

Received June 16, 2021, accepted August 25, 2021, date of publication September 16, 2021, date of current version October 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3113309

Brain Extraction From Brain MRI Images Based on Wasserstein GAN and O-Net

SHAOFENG JIANG^{1,2}, LANTING GUO¹, GUANGBIN CHENG³, XINGYAN CHEN¹,
CONGXUAN ZHANG^{2,4}, (Member, IEEE), AND ZHEN CHEN^{1,4}

¹Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang Hangkong University, Nanchang 330063, China

²Department of Biomedical Engineering, Nanchang Hangkong University, Nanchang 330063, China

³Guangdong Armed Police Corps Hospital, Guangzhou, Guangdong 510517, China

⁴Key Laboratory of Jiangxi Province for Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China

Corresponding author: Shaofeng Jiang (jsphone@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61162023, and in part by the Natural Science Foundation of Jiangxi Province under Grant 20192BAB205083.

ABSTRACT Brain extraction is an essential pre-processing step for neuroimaging analysis. It is difficult to achieve high-precision extraction from low-quality brain MRI images with artifacts and gray inconsistencies which often result in irregular hole regions in the extracted brain tissues. In addition, the U-Net based brain extraction method trends to output over-smoothed brain boundary. To remove those irregular holes in the extracted mask, we proposed a new U-Net based model for brain extraction named O-Net. O-Net replaces the skip-connection path in the U-Net with dual shortcut paths including an attention module of an O-shaped network, which uses deep semantic information to highlight the target area while retaining more image details. O-Net effectively reduces the impact of intensity differences caused by artifacts or gray inconsistencies in the brain MRI images on the extraction results. To more accurately identify brain boundary, we designed a new GAN based brain extraction method, which used above O-Net as the segmentation network. The discrimination network of the proposed GAN model adopts the residual structure to enhance the nonlinear expression ability of the network to balance the adversarial training of the two networks. To speed up the convergence of the proposed model, a segmentation loss was added to the adversarial loss to supervise the feature learning of the segmentation network. This method was compared with other popular brain extraction methods on two public datasets (IBSR18 and LPBA40). The mean Dice similarity coefficients obtained by the proposed method were 97.26% and 98.29% on IBSR18 and LPBA40 respectively. In the comparative experiment, the results of the proposed method are the best on the two public datasets. Experimental results show that the proposed model can stably output high-precision brain tissue extraction images and the influence of artifacts and gray inconsistencies is small.

INDEX TERMS Brain extraction, artifacts, gray inconsistencies, dual shortcut paths, O-Net.

I. INTRODUCTION

With the wide application of magnetic resonance imaging (MRI) equipment in clinical medicine, neuroimaging analysis becomes more and more popular in the fields of brain disease diagnosis and brain function analysis. Brain extraction is an important pre-processing step in most neuroimaging analysis, having a crucial impact on subsequent research such as registration of brain MRI, measurement of brain volume, and brain tissue classification. Therefore, an accurate and

stable brain extraction method is needed in neuroimaging analysis.

Manual extraction method has the highest accuracy, but it consumes a lot of time and energy. Although the accuracy of automatic method is lower than the manual method, it is more efficient. The automatic extraction method is affected by many factors, which is still a challenging field. First, brain MRI images usually show low contrast, low resolution, and uneven gray distribution due to the diversity of devices and imaging protocols. Furthermore, brain structure is complex, so it is difficult to determine the boundary between brain tissue regions and non-brain tissue regions. Second, artifacts caused by the patient's unconscious movement and

The associate editor coordinating the review of this manuscript and approving it for publication was Ziyang Wu.

equipment debugging are manifested as distorted, overlapped, lost, and blurred in brain MRI images. These artifacts may prevent the model from extracting the correct target features, leading to serious deviations in the results. Last, the intensity difference between individuals makes it difficult for one method to constantly output good results for different brain scans. Pathological changes in brain may cause morphological changes, which will increase the difficulty of accurately extraction. With the in-depth study of brain extraction algorithms, the problems of low resolution, low contrast and uneven gray level of image have been well solved. However, these models still can't solve the artifact problem well, and their adaptive performance is still a problem that needs to be further studied.

Since the development of the brain extraction research, a lot of automatic methods have been proposed. These methods can be divided into the classic-based [1]–[7], the atlas-based [8], [9], and the learning-based [10]–[20]. Some classic automatic methods are yet widely used because of their advantages in high calculation speed and batch processing of data. Smith [1] proposed BET which pushes the grid points to the brain boundary by a set of local adaptive forces. BET has stable performance and can handle some complex areas in brain MRI image well such as eyeballs. Therefore, many researchers have proposed improved brain extraction methods based on BET [2]–[4]. However, these classical methods are highly dependent on parameter setting, and their extraction accuracy is not enough to meet clinical requirements. Compared with the classical method, the atlas-based brain extraction method has a good performance in terms of accuracy and stability, yet this kind of method involves registration technology. If there is a large error in the registration, the subsequent operation will be affected. ROBEX proposed by Iglesias *et al.* [8] is known for stability. However, if its registration step fails, the extraction work can't continue. Moreover, the performance of the atlas-based method is also restricted by the slow registration speed.

Learning-based methods can be divided into machine learning-based and deep learning-based according to the way of feature extraction. Many machine learning algorithms have been applied to brain extraction task, such as hidden Markov algorithm [10], [11], meta-algorithm [12], K-mean [13], and Bayesian [14]. However, all most features need to be determined by experts, and then manually coded in machine learning-based methods. As the database is consciously collected and established by people, deep learning based on a large amount of data can be studied deeply and used widely. Deep learning tries to learn feature from the data itself. Among them, the convolutional neural network (CNN) extracts the shallow and deep semantic information of the image by establishing multiple hidden layers. The CNN-based method has low sensitivity to low contrast, low resolution and uneven gray-scale of brain MRI image. Kleesiek *et al.* [15] firstly used an end-to-end extraction method based on 3D convolutional network for skull stripping. As the most widely used model in medical image

segmentation, U-Net proposed by Ronneberger *et al.* [16] is the basic network of many brain extraction models. For example, Salehi *et al.* [17] proposed Auto Context U-Net, which added a local predictive brain mask to U-Net to obtain higher-precision results than Kleesiek *et al.* [15]. Hwang *et al.* [18] extended U-Net from 2D pixels to 3D voxels to make full use of more spatial information for skull separation. However, U-Net can't handle data with large different in the data sets. The predicted brain mask obtained by U-Net on a data set with severe artifacts or different intensity will have large-scale irregular missing region. Most of brain extraction methods based on U-Net are very likely to have the same problems as mentioned above.

Generative adversarial networks (GAN) are well known for the powerful data fitting capabilities and the adversarial training way that are different from others network. In the field of image segmentation, researchers make use of adversarial training to promote the segmentation network to better learn the mapping between samples, thereby improving the accuracy of segmentation. The model combining GAN and CNN has been widely used in the field of medical images segmentation [19]–[21]. Chen *et al.* [19] combined dense expansion network and GAN to realize neuron cells. Moeskops *et al.* [20] used the same idea as Chen to achieve brain tissue segmentation. Thirumagal and Saruladha [21] combined ResNet [22] and GAN to construct FCSE-GAN for accurate segmentation of brain tumor regions. At present, there are few researches on the whole brain segmentation based on combined network.

Based on the above observation and analysis, this paper proposes a new brain extraction model WGAN + O-Net, where WGAN (Wasserstein GAN) [23] performs adversarial training stably to improve the accuracy of our proposed segmentation network O-Net. O-Net introduces attention modules into U-Net to form a new shortcut to connect the corresponding feature mapping on encoding and decoding path. This structure can not only preserve the more detailed feature of the image, but also highlight the target area of each channel by using the deep semantic information. O-Net can effectively reduce the influence of intensity difference caused by artifacts or gray inconsistencies in the extraction results. At the same time, in WGAN, the residual structure is used in the discriminator to enhance the nonlinear performance of the network and balance the adversarial training of the two networks.

In the remainder of this paper, we first review the related work of the proposed brain extraction method, then give a detailed description of the proposed method, and finally verify the performance of the model through experiments on healthy brain MRI and pathological brain MRI.

II. RELATED WORK

A. ATTENTION MECHANISM

The attention mechanism in deep learning is similar to selective visual attention mechanism of human beings. Its essence

is to locate interest information and suppress useless information. The results are usually displayed in the form of probability graphs or probability feature vectors. The attention model can be divided into three categories according to the principle: spatial attention model, channel attention model, and spatial channel mixed attention model. The spatial attention model includes AG by Oktay *et al.* [24], STN proposed by DeepMind [25], and the SAM proposed by Zhu *et al.* [26]. The channel attention model includes the SENet proposed by Hu *et al.* [27], and Wang *et al.* [28] improved ECANet based on SENet. The attention modules of hybrid spatial and channel are CBAM proposed by Woo *et al.* [29] and SANet by Zhang and Yang [30].

The advantage of these attention models is that they can be easily added to the CNN network model without causing large-scale changes to the model structure. For example, the Attention U-Net proposed by Oktay *et al.* [24] directly adds attention gate (AG) between the skip-connections of U-Net to supervise the previous-level feature map with the next-level feature map. AG belongs to the spatial attention model, the importance of features at different spatial positions can be controlled in U-Net model training. This allows AG to suppress the characteristic response of irrelevant regions.

B. WGAN-DIV

The emergence of GAN has opened up the research ideas of unsupervised learning on complex distributions, but its training has the problem of model collapse. To solve this problem, many improved GANs have been proposed. Wasserstein distance [23] was proposed for the difficulty of GAN network training. The training of WGAN is easier than the original GAN, and eliminates mode collapse to a certain extent. However, WGAN introduces the Lipschitz (L) constraint.

To satisfy the L -constraint, many methods have been proposed. The original solution given by Zhang and Yang [30] is weight clipping. Gulrajani *et al.* [31] proposed gradient penalty to discriminator. Miyato *et al.* [32] proposed spectrum normalization. However, these methods have a problem: restrict discriminator to a small cluster of functions. In this case, WGAN-DIV proposed by Wu *et al.* [33] proved that the optimization objective of the discriminator in WGAN-GP is not always divergence, and proposed a theoretically complete wasserstein divergence. Wasserstein divergence can satisfy the L -constraint while retaining the excellent properties of wasserstein distance for stable training. The definition of wasserstein divergence is as follows:

$$W_{k,p}(P_r, P_q) = \max_{D \in C^1} E_{x \sim P_r}[D(x)] - E_{x \sim P_f}[D(x)] - kE_{x \sim P_{Random}}[||\nabla_x D(x)||^p] \quad (1)$$

where C^1 represents the first-order continuous function; r represents the true distribution, and q represents the composite distribution. WGAN-div has very low requirements for P_{random} distribution, and the effects of various sampling methods are similar. After a series of experiments, the author

found that the effect is best when $K = 2$ and $P = 6$. Based on the above analysis, we choose the Wasserstein divergence in WGAN-DIV as the loss function of our generative adversarial network.

III. METHOD

The general architecture of the WGAN+O-Net model we proposed is illustrated in Figure 1. The network can be divided into two parts: segmentation network and discrimination network. We introduce the above two networks in turn, and then introduce the adversarial training of the network.

A. SEGMENTATION NETWORK

In the proposed brain extraction model, the purpose of the segmentation network is to output a predictive brain mask that can replace manual labels to deceive the discrimination network. The encoding-decoding architecture of U-Net with skip connection can integrate features of different scales and has few of parameters, so it is very suitable for the data mining with simpler semantics and relatively fixed structure such as medical images. Based on U-Net, we proposed O-Net to improve brain extraction results of brain MRI with serious artifacts or inconsistent gray-scale distribution.

In the proposed network, the encoder has 8 convolutional components, each of which includes a convolution layer with kernel size of 4×4 and a stride of 2 + BN layer + Leaky Relu activation function layer + a convolution layer with kernel size of 3×3 and a stride of 1 + BN layer + Leaky Relu activation function layer. The max-pooling layer is replaced by a convolution layer with a stride of 2 to achieve down-sampling in the entire coding path. The initial number of feature channels is 64, and then increases exponentially until it reaches 512. Obviously, the proposed network is deeper than the baseline U-Net. The deepened network can provide a wider range of receptive fields and more feature maps of different scales. However, as deepening the network more spatial feature information is lost with the down-sampling. Therefore, we add an attention path between the encoder and the decoder based on the original skip connection to locate the target region features of different scale feature maps. Furthermore, two parallel skip connection paths can retain more detailed features in deep model training. As shown in Figure 1, the skip connection path and the attention path come to form a closed path which looks like a letter 'O'. Therefore, we named it O-Net.

The key components of the attention path are the attention module which is marked with a capital letter A in the middle of the red-filled circle in the Figure 1. Each attention module has two inputs (x and g). x is fed by the shallow feature map from the encoding path. g is fed by the feature map after transposed convolution of the output of the previous level in the decoding path. The output of the attention module is labeled x' . x and x' are concatenated and then undergoes a convolution with the kernel size of 3×3 . The output of the convolution become the input g of the next attention module. The detailed internal structure of attention module

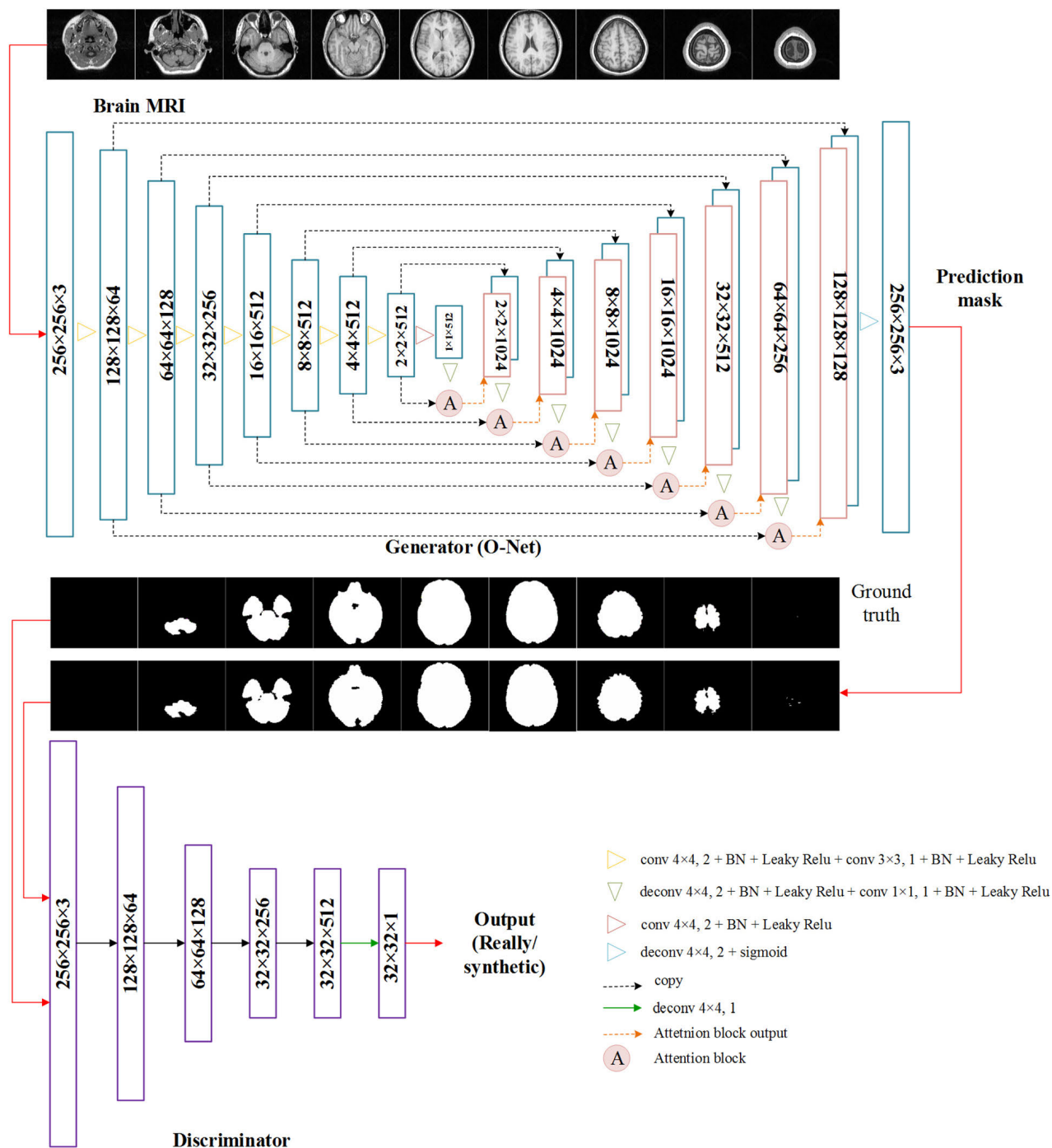


FIGURE 1. Schematic diagram of the proposed networks: 'A' in the generator represents the attention module, and the detailed structure is shown in Figure 2(a). The discriminator is composed of a superposition of residual blocks, and the detailed structure is shown in Figure 2(b).

is shown in Figure 2(a). First, the two inputs (x and g) are pixel-wise superimposed to control the importance between the feature information of the same position in x through the high-level semantic information in g . Then, the irrelevant features in the feature map are suppressed by a convolution layer with kernel size of $3 \times 3 + \text{BN}$ layer + Relu activation function layer. The value of the obtained attention matrix is normalized to between $0 \sim 1$ by a convolution layer with

a kernel of $3 \times 3 + \text{BN}$ layer + Sigmoid activation function layer. Finally, the obtained attention coefficient matrix of each channel is multiplied by the corresponding channel of x to output a new feature map x' . The attention module uses deeper semantic information to guide the feature map of the current layer to adjust feature weights. Therefore, the size of the coding path down-sampling to the feature map is set to 1×1 to assist the model to obtain a wider receptive field.

The attention module proposed in this paper draws on the idea of attention gate (AG) in attention U-Net, but it is different from AG in internal structure and attention matrix. In the module for suppressing useless feature and normalization, we use a convolutional layer with a kernel size of 3×3 to obtain more features to distinguish it from the feature information provided by another skip-connection path and provide more feature for the decoding path. At the same time, this operation also extends the receptive field of the attention module to a deeper level so that it is not limited to the features of the previous layer and the current layer. The number of characteristic channels of AG changes as follows: C (input) $\rightarrow C / 2$ (after the first convolution) $\rightarrow 1$ (after the second convolution). The number of channels of the proposed attention model changes as follows: C (input) $\rightarrow C / 2$ (after the first convolution) $\rightarrow C$ (after the second convolution). The scale of the attention coefficient matrix obtained by AG is $H \times W \times 1$, which means that the feature map of each channel of x employs the same attention matrix to adjust the weight space information. Although the attention matrix output by the AG integrates the information of all channels, using the same attention for all channels will lead to deviations in the range of the prominent target area. The attention coefficient matrix scale obtained by the proposed attention model is $H \times W \times C$, which means that the feature map on each channel has its own corresponding attention matrix. Adjust the proportion of features according to different semantic information represented on different channels. This targeted adjustment helpful to better understand the feature distribution of the image.

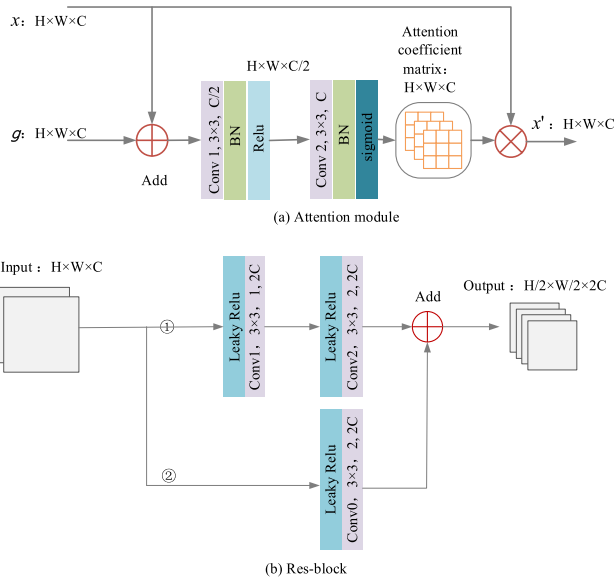


FIGURE 2. Details of the proposed network architecture: (a) Attention module, where g is the target feature map, x is the source feature map (b) residual block, where x_l represents the input of each residual layer.

The purpose of the decoder is to restore the low-resolution feature map with high-level semantic information to the same resolution as the input image. The network structure is

symmetrical, so the decoder needs to be up-sampled 8 times. After the two input feature maps of the decoder are concatenated by channel, they are sent to a convolution with kernel size of $3 \times 3 + \text{BN layer} + \text{Leaky Relu activation function layer}$ for dimension reduction and feature fusion. The output is used as the input of the next attention module after transposed convolution. The dual-skip path facilitates propagation and update of gradients and provides more detailed feature information. Therefore, the proposed decoding architecture makes the edge of the output prediction mask more fine. The proposed O-Net has a feature map with a larger receptive field to deal with artifacts, which are usually different in images and do not exist only in a certain region of the image. At the same time, the attention module in the model obtains image context information to increase the weight of brain tissue regions. The purpose of the final model is to weakening the influence of artifacts.

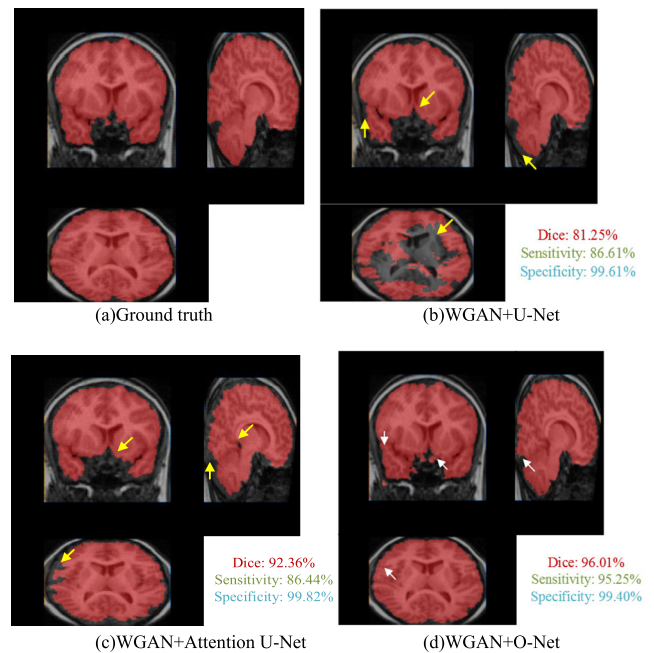


FIGURE 3. Overlay of the prediction mask and the original image for each improvement stage of the model. The brain MRI image is from IBSR18, which has artifacts. In the figure, the yellow arrows point to the defects of the prediction mask, and the white arrows indicate that the prediction masks of these parts have been greatly improved compared with the models of other stages.

To show how O-Net can improves brain extraction, we show the outputs of a typical image in three processing ways in Figure 3. They all used the WGAN-DIV, but with different generative networks. The WGAN+U-Net, WGAN+Attention U-Net, and WGAN+O-Net used the U-Net, the Attention U-Net, and the O-Net as the generative network respectively. In Figure 3(b), affected by the artifacts in the image, a lot of brain tissues are missed in the middle of the extraction results obtained by WGAN+U-Net. By adding the attention module into the U-Net, the large-scale missed areas in Figure 3(b) are recovered in Figure 3(c), but the accuracy has not been significantly improved.

By providing more detailed information for the up-sampling of the network, our proposed O-Net not only avoided holing in the prediction mask but also significantly improved the accuracy.

B. DISCRIMINATION NETWORK

The discrimination network learns the difference between the ground truth and the output of the segmentation network to effectively punish the segmentation network. It provides a learning signal for the segmentation network so that the segmentation network can output the predictive brain mask closest to the ground truth. Therefore, the stronger the discriminating ability of the discrimination network is, the better the segmentation network is. We deepened the depth of the discrimination network as well as the segmentation network to enhance the discrimination ability and balance the adversarial training of the two networks. The modified convolution residual blocks (ResNet v2 [34], being qualified to build more deeper network) were used to form the layers in the discrimination network in Figure 1. The internal structure of the modified res-block is shown in Figure 2(b). The residual block consists of two convolutional paths including multiple Leaky Relu, convolutional layer with kernel size of 3×3 and 2×2 max-pooling convolutional components. The output of the left path including two 3×3 convolutional layers have the same receptive field as the output of one 5×5 convolutional layer, but having fewer parameters. The outputs of the two convolutional paths are added, thus the residual block can provide multiple receptive fields. Instead of using the BN layer in the discriminator, we removed the last activation layer in the discriminator to ensure that the model can be trained stably.

C. ADVERSARIAL TRAINING

The generative adversarial network is trained alternatively to minimize the objective function. Adversarial training is implemented iteratively by two iterative steps. First, the adversarial model is trained in order to improve the model's ability to discern the authenticity. The loss of discriminator is defined as:

$$L_D = \arg \max_{D \in C^1} E_{x \sim P_r} [D(x)] - E_{G(z) \sim P_g} [D(G(z))] - k E_{\hat{x} \sim P_{\hat{x}}} [(|\nabla_x D(\hat{x})|^p)] \quad (2)$$

where C^1 refers to the first-order continuous function family, \hat{x} is the sampling point on the line between the real distribution P_r and the generator distribution P_g , ∇ is the gradient operation in connection with the discriminator.

The generator of the model is trained in the second step. In order to speed up the convergence of the model and better cooperate with the segmentation network, we added the segmentation loss function L_S to generator's loss function.

$$L_G = \arg \min_G E_{G(z) \sim P_g} [D(G(Z))] + L_S \quad (3)$$

$$L_S = \sum_{n=1}^N l_{mce}(G(z), x) \quad (4)$$

where l_{mce} is class cross-entropy and x is the true distribution data. Cost minimization on 10 epochs was performed using ADAM optimizer with an initial learning rate of 0.0001 on both the segmentation network and the discrimination network. All model performance evaluation experiments were done on workstations with Nvidia eforce GTX1080Ti.

IV. EXPERIMENTS

A. DATASETS

We evaluated and verified the performance of proposed model on three brain MRI image data sets, two of which are public and the other is private.

IBSR18: come from the International Brain Tissue Segmentation Image Library, dedicated to the study of brain extraction algorithms. The 18T1 scans of IBSR18 come from healthy human with a resolution of $0.94 \times 1.5 \times 0.94$ mm. The 2D image size in three directions of this dataset (cross section, coronal plane and sagittal plane) are 256×128 , 256×256 and 128×256 , and the corresponding numbers are 4068, 2304, and 4068. Overall quality Poor, with severe artifacts.

LPBA40: come from the LONI website of the University of Southern California Los Angeles. The 40T1 scans of LPBA40 come from healthy humans with a resolution of $0.86 \times 1.5 \times 0.86$ mm. The 2D image size in three directions of this dataset (cross section, coronal plane and sagittal plane) are 256×124 , 256×256 and 124×256 , and the corresponding numbers are 10240, 4952 and 10240. Overall quality is clearer and better.

The brain MRI image is a small sample data set, so the different division of the data would greatly affect the effect of the model. To reflect the true level of the model as much as possible, we used the K-fold cross-validation method for sampling, in which LPBA40 was 10-fold cross-validation (training: test = 9:1) and IBSR18 was 9-fold cross-validation (training: test = 8:1).

Dataset with lesions: come from the hospital, which is a set of MRI images with meningiomas, cerebral ischemic stroke, and pituitary tumor. There are 30T1 scans sets, and each type of lesion contains 10T1 scans sets. Each scan set contains only 15 slices. The image dimensions of these slices are 256×256 . The labels corresponding to the data set were marked by the researchers under the guidance in this paper. In order to avoid uncertainty, they were only used for cross-dataset experiments to test the adaptability of the model.

B. EVALUATION METRICS

To quantitatively evaluate the extraction methods, three evaluation indexes Dice, sensitivity, and specificity, were calculated Dice coefficient was used to measure the similarity

between prediction mask and ground truth:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (5)$$

where TP is true positive, FP is false positive, and FN is false negative. TP is the total number of pixels correctly classified as brain tissue in the prediction label. FP is the total number of pixels incorrectly classified as brain tissue in the prediction label. FN is the total number of pixels incorrectly classified as non-brain tissue.

Sensitivity represents the ability of brain extraction methods to correctly recognize brain tissue:

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

Specificity represents the ability of brain extraction methods to correctly recognize non-brain tissues.

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

The value of Dice coefficient, sensitivity, and specificity range from 0 to 1. The larger the values of these three evaluation coefficients are, the more accurate the brain tissue extraction results are.

C. RESULTS

We evaluated the performance of the proposed models (WGAN+O-Net) through several comparative experiments with some popular brain extraction algorithms and similar model. The brain extraction methods involved in the comparative experiment include the baseline model (U-Net [16]), the deep learning-based method specially proposed for the brain extraction (Kleesiek *et al.* [15], Auto Context U-Net [17]), and three non-deep learning methods (HFEM-E [35], BET [2], and ROBEX [8]). Similar models include WGAN+Attention U-Net, WGAN+U-Net, and Pix2pix+U-Net. The evaluation results of each method on IBSR18 and LPBA40 are listed in Table 1 respectively.

Compared with other algorithms specifically proposed for brain extraction tasks, WGAN+O-Net obtained the highest Dice score in the two public datasets. On IBSR18, the Dice score of WGAN+O-Net is significantly improved compared with other methods (U-Net: 2.4%, Auto Context U-Net: 4.57%, Kleesiek: 0.94%, HFEM-E: 1.76%, BET: 4.91%, and ROBEX: 5.58%). On LPBA40, the difference between the scores of WGAN+O-Net and all other methods is more than 0.5% (U-Net: 0.61%; Auto Context U-Net: 0.56%; Kleesiek: 1.33%; BET: 4.77%; and ROBEX: 1.73%). We obtained the p-value after paired T-test to show the statistical difference between the proposed method and other methods in three evaluation indexes. On the two public datasets, all p-values

TABLE 1. The mean and standard deviation of the three evaluations of the two public datasets (LPBA40 and IBSR18) with different brain extraction methods. Figures 4 and 6 are supplementary figures of table 1.

	IBSR18		
	Dice (%)	Sensitivity (%)	Specificity (%)
HFEM-E	95.5(±0.015)	96.9(±0.054)	97.6(±0.0041)
P-value	1.9×10^{-03}	3.7×10^{-03}	—
BET	92.35(±0.037)	98.49(±0.013)	97.93(±0.009)
P-value	2.1×10^{-03}	1.116×10^{-01}	5.01×10^{-02}
ROBEX	91.68(±0.029)	99.28(±0.005)	97.42(±0.016)
P-value	1.0593×10^{-05}	3.1734×10^{-06}	3.3×10^{-03}
U-Net	94.86(±0.014)	94.64(±0.038)	99.20(±0.006)
P-value	3.3159×10^{-05}	1.28×10^{-02}	4.188×10^{-01}
Auto Context-U-Net	92.69(±0.012)	97.93(±0.110)	98.10(±0.003)
P-value	6.0541×10^{-07}	8.2836×10^{-05}	7.21×10^{-02}
Kleesiek	96.32(±0.010)	94.25(±0.030)	99.36(±0.003)
WGAN+At-U-Net	93.20(±0.006)	93.20(±0.027)	99.26(±0.0016)
WGAN+U-Net	96.00(±0.015)	95.34(±0.032)	99.54(±0.002)
Pix2pix+U-Net	95.50(±0.0094)	96.33(±0.0105)	97.41(±0.0137)
WGAN+O-Net	97.26(±0.006)	97.76(±0.009)	98.87(±0.008)
	LPBA40		
	Dice (%)	Sensitivity (%)	Specificity (%)
BET	93.52(±0.028)	98.59(±0.005)	97.76(±0.012)
P-value	4.0173×10^{-06}	2.609×10^{-07}	5.4354×10^{-06}
ROBEX	96.56(±0.003)	95.81(±0.008)	99.52(±0.001)
P-value	1.5915×10^{-06}	5.229×10^{-01}	4×10^{-03}
U-Net	97.61(±0.007)	97.30(±0.015)	99.64(±0.001)
P-value	7.5×10^{-03}	7.2×10^{-03}	5.62×10^{-01}
Auto Context-U-Net	97.73(±0.003)	98.31(±0.006)	99.48(±0.001)
P-value	8.3803×10^{-04}	1.742×10^{-01}	8.8495×10^{-04}
kleesiek	96.96(±0.01)	97.46(±0.03)	99.41(±0.003)
WGAN+At-U-Net	97.33(±0.0039)	99.27(±0.0035)	99.18(±0.0019)
WGAN+U-Net	97.44(±0.0035)	99.32(±0.0024)	99.19(±0.0022)
Pix2pix+U-Net	97.79(±0.0042)	98.60(±0.0080)	99.44(±0.0020)
WGAN+O-Net	98.29(±0.004)	99.00(±0.0004)	99.67(±0.002)

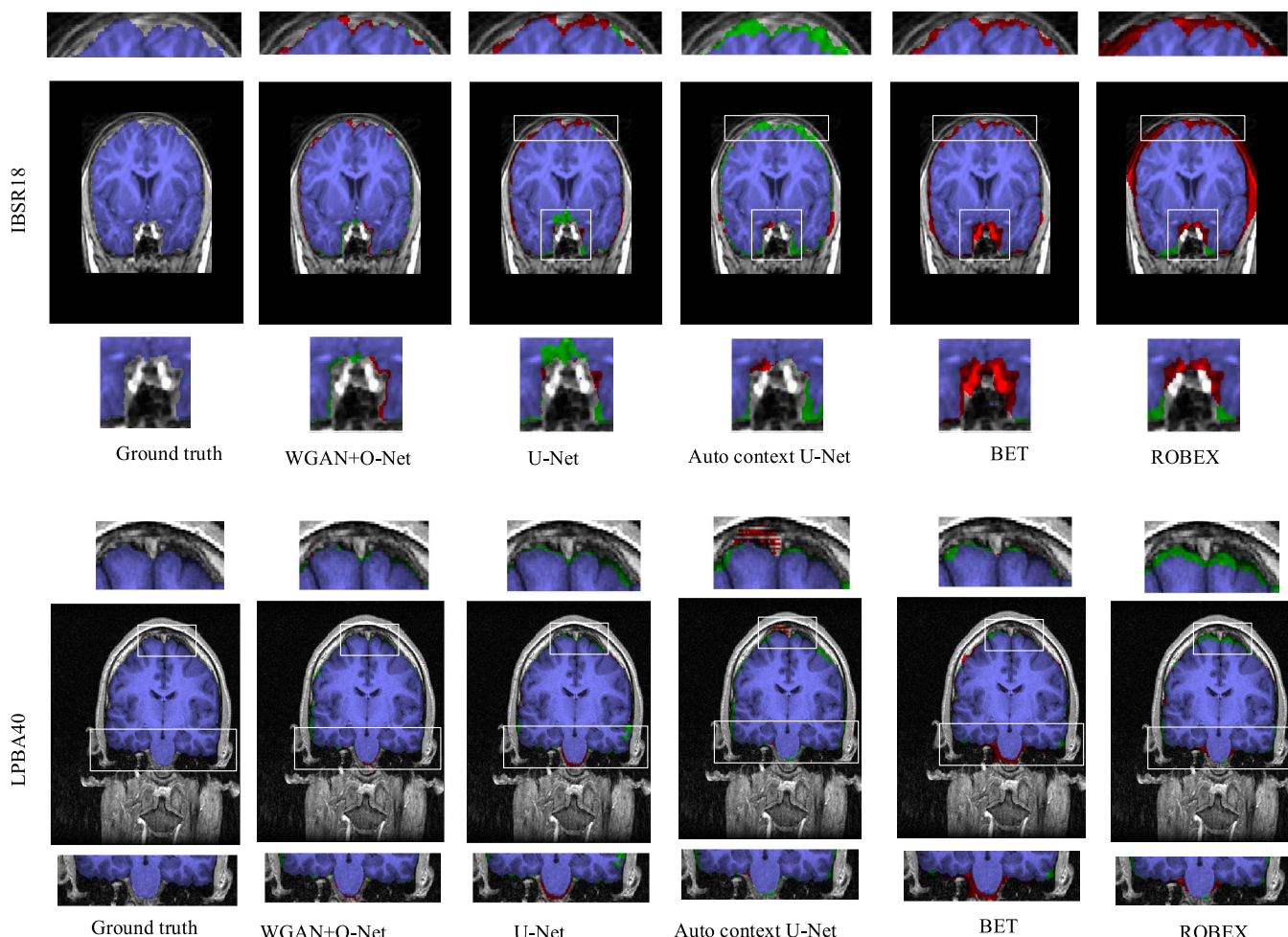


FIGURE 4. The overlays of two typical MRI scans from LPBA40 and IBSR18 and their corresponding prediction masks produced by various brain extraction methods Blue pixels represent the segmentation results of the corresponding brain extraction methods. Green pixels indicate the false negatives and red pixels indicate the false positives. Regions that are difficult to segment are intercepted and zoomed with rectangles. WGAN+O-Net produces more accurate brain segmentation compared with other brain extraction methods.

in Dice that can be calculated are lower than 0.01, which shows that WGAN+O-Net has a very significant difference in accuracy with other comparison methods. In terms of sensitivity, ROBEX performed best on IBSR18, followed by BET; our method performed best on LPBA40, followed by BET. In terms of specificity, Kleesiek performed best on IBSR18, followed by U-Net; our method performed best on LPBA40, followed by U-Net. Overall, WGAN+O-Net has the best sensitivity and specificity on LPBA40, but it does not perform well on IBSR18. However, both our method and the second-ranked method have p-values greater than 0.05 on IBSR18, which indicates that there is no significant difference between them.

Compared with similar models, WGAN+O-Net still obtains the highest value of Dice on the two datasets on Dice. WGAN+O-Net is at least 1.26% higher than the second highest WGAN+U-Net on IBSR18, and at least 0.5% higher than the second highest Pix2pix+U-Net on LPBA40. Observing the values in Table 1 shows that all methods perform better

TABLE 2. The mean value of the evaluation results of different algorithms on across-dataset experiment.

method	Dice (%)	Sensitivity (%)	Specificity (%)
Our	95.51	94.61	99.34
Auto U-Net	79.17	65.59	99.98
U-Net	92.25	86.85	99.65
BET	82.44	91.02	94.54
ROBEX	51.45	88.16	69.14

on LPBA40 than on IBSR18. The image quality affects the extraction methods resulting in unstable output. Compared with other methods, our method has better stability on the two data sets.

Figure 5 shows the box plots of the evaluation values of different brain extraction methods in the comparative experiment of the two public datasets. The purpose of the box plots is to intuitively reflect the performance of different algorithms on brain extraction tasks. Since the evaluation values of Kleesiek and HFEM-E are directly quoted from

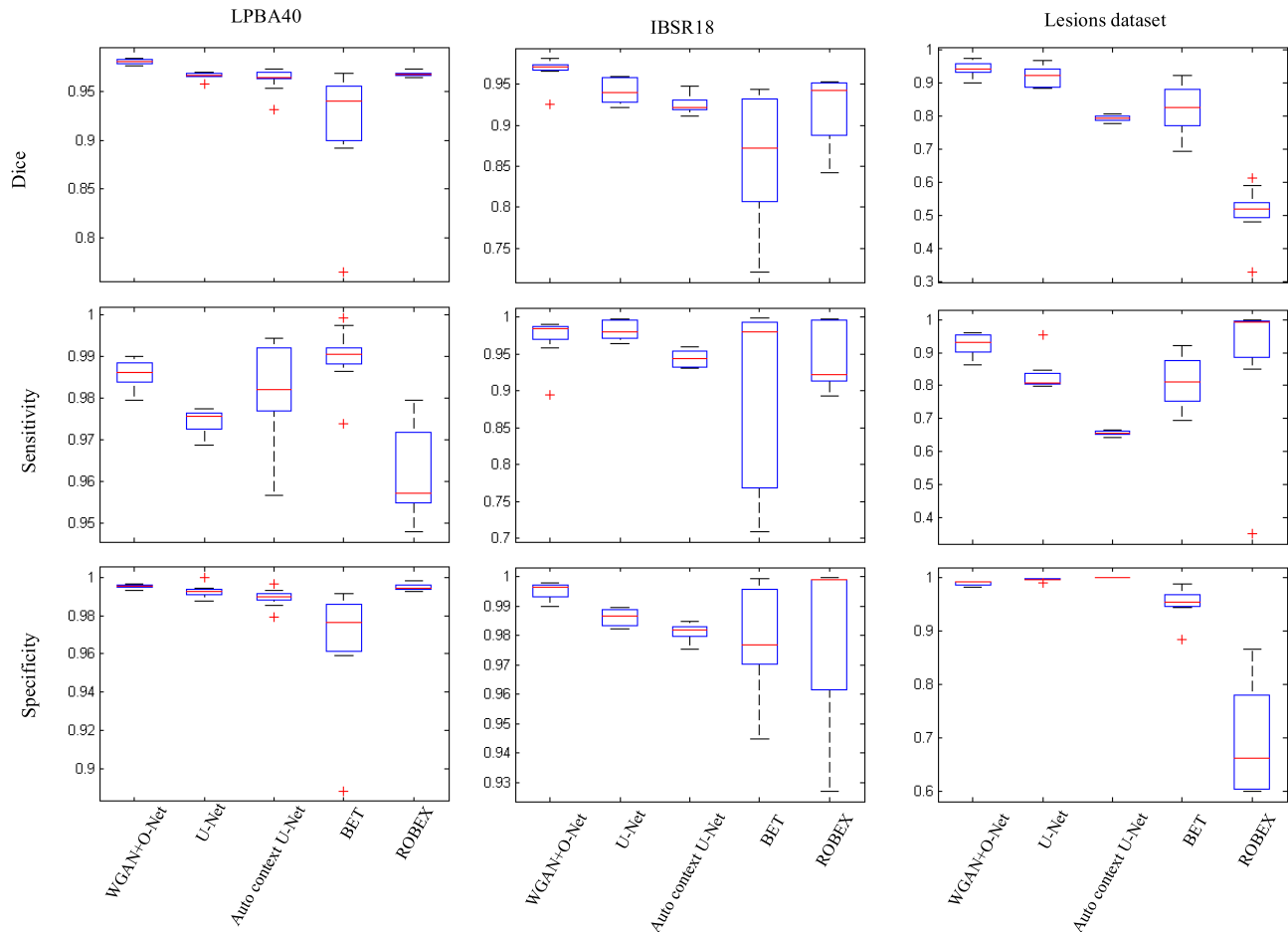


FIGURE 5. Box plots based on evaluation data obtained by different algorithms on three datasets.

the papers, box plots of these two brain extraction methods is not added in Figure 5. In the box plots of WGAN+O-Net on all datasets, every square box maintains a high level; the distribution of normal values is concentrated; most square boxes are symmetrical and have very few outliers. The boxes in the plots of U-Net are also relatively concentrated. Expect for sensitivity, there are few outliers, but their values are not high. For LPBA40, Auto context U-Net has a discrete numerical distribution compared to the other two deep learning models. On the IBSR18, the number of digits in the box plot of each indicator is not as good as that of BET. On LPBA40 and IBSR18, the box plots clearly show that the traditional brain extraction methods (BET and ROBEX) has many abnormal values, heavy tails, and a large deviations from the median. This indicates that their extraction results are usually not as good as those based on deep learning.

The above analysis of comparative experiments shows the superiority of our method in accuracy. Next, we verified the adaptability of the model by cross-dataset experiments. This experiment used healthy sample scans LPBA40 as the training set and diseased sample scans as the test set. The training set and test set with huge differences are selected

in order to explore the adaptability limit of the model as much as possible. Table 2 shows the evaluation results of the predicted masks obtained by different methods in the cross-dataset experiment. Our method achieves the highest Dice coefficient value, which is at least 3.26% higher than the best among other methods. At the same time, the highest sensitivity is obtained, which is at least 3.59% higher than other methods, while the specificity is lower than the highest value of 0.78%. Although Auto Context U-Net method obtained a specificity value of nearly 1, its sensitivity is too low, which means that its brain extraction results are too conservative.

Figure 6 shows the results of experiments across datasets. There is no prediction mask of ROBEX in the figure 5, because it failed to register brain MRI images of lesions, method. The first row in Figure 6 is ground truth (GT). A round and white area can be clearly observed on the coronal and sagittal planes of GT. This area is the focal area of meningioma. which is the main problem of atlas-based brain extraction The gray scale of this lesion area is obviously whiter than that of brain tissue, so it is easy to be recognized by the model as non-brain tissue. It was obvious that only part of the lesions can be identified by U-Net.

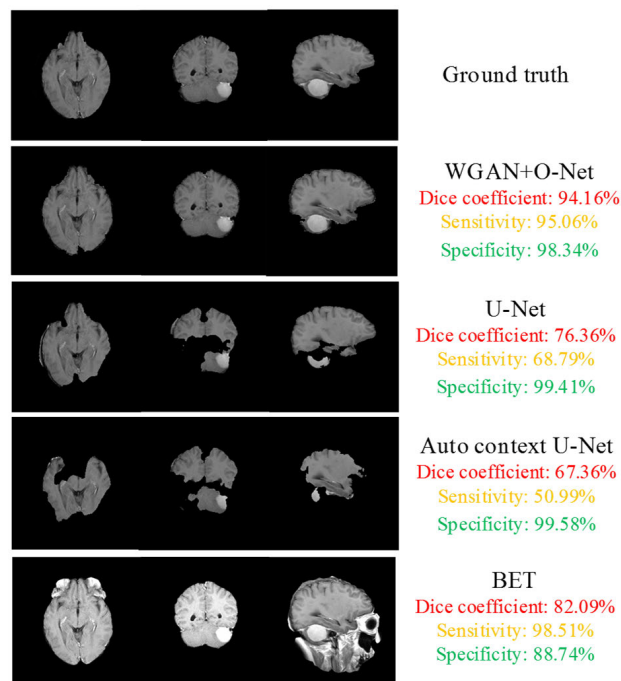


FIGURE 6. Prediction masks of various brain extraction methods in dataset with lesions.

TABLE 3. The mean value of the evaluation results of different algorithms on ablation experiment.

	Dice(%)	Sensitivity(%)	Specificity(%)
Our	97.59	98.25	99.55
Our (L_S)	41.54	27.33	99.35
Our (D)	93.06	92.16	99.11
Our (G)	94.89	98.88	98.96
Our (G, D)	81.25	86.61	99.61

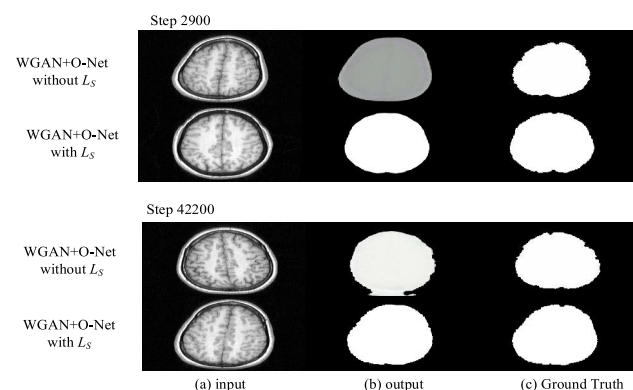


FIGURE 7. The output image of the training process when different objective functions are applied to our model.

The result of Auto Context U-Net that rely on prior information was worse. The extraction results of BET left too many non-brain tissue parts such as eyeballs. Although the result of WGAN+O-Net is not as smooth as ground truth, its segmentation accuracy is the highest. It can completely segment the tumor region with the highest failure probability in the sagittal plane.

TABLE 4. Evaluation results of different network structures on IBSR18.

	Dice (%)	Sensitivity (%)	Specificity (%)
U-Net	94.86 (±0.014)	94.64 (±0.038)	99.20 (±0.006)
Deep-U-Net	91.26 (±0.780)	88.41 (±0.14)	99.39 (±0.006)
WGAN+U-Net	96.00 (±0.015)	95.34 (±0.032)	99.54 (±0.002)
WGAN+	97.30 (±0.007)	98.09 (±0.013)	99.54 (±0.002)
Deep-U-Net	97.26 (±0.006)	97.76 (±0.009)	98.87 (±0.008)
WGAN+O-Net			

The high accuracy and adaptability of the model proposed are proved by the above two experiments. In order to prove that these superior properties contribute to our improvement of the model, we have done ablation experiments. The improvements of each module were removed individually from the model in turn, and then learning process was done on the same dataset. The mean values of the evaluation results of different algorithms on ablation experiment are listed in Table 3. First, the accuracy of the model decreased to 41.54% after the segmentation loss function (L_S) was removed. This result from one side indicates that WGAN + O-Net has most convergence. Figure 7 is a supplementary diagram of Table 3, which records the output results of the segmentation network in the two loss function states at a certain moment in the training process. In the 2900th round of training, WGAN+O-Net with L_S already outputted the approximate outline of the predicted brain mask, while the model without L_S only had a grayscale distribution that was much different from the label. In the 42200th round of training, the model with L_S in the loss function can already output the prediction mask with edge details, while the model without L_S has obvious adhesion. Therefore, it is verified that adding the loss of the segmentation network on the basis of the adversarial loss can greatly improve the convergence speed of the model and reduce the training time. WGAN+O-Net(D) means to replace the residual block in the discrimination network with a common convolutional layer. The model extraction accuracy dropped by 4.03% after the replacement. WGAN+O-Net(G) with U-Net. The accuracy of the model decreased by 2.7%, After the replacement. The above two datasets show that the segmentation network O-Net and the residual block of the discrimination network are effective contributions to the model. The results of ablation experiments show that the improvement of each part of the model has a positive impact on the brain extraction task and makes an effective contribution to the improvement of accuracy.

Compared with the basic model U-Net, the new brain extraction model not only adds a discrimination network, but also increases the depth of the network. Obviously, WGAN+O-Net has more parameters than U-Net. Although the ablation experiments have confirmed that each part of the improvement has a positive contribution to the improvement of brain extraction accuracy, it cannot explained whether the

improvement of accuracy and artifact suppression is due to the improvement of structure or the increase of parameters. To further verify the contribution of model structure to brain extraction research, we evaluated and analyzed the different depth and structure of the model, and the results are shown in Table 4.

In the Table 4, Deep U-Net refers to a model that increases the 4 down-sampling of U-Net to 8 down-sampling. Deep-U-Net adds more feature maps of different scales and some parameters compared to U-Net. Theoretically, Deep-U-Net have better performance. However, Table 4 shows that the Dice coefficient of Deep-U-Net is 91.26%, which is 3.6% lower than of U-Net. Simply increasing the convolutional layer did not achieve performance improvement. When both U-Net and Deep-U-Net were combined with WGAN, the extraction accuracy increased significantly. Note that the discrimination network and loss function of WGAN here refer to the improved one. The Dice coefficient value of WGAN+U-Net is increased by 1.14% compared with U-Net; the Dice coefficient value of WGAN+Deep-U-Net is increased by 6.04% compared with Deep-U-Net. This result reflects that the adversarial training can promote the target segmentation network to obtain better extraction accuracy, alleviate the perform of network degradation, and make the deep network play its due performance.

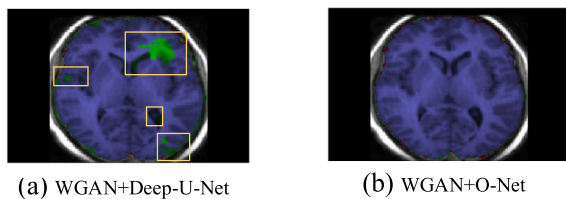


FIGURE 8. Output of WGAN+Deep-U-Net and WGAN+O-Net in the same slice scan. Regions that are difficult to segment are intercepted and zoomed.

In fact, WGAN+Deep-U-Net achieves the highest Dice coefficient value, which is even 0.04% higher than the proposed WGAN+O-Net model in this experiment. However, the results of WGAN+Deep-U-Net on samples with severe artifacts or uneven gray levels show holes, as shown in Figure 8(a). Figure 8(b) is the extraction result of WGAN+O-Net on the same brain MRI slice image. The difference between the WGAN+O-Net model and the WGAN+Deep-U-Net model is whether the dual shortcut path with channel corresponding attention module is used. Although the extraction accuracy of the overall evaluation is slightly worse than that of WGAN+Deep-U-Net, the proposed WGAN+O-Net can better handle low-quality images such as artifacts or uneven gray-scale distribution. This proves that the proposed O-Net can indeed suppress artifacts.

V. DISCUSSION

The performance of these brain extraction methods evaluated in this article is different on the two datasets. ROBEX performed normally on LPBA40 without major error

recognition, but it was affected by artifacts and inconsistent gray-scale distribution. Its extraction results retained large-scale skulls on IBSR18. BET shows strong stability on the two datasets although the extraction results are not good. The boundary of the result extracted by BET is too smooth, and a small part of the brain can easily be misidentified. The sensitivity and specificity evaluation value of the latest method HEFM-E rank low among these methods, indicating that its brain tissue recognition ability and non-brain tissue recognition ability are not good enough. Compared with traditional brain extraction methods, CNN-based methods cause no serious error. U-Net is able to achieve good extraction results on brain MRI datasets with good image quality, such as LPBA40, but it can cause unstable results for datasets with uneven quality such as IBSR18. U-Net largely depends on image quality. When it processes images with artifacts or partial gray distributions that are inconsistent with most other images, the extraction results will show large-scale irregular missing areas. Like U-Net, the results obtained by auto context U-Net also depend on the quality of the dataset. This brain extraction method trends to exclude some brain tissues (see Figure 5). The reason for this may be that the presence of artifacts will affect the effect of the automatic context architecture.

Although based on U-Net, the stability of the proposed WGAN+O-Net is better than U-Net. WGAN+O-Net can achieve a further improvement in accuracy on a good-quality data set (LPBA40), and can also process images with artifacts or uneven grayscale on a poor-quality dataset IBSR18 (see Figure 3 and Figure 8). WGAN+O-Net achieved the highest accuracy on both public datasets, and its results are significantly different from other comparison methods. The adaptability of WGAN+O-Net has also been verified by cross-data set experiments with very large differences between the training set (LPBA40) and the test set (brain data set with lesions). In the test scan images, the lesion may squeeze the brain tissue and blur the brain boundary, which makes it easy for the model that has not learned the characteristics of the lesion area to make wrong judgment. WGAN+O-Net can locate the brain tissue by increasing the focus of the target region so as to improve the model's sensitivity of the model to foreground pixels, and more scale participation makes it have a more comprehensive receptive field. At the same time, the model also has a strong fitting ability to generate a confrontation network. Therefore, high-precision extraction (average 95.51%) can still be maintained for experiments with huge differences across datasets. WGAN+O-Net has a high rate of correct recognition of the lesion area (see Figure 4).

Although WGAN+O-Net has the above advantages, it also has some disadvantages. One is that the method cannot yet achieve cross-modal extraction, that is, the model trained on the T1 weighted MRI images cannot be applied to the T2 weighted images. The other is that we only studied on 2D brain MRI slice images. Thus, the future work will focus on how to realize brain extraction on multi-modal MRI images

based on GAN and expand the 2-D based brain extraction method to 3-D based extraction method.

VI. CONCLUSION

In this study, we presented a new brain extraction model WGAN+O-Net which has the ability to suppress artifacts to prevents the large-scale irregular regions in the prediction of brain mask. In comparison to existing brain extraction methods, WGAN+O-Net has a more accurate prediction of brain mask. Moreover, this method can still maintains a good extraction results on data with large differences.

REFERENCES

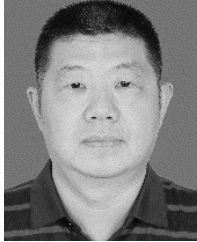
- [1] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [2] X. F. Wang, S. D. Nie, and Y. J. Wang, "Improved brain extraction tool algorithm for brain parenchyma segmentation," *J. Med. Phys.*, 2016.
- [3] J. Qiu and W. Cheng, "Brain tissues extraction based on improved brain extraction tool algorithm," in *Proc. IEEE Int. Conf. Comput. Commun.*, Oct. 2016, pp. 553–556.
- [4] Y. Luo, B. Gao, Y. Deng, X. Zhu, T. Jiang, X. Zhao, and Z. Yang, "Automated brain extraction and immersive exploration of its layers in virtual reality for the rhesus macaque MRI data sets," *Comput. Animation Virtual Worlds*, vol. 30, no. 1, p. e1841, Jan. 2019.
- [5] J. Valente, P. M. Vieira, C. Couto, and C. S. Lima, "Brain extraction in partial volumes T2*7T by using a quasi-anatomic segmentation with bias field correction," *J. Neurosci. Methods*, vol. 295, pp. 129–138, Feb. 2018.
- [6] H. Chang, W. Huang, C. Wu, S. Huang, C. Guan, S. Sekar, K. K. Bhakoo, and Y. Duan, "A new variational method for bias correction and its applications to rodent brain extraction," *IEEE Trans. Med. Imag.*, vol. 36, no. 3, pp. 721–733, Mar. 2017.
- [7] S. Jiang, Y. Wang, X. Zhou, Z. Chen, and S. Yang, "Brain extraction using active contour neighborhood-based graph cuts model," *Symmetry*, vol. 12, no. 4, pp. 1–14, 2020.
- [8] J. E. Iglesias, C. Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1617–1634, Sep. 2011.
- [9] X. Han, R. Kwitt, S. Aylward, S. Bakas, B. Menze, A. Asturias, P. Vespa, J. Van Horn, and M. Niethammer, "Brain extraction from normal and pathological images: A joint PCA/image-reconstruction approach," *NeuroImage*, vol. 176, pp. 431–445, Aug. 2018.
- [10] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [11] T. X. Pham, P. Siarry, and H. Oulhadj, "Segmentation of MR brain images through hidden Markov random field and hybrid metaheuristic algorithm," *IEEE Trans. Image Process.*, vol. 29, pp. 6507–6522, 2020.
- [12] F. Shi, L. Wang, Y. Dai, J. H. Gilmore, W. Lin, and D. Shen, "LABEL: Pediatric brain extraction using learning-based meta-algorithm," *NeuroImage*, vol. 62, no. 3, pp. 1975–1986, 2012.
- [13] J. Liu and L. Guo, "A new brain MRI image segmentation strategy based on K-means clustering and SVM," in *Proc. 7th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, vol. 2, Aug. 2015, pp. 270–273.
- [14] G. Zhao, F. Liu, J. A. Oler, M. E. Meyerand, N. H. Kalin, and R. M. Birn, "Bayesian convolutional neural network based MRI brain extraction on nonhuman primates," *NeuroImage*, vol. 175, pp. 32–44, Jul. 2018.
- [15] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [17] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2319–2330, Nov. 2017.
- [18] H. Hwang, H. Z. U. Rehman, and S. Lee, "3D U-Net for skull stripping in brain MRI," *Appl. Sci.*, vol. 9, no. 3, p. 569, 2019.
- [19] K. Chen, D. Zhu, J. Lu, and Y. Luo, "An adversarial and densely dilated network for connectomes segmentation," *Symmetry*, vol. 10, no. 10, p. 467, Oct. 2018.
- [20] P. Moeskops, M. Veta, M. W. Lafarge, K. A. Eppenhof, and J. P. Pluim, "Adversarial training and dilated convolutions for brain MRI segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, Quebec, QC, Canada, 2017, pp. 56–64.
- [21] T. E. and K. Saruladha, "Design of FCSE-GAN for dissection of brain tumour in MRI," in *Proc. Int. Conf. Smart Technol. Comput., Electr. Electron. (ICSTCEE)*, Oct. 2020, pp. 1–6.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [23] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: <http://arxiv.org/abs/1701.07875>
- [24] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [25] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. NIPS*, vol. 28, 2015, pp. 2017–2025.
- [26] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6688–6697.
- [27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [28] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks,'" Tech. Rep.
- [29] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
- [30] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," 2021, *arXiv:2102.00240*. [Online]. Available: <http://arxiv.org/abs/2102.00240>
- [31] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 1–20.
- [32] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: <http://arxiv.org/abs/1802.05957>
- [33] J. Wu, Z. Huang, J. Thoma, D. Acharya, and L. Van Gool, "Wasserstein divergence for GANs," in *Proc. Eur. Conf. Comput. Vis.*, vol. 11209, 2018, pp. 653–668.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Oct. 2016, pp. 630–645.
- [35] K. Ezhilarasan, S. Praveenkumar, K. Somasundaram, T. Kalaiselvi, S. Magesh, S. Kiruthika, and A. Jeevarekha, "Automatic brain extraction from MRI of human head scans using Helmholtz free energy principle and morphological operations," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102270.



SHAOFENG JIANG was born in Yingtan, Jiangxi, China, in 1978. He received the B.S. degree in electronic engineering from Beihang University, Beijing, China, in 1999, the M.S. degree in measurement technology and instrumentation from Nanchang Hangkong University, Nanchang, China, in 2005, and the Ph.D. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2008. Since 2009, he has been an Assistant Professor with the Department of Biomedical Engineering, Nanchang Hangkong University. His research interests include medical image processing, pattern recognition, and artificial intelligence.

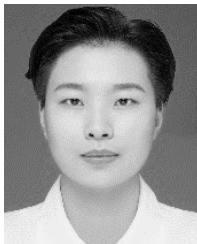


LANTING GUO received the bachelor's degree in biomedical engineering from Nanchang Hangkong University, China, in 2018, where she is currently pursuing the master's degree in instrumentation engineering. Her current research interests include image processing and computer vision.

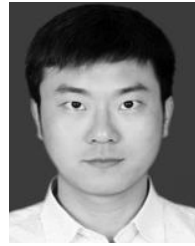


GUANGBIN CHENG was born in Shihezi, Xinjiang, China, in 1969. He received the B.S. degree in automatic control engineering from Northeastern University, Shenyang, China, in 1991, the M.S. degree in biomedical engineering from China Medical University, Shenyang, in 1998, and the Ph.D. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2008.

Since 2009, he has been a Senior Engineer with the Department of Medical Engineering, Guangdong Armed Police Corps Hospital, Guangzhou. His research interests include medical image processing, pattern recognition, artificial intelligence, and clinical medical equipment.



XINGYAN CHEN received the bachelor's degree in biomedical engineering from Nanchang Hangkong University, China, in 2019, where she is currently pursuing the master's degree in instrumentation engineering. Her current research interests include image processing and computer vision.



CONGXUAN ZHANG (Member, IEEE) received the Ph.D. degree in measurement technology and instruments from Nanjing University of Aeronautics and Astronautics, Nanjing, in 2014. From 2018 to 2019, he was a Visiting Scholar with the Department of Biomedical Engineering, University of Kansas. He is currently an Assistant Professor with the School of Measuring and Optical Engineering, Nanchang Hangkong University, China. His current research interests include image processing and computer vision.



ZHEN CHEN received the Ph.D. degree in mechanical design and theory from Northwestern Polytechnical University, Xi'an, in 2003. From 2006 to 2007, he was a Visiting Scholar with the Department of Biomedical Engineering, University of Kansas. He is currently a Professor with the School of Measuring and Optical Engineering, Nanchang Hangkong University, China. His current research interests include image understanding and measurement.

...