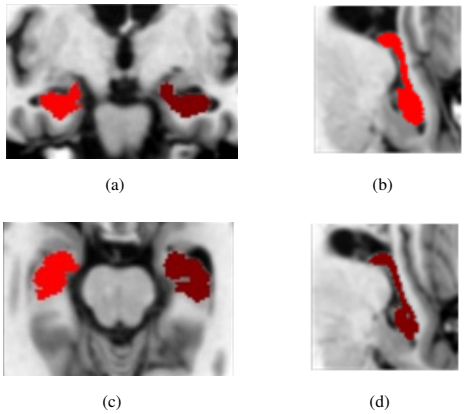


Fig. 4 Pre-processing results: (a) the coronal plane, (b) the left hippocampus of the sagittal plane, (c) the cross-sectional hippocampus, and (d) the right hippocampus of sagittal plane.

This study used data Figs. 4b and 4d for training.

3.2 Loss function

The generative model uses the mean absolute value error (MAE) as the loss function. The average absolute value error, also known as L1 loss, is the sum of the absolute value of the difference between the target value and the predicted value. The MAE calculation is shown in Eq. (4).

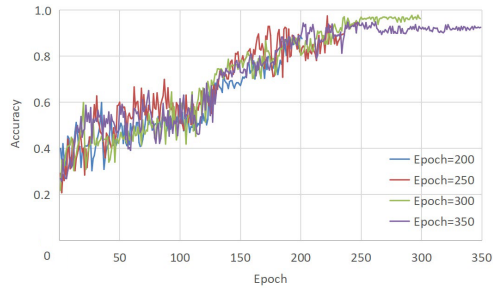
$$MAE = \sum_{i=1}^{n} |y_i - y_i^p| \tag{4}$$

The adversarial model uses the mean square error (MSE) as the loss function. The MSE is a commonly used regression loss function, which represents the sum of squares of the distance between the predicted value and the true value. The MSE calculation is shown in Eq. (5).

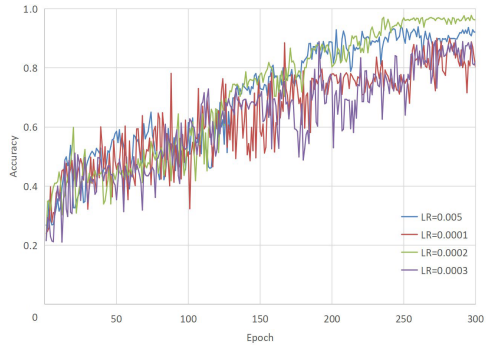
$$MSE = \sum_{i=1}^{n} (y_i - y_i^p)^2 \tag{5}$$

3.3 Parameter settings

The software environment is Keras 2.2.4, which uses the Nadam optimizer; beta 1 = 0.5, and the batch size is 128. Figure 5 shows the selection of epoch and learning rate (LR). Figure 5a shows that the epoch does not reach convergence when set to 200 and 250. The accuracy rate is the highest when set to 300. When set to 350, the accuracy rate is not improved but decreased. The loss of the computer is large, and the effect is not ideal; thus, the epoch is set to 300. Next is the selection of learning rate. Under the premise of controlling the epoch to 300, Fig. 5b reveals that the learning rate reaches the highest accuracy at 0.0002; thus, the learning rate is set to 0.0002. To verify the performance of the algorithm proposed in this study, a ten-fold cross-validation method is used. The experimental hardware environment is a NVIDIA GTX1080Ti single GPU and an Intel Core i7 processor.



(a) Selection of parameter Epoch



(b) Selection of parameter LR

Fig. 5 Parameter setting.

3.4 Evaluation index

To evaluate quantitatively the performance of the model proposed in this study, we chose the dice similarity coefficient (DSC), sensitivity (SEN), positive predictive value (PPV), and the Jaccard coefficient as the evaluation indicators of the hippocampus segmentation results. These evaluation indexes are widely used in the field of medical image segmentation.

$$DSC(A, B) = 2 \times \frac{|A \cap B|}{|A| + |B|} \tag{6}$$

$$SEN(A, B) = \frac{|A \cap B|}{|B|} \tag{7}$$

$$PPV(A, B) = 2 \times \frac{|A \cap B|}{|A|} \tag{8}$$

$$Jaccard(A, B) = 2 \times \frac{|A \cap B|}{A \cup B} \tag{9}$$

Equations (6)–(9) are calculation methods, where A represents the result of the hippocampus segmented by the network model in this study, B represents the result of the expert manually segmenting the hippocampus, $A \cap B$ represents the voxel region at the intersection of the algorithm's segmentation region and the expert's manual segmentation region, and $A \cup B$ represents all the voxel regions of the segmented area of the algorithm in this study and the expert-segmented area. DSC represents the degree of overlap between the experimental segmentation results and the real hippocampus, SEN represents the proportion of correctly segmented hippocampus to the real hippocampus, PPV represents the proportion of correctly segmented hippocampus to the segmented hippocampus, and the Jaccard coefficient is the ratio of sample set intersection to sample set. The closer these three indicators are to 1, the higher the similarity. Conversely, the closer the value is to 0, the lower the similarity.

4 Experimental Result

The performance of the proposed network model was quantitatively evaluated using the ADNI dataset. To verify the performance of the algorithm proposed in this study, the experiment used ten-fold cross-validation to analyze the segmentation results. Self-contrast experiments and comparisons with other methods were conducted. Using DSC, SEN, and PPV as evaluation indicators can quantitatively evaluate the accuracy of the segmentation results.

4.1 Influence of network model on experimental results

To verify the performance of the model, GAN, Res_GAN, SEblock_GAN, and Res_SEblock_GAN were compared. GAN is the original method of generating an adversarial network; Res_GAN is a generative adversarial network that uses a residual network in the coding structure of the generative model; SEblock_GAN is a generative adversarial network that adds a channel attention mechanism to the decoding structure of the generative model; and Res_SEblock_GAN is the method proposed in this study.

Figure 6 shows the segmentation results of each model; the first column is the image to be segmented, the second column is the expert segmentation result, the third column is the segmentation result of the basic GAN, the fourth column is the segmentation result of the Res_GAN, the fifth column is the segmentation result of the SEblock_GAN, and the sixth column is the segmentation result of the Res SEblock_GAN. We randomly listed the segmentation results of the hippocampus at different layers. As shown in Fig. 6, Res_SEblock_GAN could identify the segmentation details well, and the segmentation result was close to the expert segmentation result. Each row represents the segmentation results of different methods under this layer. Given that the original GAN does not introduce the attention mechanism, the segmentation effect in the detailed boundary area of the hippocampus is not ideal. By introducing the attention mechanism, the segmentation effect of the details was considerable improved, which was close to the result of expert segmentation. This result shows that Res_SEblock_GAN can pay attention to the details of the edge part by adjusting the weight of the feature channel, thereby expanding the effective features, suppressing useless information, and achieving optimal segmentation results.

Table 1 quantitatively evaluates the experimental results. Experiments show that the introduction of residual blocks and attention mechanisms in generative adversarial networks could effectively and accurately complete the task of hippocampus segmentation.

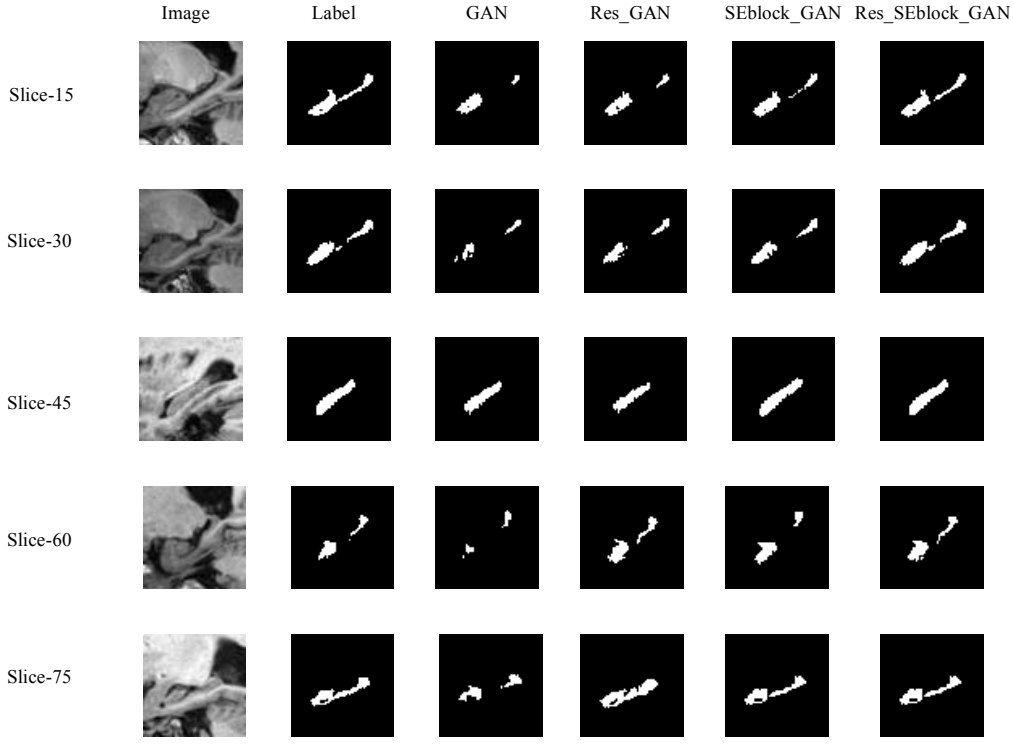


Fig. 6 Self-contrasting experiment results.

Table 1 Evaluation index results of each model.

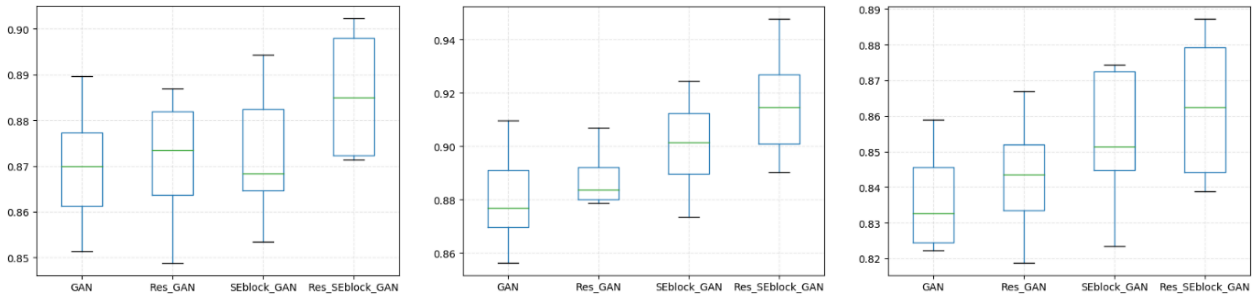
Method	DSC	SEN	PPV
GAN	87.34	87.42	88.96
Res GAN	88.03	88.14	89.89
SEblock_GAN	89.14	88.29	89.70
Res_SEblock_GAN	89.46	88.71	92.36

To compare the methods intuitively, the test results of Res_SEblock_GAN and the model before improvement were analyzed, as shown in Fig. 7. As shown in Fig. 7, Res_SEblock_GAN's DSC, PPV, and SEN indicators are better than that of the original GAN, indicating that the method in this study can suppress or reduce useless features while amplifying the effective features through the attention mechanism in the generation model to improve segmentation accuracy.

4.2 Comparative analysis with other algorithms

To verify the performance of the algorithm in this study, the segmentation algorithms of various methods on the ADNI dataset were compared. We randomly listed the segmentation results of the hippocampus at different layers.

As shown in Fig. 8, the first row is the label map segmented by experts, the second row is the segmentation result of the UNet network, the third row is the segmentation result of the Attention UNet, and the fourth row is the segmentation result of the network model proposed in this study. The red part represents the hippocampus result of expert segmentation, and the yellow part represents the hippocampus contour of the network segmentation. To compare the segmentation effects of each method clearly, the segmentation results of each segmentation method and that of the expert are merged. Figure 8 shows that in the method proposed in this study, the yellow line and red parts are more consistent than the other methods. For example, the position of the vacancy in the middle of the 65th floor is evident. The method proposed in this study captures



(a) DSC (b) PPV (c) SEN

Fig. 7 DSC, PPV, and SEN evaluation indicators of four models.

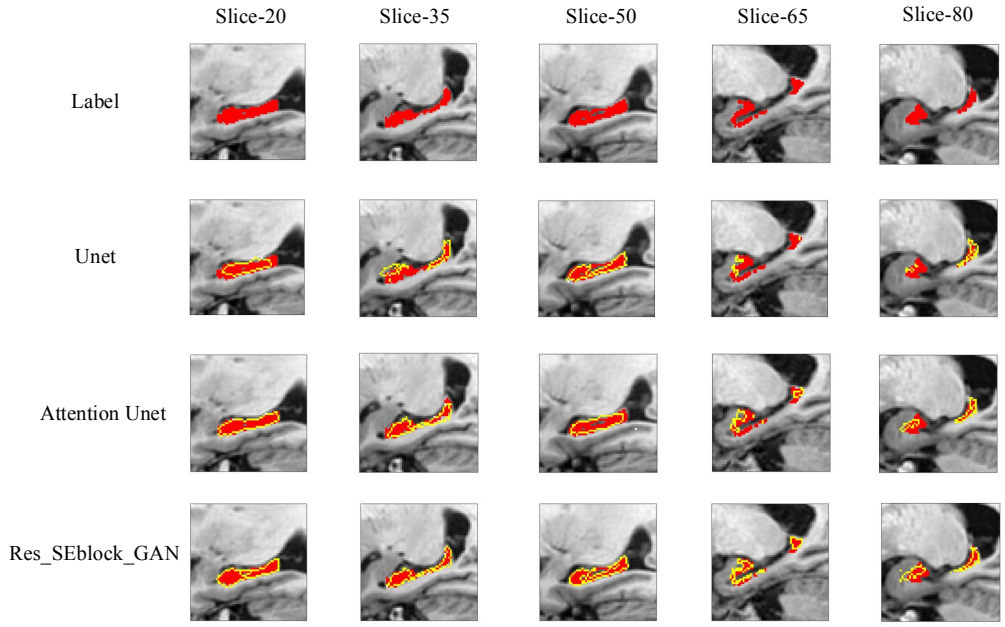


Fig. 8 Segmentation results of several models.

the details of the hippocampus and the overall outline. Figure 8 also shows that the network model proposed in this study can accurately achieve hippocampus segmentation.

Table 2 shows the comparison of the Dice and Jaccard coefficients between the method in this study and the method mentioned in Refs. [11, 35–37]. The Dice and Jaccard coefficients of the UNet model are 85.40%. The Dice and Jaccard coefficients of Attention UNet can reach 87.70% and 85.06%, respectively. The introduction of the residual block and attention mechanism increased the evaluation index by 2.3% and 3.03%, respectively. In the end, the evaluation index of Res_SEblock_GAN can reach 89.46% and 85.18%. By combining the adversarial model, the accuracy rate was improved. Experiments showed that the introduction of the residual block and attention mechanism in the generation process of the adversarial model can effectively and accurately complete the task of hippocampus segmentation. Through the proposed method, optimal segmentation precision is achieved. Experimental results showed that the proposed method can achieve good results in hippocampus segmentation.

5 Conclusion

To segment the hippocampus accurately and efficiently, a hippocampus segmentation method based on generative adversarial networks was proposed. In the generative model, ResNet34 was used as the coding structure to extract the shallow features of the image, thus reducing the processing time to a certain extent. The attention mechanism was added to the decoding structure to enhance the useful features and suppress the features that were not useful for the current task. Combined with the iterative training of the adversarial model, the generative model achieved the optimal segmentation state. Experiments show that the proposed model improves the accuracy of hippocampus segmentation; thus the effectiveness of the proposed model is verified. Compared with existing methods, the proposed model has the following advantages:

The proposed generative adversarial network can maintain the spatial consistency of the segmentation results generated by the generative model. The generated segmentation results are similar to real label maps, which can achieve the effect of false and true, and the edges of the segmentation area are smooth.

(1) ResNet34 is used as the coding structure of the generative model to extract the shallow features of the image, thereby reducing the processing time to a certain extent.

(2) By introducing an attention mechanism in the decoding part of the generative model, the network starts from the global information to amplify valuable feature channels selectively and suppress useless feature channels.

(3) The data used in this study are MRI images of the normal human hippocampus, exploring the difference between the hippocampus of the Alzheimer's syndrome patient and the normal human hippocampus. The automatic diagnosis of cases will be the focus of subsequent research.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Nos. 61873178 and 61976150), Natural Science Foundation of Shanxi Province (Nos. 201801D21135 and 201901D111091), Key Research and Development Projects of Shanxi Province (No. 201803D421047).

References

[1] S. Pei, J. Guan, and S. Zhou, Fusion analysis of resting-state networks and its lication to Alzheimer's disease, *Tsinghua Science and Technology*, vol. 24, no. 4, pp. 456–467, 2019.

[2] D. L. Collins and J. C. Pruessner, Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion, *NeuroImage*, vol. 52, no. 4, pp. 1355–1366,

Table 2 Comparison of accuracy of seven methods.

Method	Model	Dice coefficient result	Jaccard coefficient result
U-net	U-net	0.8540	0.8203
Morra et al.[35]	Ada-SVM	0.8180	0.8349
Chen et al.[11]	U-Seg-Net	0.8605	0.8413
Tong et al.[36]	DDLS	0.8720	0.8149
Song et al.[37]	Progressive SPBL	0.8835	0.8473
Attention Unet	Attention Unet	0.8770	0.8506
Our method	Res_SEblock_GAN	0.8946	0.8518

2010.

[3] A. R. Khan, N. Cherbuin, W. Wen, K. J. Anstey, P. Sachdev, and M. F. Beg, Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on hippocampus segmentation, *NeuroImage*, vol. 56, no. 1, pp. 126–139, 2011.

[4] F. Van der Lijn, T. den Heijer, M. M. B. Breteler, and W. J. Niessen, Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts, *NeuroImage*, vol. 43, no. 4, pp. 708–720, 2008.

[5] D. Zarpalas, P. Gkontra, P. Daras, and N. Maglaveras, Hippocampus segmentation by optimizing the local contribution of image and prior terms, through graph cuts and multi-atlas, in *Proc. 9th Int. Symp. Biomedical Imaging* (*ISBI*), Barcelona, Spain, 2012, pp. 1168–1171.

[6] M. Hajiesmaeili, B. Bagherinakhjavanlo, J. Dehmeshki, and T. Ellis, Segmentation of the Hippocampus for detection of Alzheimer's disease, in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M. H. Choi, S. Mantler, J. Schulze, D. Acevedo, et al., eds. Berlin, Germany: Springer, 2012, pp. 42–54.

[7] S. Pei, J. Guan, and S. Zhou, Fusion analysis of resting-state networks and its lication to Alzheimer's disease, *Tsinghua Science and Technology*, vol. 24, no. 4, pp. 456–467, 2019.

[8] E. A. A. Alaoui, S. C. K. Tekouabou, S. Hartini, Z. Rustam, H. Silkan, and S. Agoujil, Improvement in automated diagnosis of soft tissues tumors using machine learning, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 33–46, 2021.

[9] E. Shelhamer, J. Long, and T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 39, no. 4, pp. 640–651, 2017.

[10] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Proc. 18th Int. Conf. Medical Image Computing and Computer-Assisted Intervention MICCAI*, Munich, Germany, 2015, pp. 234–241.

[11] Y. N. Chen, B. B. Shi, Z. W. Wang, P. Zhang, C. D. Smith, and J. D. Liu, Hippocampus segmentation through multi-view ensemble ConvNets, in *Proc. 14th IEEE Int. Symp. Biomedical Imaging* (*ISBI 2017*), Melbourne, Australia, 2017, pp. 192–196.

[12] Y. N. Chen, B. B. Shi, Z. W. Wang, T. Sun, C. D. Smith, and J. D. Liu, Accurate and consistent hippocampus segmentation through convolutional LSTM and view ensemble, in *Int. Workshop on Machine Learning in Medical Imaging*, Q. Wang, Y. H. Shi, H. I. Suk, and K. Suzuki, eds. Cham, Germany: Springer, 2017, pp. 88–96.

[13] L. Cao, L. Li, J. F. Zheng, X. Fan, F. Yin, H. Shen, and J. Zhang, Multi-task neural networks for joint hippocampus segmentation and clinical score regression, *Multimed. Tools Appl.*, vol. 77, no. 22, pp. 29669–29686, 2018.

[14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in *Medical Image Computing and Computer-Assisted Intervention*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds. Cham, Germany: Springer, 2016, pp. 424–432.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in *Proc. 27th Int. Conf. Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 2672–2680.

[16] M. Mirza and S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784, 2014.

[17] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets, in *Proc. 30th Int. Conf. Neural Information Processing Systems*, Barcelona, Spain, 2016, pp. 2180–2188.

[18] M. Arjovsky, S. Chintala, and L. Bottou, Wasserstein generative adversarial networks, in *Proc. 34th Int. Conf. Machine Learning* (*ICML*), Sydney, Australia, vol. 70, 2017, pp. 214–223.

[19] T. Che, Y. R. Li, A. P. Jacob, T. Bengio, and W. J. Li, Mode regularized generative adversarial networks, in *Int. Conf. Learning Representations* (*ICLR*), Toulon, France, 2017.

[20] Z. Zhang, G. Fu, R. Ni, J. Liu, and X. Yang, A generative method for steganography by cover synthesis with auxiliary semantics, *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 516–527, 2020.

[21] X. Wu, K. Xu, and P. Hall, A survey of image synthesis and editing with generative adversarial networks, *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 660–674, 2017.

[22] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, Semantic segmentation using adversarial networks, presented at Computer Science-Computer Vision and Pattern Recognition, NIPS Workshop on Adversarial Training, Barcelona, Spain, 2016.

[23] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. L. Huang, SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation, *Neuroinformatics*, vol. 16, no. 3, pp. 383–392, 2018.

[24] T. Neff, C. Payer, D. Štern, and M. Urschler, Generative adversarial network based synthesis for supervised medical image segmentation, in *OAGM & ARW Joint Workshop*, Vienna, Austria, 2017.

[25] A. K. Mondal, J. Dolz, and C Desrosiers, Few-shot 3d multi-modal medical image segmentation using generative adversarial learning, arXiv preprint arXiv: 1810.12241, 2018.

[26] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. Van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[27] B. Murugesan, K. Sarveswaran, R. S. Vijaya, S. M. Shankaranarayana, K. Ram, and M. Sivaprakasam, A context based deep learning approach for unbalanced medical image segmentation, presented at 2020 IEEE 17th Int. Symp. Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020.

[28] S. Izadi, Z. Mirikharaji, J. Kawahara, and G. Hamarneh, Generative adversarial networks to segment skin lesions, presented at 2018 IEEE 15th Int. Symp. Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 881–884.

[29] Z. J. Li, Y. Y. Wang, and J. H. Yu, Brain tumor segmentation using an adversarial network, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, eds. Cham, Germany: Springer, 2018, pp. 123–132.

[30] Y. G. Shi, K. Cheng, and Z. W. Liu. Hippocampal subfields segmentation in brain MR images using generative adversarial networks, *Biomed. Eng. Online*, vol. 18, no. 1, p. 5, 2019.

[31] B. Hui, Y. Liu, J. Qiu, L. Cao, L. Ji, and Z. He, Study of texture segmentation and classification for grading small hepatocellular carcinoma based on CT images, *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 199–207, 2021.

[32] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, presented at 2017 IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242–2251.

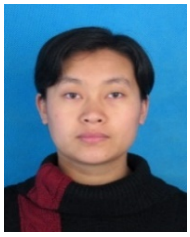
[33] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, Deep residual learning for image recognition, presented at 2016 IEEE Conf. Computer Vision & Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770–778.

[34] J. Hu, L. Shen, S. Albanie, G. Sun, and E. H. Wu, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, 2020.

[35] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*), Boston, MA, USA, 2015, pp. 1–9.

[36] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, and D. Rueckert, Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling, *NeuroImage*, vol. 76, pp. 11–23, 2013.

[37] Y. T. Song, G. R. Wu, K. Bahrami, Q. S. Sun, and D. G. Shen, Progressive multi-atlas label fusion by dictionary evolution, *Med. Image Anal.*, vol. 36, pp. 162–171, 2017.



Hongxia Deng recived the PhD degree from Taiyuan University of Technology in 2013. Since 2013, she has been an associate professor in the School of Computer Science and Technology, Taiyuan University of Technology. Since September 2019, she has studied in the School of Computer Science, Zhejiang University as a visiting scholar in China. Her research interest includes computer vision, image processing, big data intelligent analysis, and brain science.

Yuefang Zhang is a master student at Taiyuan University of Technology now. Her main research interests include computer vision, and medical image processing.

Ran Li is a master student at Taiyuan University of Technology now. Her main research interests include computer vision and image processing.

Chunxiang Hu is a master student at Taiyuan University of Technology now. Her main research interests include computer vision and medical image processing.

Zijian Feng is a master student at Taiyuan University of Technology now. His main research interests include computer vision and image processing.

Haifang Li recived the PhD degree from Taiyuan University of Technology in 2009. Her research interests include big data intelligent analysis, computer vision, and affective computing.