IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 17, 2019, accepted July 14, 2019, date of publication July 16, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929270

Liver Semantic Segmentation Algorithm Based on Improved Deep Adversarial Networks in Combination of Weighted Loss Function on Abdominal CT Images

KAIJIAN XIA[1,2,3], HONGSHENG YIN[1], PENGJIANG QIAN[3,4], (Member, IEEE), YIZHANG JIANG[3,4], (Member, IEEE), AND SHUIHUA WANG[5]

[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]Changshu Affiliated Hospital of Soochow University, Changshu No.1 People's Hospital, Changshu 215500, China
[3]Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi 214122, China
[4]School of Digital Media, Jiangnan University, Wuxi 214122, China
[5]School of Architecture Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, U.K.

Corresponding author: Kaijian Xia (xiakaijian@163.com)

This work was supported in part by the Jiangsu Committee of Health on the Subject under Grant H2018071, in part by the National Natural Science Foundation of China under Grant 61702225 and Grant 61772241, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160187, in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX- 127, in part by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002, and in part by the Open Fund Project of Jiangsu Key Laboratory of Media Design and Software Technology (Jiangnan University) under Grant 19ST0205.

ABSTRACT Due to the space inconsistency between benchmark image and segmentation result in many existing semantic segmentation algorithms for abdominal CT images, an improved model based on the basic framework of DeepLab-v3 is proposed, and Pix2pix network is introduced as the generation adversarial model. Our proposed model realizes the segmentation framework combining deep feature with multi-scale semantic feature. In order to improve the generalization ability and training accuracy of the model, this paper proposes a combination of the traditional multi-classification cross-entropy loss function with the content loss function of generator output and the adversarial loss function of discriminator output. A large number of qualitative and quantitative experimental results show that the performance of our proposed semantic segmentation algorithm is better than the existing algorithms, and can improve the segmentation efficiency while ensuring the space consistency of the semantics segmentation for abdominal CT images.

INDEX TERMS Semantic segmentation, generation adversarial networks, weighted loss function, multi-scale features, game adversarial, atrous space pyramid pooling.

I. INTRODUCTION

Primary liver cancer, especially hepatocellular carcinoma, is one of the common malignant tumors, and is one of the leading causes of cancer death in the world. According to the statistics of the 2015 World Health Organization[1], liver cancer has become the second disease in global cancer mortality. The prevention and treatment of liver diseases are imminent and have become a hot spot and focus of the world. In the primary liver cancer, the texture is hard, the edges are irregular, and the surface irregularities are characterized by large or small nodules. Therefore, the detection of liver lesions provides an important basis for the subsequent clinical and treatment planning, and there are more and more requirements for detection and diagnosis in the clinic [2].

The rapid development of medical imaging technology provides a new means for the identification of primary liver cancer. Doctors can observe the signs of the lesion from the image, analyze and diagnose it. However, the existing medical imaging technology for liver examination rely heavily on the experience and technology of the operator, and often has the disadvantages of strong subjectivity, low reproducibility, high labor intensity, and low efficiency [3]. Therefore, the assisted detection technology is adopted for primary liver cancer. It is of great significance in clinical applications. Intelligent medical image processing needs to accurately

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

locate the spatial location, size and other states of the lesion and its corresponding relationship with surrounding tissues, assist the medical staff to qualitatively and even accurately quantify the diseased tissues and organs, and then the liver status and treatment plan have a more accurate judgment. However, most of the existing medical image-based processing algorithms are not universal, and the treatment effects for different human organs are quite different [3]. Therefore, the establishment of a robust, objective, repeatable, efficient and high-accuracy method for detecting liver lesions has important clinical significance for the prevention and treatment of liver diseases. In recent years, domestic and foreign studies have begun to actively explore the research methods of liver damage imaging and achieved some stage results.

Semantic segmentation is one of the challenging research topics in the field of computer vision. Its purpose is to assign a label to each pixel in the image, and to decompose an overall scene into several separate entities, which helps to infer the different behaviors of the target and finally solve the higher levels of visual problems, including autonomous driving, augmented reality, etc. With the development of artificial intelligence technology, semantic segmentation has attracted more and more scholars' attention and proposed a series of effective solutions. However, the semantic segmentation model is still very challenging to obtain accurate segmentation in target localization and segmentation, mainly due to complex background, multi-scale variation, boundary blur for abdominal CT images.

In recent years, deep learning algorithms represented by convolutional neural networks (CNN) have achieved significant performance improvements in image semantic segmentation, but these methods always suffer from spatial inconsistencies between the benchmark template and the segmentation results, which are partially attributed to the random error generated by the independent prediction process of the tag variable. Therefore, scholars have proposed a number of post-processing methods to enhance spatial consistency in predictive label maps, refine segmented label masks and eliminate significant boundary errors for liver segmentation. Chen *et al.* added a fully connected conditional random field based on deep semantic pixel classification to enhance the spatial consistency of the segmentation result [4]. The DeepLab method introduced a fully connected conditional random field in the last layer of the deep network, and combined the response results at different scales to enhance the performance of target positioning. This method can be widely used for information combination of high-level deep features and low-level local features [5]. Despite of these improvements in the use of post-processing methods, the DeepLab model is still limited to the use of point-pair random field model to fuse feature information with a priori information. In [7], Luc *et al.* proposed a generative adversarial networks method to train the segmentation model. The average performance of the method is good, but the semantic segmentation performance is general in some specific scenarios.

Due to the multi-scale characteristics of the object image and the low resolution of the feature map, the deep network mainly uses the maximum pooling and down-sampling methods to obtain feature invariance, which leads to loss of positioning accuracy [6]. In general, a deep network is connected to a convolutional layer with several fully connected layers. The purpose is to map the feature map generated by the convolutional layer into a fixed length feature vector [8]. Since the network output is a probability-based feature map, the network is suitable for pixel-level binary classification and regression tasks [9]. However, for the semantic multi-objective segmentation model, the fully concatenated convolution network may input an image of any size, compress it by layer-by-layer convolution, and then up-sample the final feature image with the deconvolution layer to reconstruct it to the same size of the input image. This makes the final saliency map with deviation. The general solution is to preserve the original spatial information while generating a prediction probability for each pixel, and finally achieve pixel-by-pixel classification, but this increases the spatial and temporal complexity of the network [10]. In [11], Wang *et al.* proposed a novel feature transformation network that connected convolutional networks with deconvolution networks to enhance the representation of shared features so as to recover spatial information from low-resolution feature maps. The DeepLab-V3 model proposed in [12] introduced the idea of atrous convolution, and realized the exponential expansion of the receptive field on the basis of ensuring the resolution of the convolution feature. At the same time, under the framework of cascading module and space pyramid pooling, multi-scale semantic information was extracted to improve the segmentation effect [12]. However, the DeepLab V3 network not only removed the fully-connected module in the last layer to obtain accurate local spatial consistency information, but also eliminated global information, resulting in incomplete segmentation results.

Since the classical spatial pyramid scale structure can process images with any size and scale, it not only improves the accuracy of classification, but also improves the detection efficiency. In order to obtain more scales of irregular object information, different levels of features can be cascaded through the atrous space pyramid pooling (ASPP), and the understanding ability of image is enhanced by blending local and global semantic features. Based on the basic framework of DeepLab v3, this paper introduced Pix2pix network as the Generative Adversarial Networks model, and realized the segmentation architecture based on deep features and multi-scale semantic features. The architecture used a generator, two discriminators and a semantic network to correct semantic segmentation results and reduce spatial structure inconsistency. In order to increase the generalization ability and training precision of the model, this paper proposed to combine the traditional multi-class cross entropy loss function with the content loss function of the generator output and the adversarial-loss function of the discriminator output to construct a weighted loss function. The results of

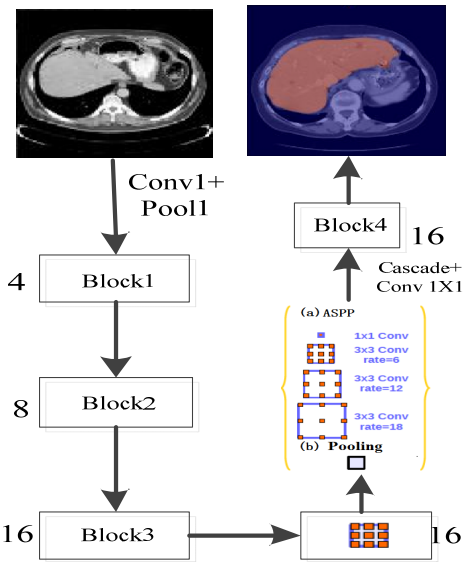


FIGURE 1. ASSP-based DeepLab V3 architecture.

qualitative and quantitative experiments showed that the performance of the semantic segmentation algorithm proposed in this paper exceeds the existing segmentation algorithm, and the segmentation efficiency is improved while ensuring the consistency of semantic segmentation. The rest of this paper is organized as follows: the related work is presented in section II. The section III is dedicated to the introduction of the main contributes of the study about how to improve the efficiency of the improved adversarial network, and the improved algorithm is developed. The experimental results are presented in section IV. The paper ends up with conclusions and perspectives.

II. RELATED WORKS

Semantic segmentation is a very active research direction in the field of computer vision, and many excellent algorithms have been proposed. Before the arrival of the deep network, the semantic segmentation method relied mainly on artificially designed features to classify pixels independently. Specifically, the features of the image segmentation are sent to a well-designed classifier, such as SVM, random forest and Boosting, to predict the category of the central pixel of the image block. With the rapid development of convolutional neural networks in image classification, the deep feature extracted by deep model for semantic segmentation can improve the accuracy of segmentation. Most of the existing algorithms use deep network training on a block-by-block basis to achieve accurate boundary prediction by super-pixel refinement of random fields or local classifiers.

The DeepLab V3 network is currently a better semantic segmentation algorithm. The model architecture is shown in Figure 1. The deep module of the structure is generally improved by the VGG-16 or ResNet-1 architecture. This model replaces all the fully connected layers in DeepLab with convolutional layers, and reduces the resolution of the feature map to 1/8 of the original image by atrous convolution; then the feature map is enlarged by 8 times through the bi-linear interpolation algorithm so as to reconstruct the original size. Finally, the final feature map is input into the fully-connected condition random field to be refined so as to further improve the segmentation result. It can be seen that under the framework of the cascade module and the space pyramid, the atrous convolution module increases the receptive field of the filter to fuse the multi-scale semantic information, and the network also introduces convolution modules with different learning rates to enhance the feature representation capability.

III. OUR PROPOSED SEMANTIC FRAMEWORK

It is well known the DeepLab V3 network not only removes the fully-connected module in the last layer to obtain local spatial consistency information, but also eliminates global information, resulting in incomplete segmentation results. Therefore, according to the basic framework of DeepLab-v3, this paper introduces Pix2pix network as the Generative Adversarial Networks (GAN) model, and realizes the segmentation architecture combining deep features and multi-scale semantic features. The basic framework is shown in Figure 2. The model proposed in this paper consists of three modules: (1) basic semantic segmentation model, which is mainly constructed by DeepLabv3 network; (2) generator for reconstructing the generated image from the training samples; and (3) discriminator for identifying the generated image and real image. The generator and discriminator form a generative adversarial network that is first pre-trained using the reference mask and its original image. In the pre-training phase, the reference mask is used as an input, and the generator is driven to produce a reconstructed image that is difficult to distinguish from the real image. Then, a pre-trained generator and discriminator are used to characterize the change of the loss function during training. Therefore, the semantic segmentation framework proposed in this paper uses WGAN as a loss function to optimize the basic semantic segmentation network in an adversarial manner.

A. IMPROVED GENERATIVE ADVERSARIAL NETWORKS MODEL

The Generative Adversarial Networks (GAN) consists mainly of two modules, that is, the generator and the discriminator. The generator is mainly used to learn the real image distribution so that the image generated by it is more realistic, which can fool the discriminator. While the discriminator then judges if the input image is fake. The whole learning process is that the generator generates a more realistic image, and the discriminator accurately recognizes the real and fake image. This process is equivalent to a two-person game, and a dynamic equilibrium is finally achieved through continuous confrontation. That is to say, the generator gets close to the real image distribution, and the discriminator cannot recognize the real and fake images. This training process for GAN

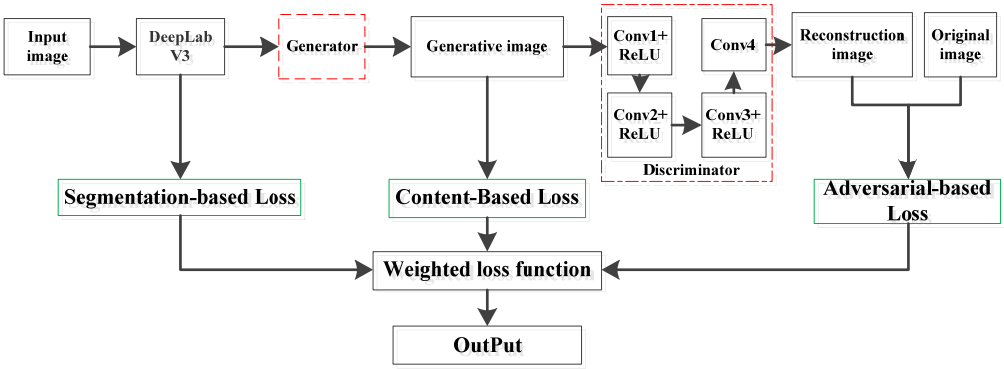


FIGURE 2. Our proposed semantic framework.

can be defined as follows:

$$\min_{G} \max_{D} E_{I\sim P_{data}(I)}[\log D(I)] + E_{I^R\sim P_g(I)}[\log(1 - D(I^R)] \tag{1}$$

The generator simultaneously takes the output of the prediction layer p_{seg} and the original image I as inputs, and generates an image similar to the original image. Inspired by the Pix2pix network, the generator consists of four convolutional layers and four deconvolution layers, followed by a dropout layer to prevent network over-fitting. During training, it is only necessary to randomly sample the parameters of the weight layer according to a certain probability $p = 0.5$, and use the corresponding sub-network as the object network for the current update. The discriminator network is composed of 4 convolution layers, each layer has a ReLU behind it as an activation function.

As described above, the generator is constructed as a codec network with convolution and deconvolution layers, and the sampling scale is gradually reduced by a series of codecs until the bottleneck layer is reached, and then the process is reverse reconstructed to the original or first layer. In order to avoid information loss in the codec process, a skipped-connection is added between the *i*-th layer and the *(n-i)*-th layer, where *n* represents the total number of layers in the generator. Each skipped-connection simply cascades the output features of the *i*-th layer with the features of the *(n-i)*-th layer.

As we all know, the l_1 loss function and the l_2 loss function may make the generated image relatively fuzzy. Although these loss functions do not achieve clear high frequency characteristics, the low frequency characteristics of the image are still accurately captured in many cases. Since the l_1 loss function enforces low frequency constraints, the model only needs to simulate the high frequency features of the local patches of the image. Since GAN only processes low-frequency components, it does not need to process the entire image. Therefore, this paper selects PatchGAN as the Markov discriminator, and judges the patch with the size $N \times N$ in the image. All responses to the discriminator are averaged to provide a final output.

B. SEGMENTATION MODEL

As we all know, the goal of the semantic segmentation model is to generate a confidence map $p_{seg} \in \mathrm{R}^{c*w*h}$, in which c is the dataset class number, and w, h is the width and height of the predicted saliency map respectively. Then, the prediction result is proposed by max pooling and the deep model is constructed through the fully connected network so that images of different sizes can be processed.

Therefore, DeepLab V3 is selected as the basic segmentation model, the purpose of which is to generate a confidence map p_{seg}, and use *argmax* operation to obtain the final prediction mask, where each value indicates the label response of the input pixel.

C. WEIGHTED LOSS FUNCTION

The semantic segmentation framework of this paper adopts a hybrid loss function l, which is mainly composed of three parts: segmentation-based loss l_s, content-based loss l_c and adversarial-based loss l_a. The mixed loss function can express the following equation:

$$l = l_s + \lambda_1 l_c + \lambda_2 l_a \tag{2}$$

where λ_1 and λ_2 are two empirical weight parameters. In this paper, multi-class cross entropy loss is used to evaluate semantic segmentation performance. The loss term is defined as follows:

$$l_s = -\frac{1}{M}\sum_{j=1}^{M}\sum_{x}^{N}\sum_{i}^{C} Y_{xi}^{(j)} \log(P_{xi}^{(j)}) \tag{3}$$

where P_{xi} is calculated by the segmentation model, indicating the probability of assigning a label i to a pixel x; Y_{xi} is the probability of label for a reference template; M, N, C are denoted as the number of samples, the total number of pixels, and the number of class on data set respectively.

The content-based loss function is used to calculate the quality of the reconstructed image I^R generated by the generator network. Therefore, the loss function is calculated pixel by pixel as follows:

$$l_c = L_{l_1}(G) = E_{Y,I\sim P_{data}(Y,I),z\sim P_z(z)} \|I - G(Y, z)\|_1 \tag{4}$$

Adversarial-based loss function reflects the image quality reconstructed by the generator. This paper uses the loss term of the Wasserstein GAN network, and its equation is written as follows:

$$l_a = -E_{Y\sim P_{data}(Y,I),Z\sim P_z(Z)}D(G(Y,\mathrm{Z})) \tag{5}$$

The objective function of Adversarial network can be expressed as:

$$\begin{aligned}\mathrm{L}_{cG}(G,D) = {} & E_{Y,z\sim P_{data}(Y,z)}(Y,z)[D(Y,z)] \\ & - E_{Y\sim P_{data}(Y,z),I\sim P_z(I)}[1-D(Y,G(Y,I))]\end{aligned} \tag{6}$$

where G is the loss term that maximizes the generative adversarial map, namely, $\mathrm{G}^* = \arg\min_G \max_D L_{cG}(G,D)$. The experimental results show that the GAN target loss function and the traditional loss function l_1 work together to generate images, which can improve the accuracy of learning. This shows that in the case where the discriminator's objective function is unchanged, the generator not only needs to deceive the discriminator, but also the absolute value error loss function is close to the true reference value. Therefore, the final loss function term of our proposed model in this paper can be expressed as the following equation:

$$\mathrm{G}^* = \arg\min_G \max_D L_{cG}(G,D) + \lambda L_{l1}(G) \tag{7}$$

If there is no input signal z, the network can still learn the mapping from x to y, but it will produce an uncertainty output, so it can't match any distribution except the δ function. The generator simply ignores noise, so this strategy does not enhance the model's generalization capabilities. For the final semantic model, its dropout layer can be consider as noise. Despite there is the dropout noise, the results show that the output of the model has only a small randomness, which can capture the complete entropy of the conditional distribution and achieve accurate semantic segmentation.

D. TRAINING PROCESS

The semantic segmentation framework proposed in this paper is optimized by the loss function in equation (2). The forward propagation training process is as follows: First, the GAN model is trained to learn the mapping relationship between the benchmark map and the original image. The benchmark mask map I_i^{GT} is the input of the generator network G to obtain the generated image I^R, and the corresponding expression is described as follows:

$$I_i^R = f(I_i^{GT};\pi) \tag{8}$$

where, π represents the parameters of the generator G; the assignment probability of the discriminator D is as follows:

$$p = D(I_i, I_i^R, \varphi) \tag{9}$$

where, I_i is the original image and φ is the parameter of the discriminator D. The learning process of the deep network is to iteratively optimize the parameters φ and π through the loss function. The network parameters are initialized and forward propagation is used to obtain the loss value loss(l_c, l_a) for each time. In each iteration, select a small portion of the image from the training set to learn, and then update each parameter:

$$\pi \leftarrow \pi - \tau\nabla_\pi(l_c + l_a) \tag{10}$$

$$\varphi \leftarrow \varphi + \tau\nabla_\pi l_a \tag{11}$$

where, τ represents the learning rate of the training process.

After the GAN model completes its training, the segmentation model can be trained using the proposed framework. First, the parameters of the GAN model are initialized from a pre-trained weight file. Then, the mask image I_i^M is obtained from the segmentation model,

$$I_i^M = \phi(I_i;\theta) \tag{12}$$

where, θ is the parameter of the segmentation model. In the semantic segmentation step, the parameters of the GAN model are fixed and will not be updated. We use Back Propagation (BP) for learning and stochastic gradient descent (SGD) to learn and optimize the parameter θ so as to minimize the loss function $loss(l_a; l_c; l_s)$. In addition, the forward propagation is adopted to obtain the loss value of each iteration. The update strategy of the parameter θ is consistent with the parameter, and its update equation is denoted as follows:

$$\theta \leftarrow \theta - \tau\nabla_\theta(l_a + l_c + l_s) \tag{13}$$

IV. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

A. EXPERIMENTAL DATA SET

To evaluate the effectiveness of the proposed semantic segmentation model, the self-built data-set and the public data-set are used for deep learning training and testing. The self-built data-set collects 7000 medical abdominal image data from several Third-Class A Hospitals in Jiangsu Province and some data have been labeled by imaging specialists. The public data set is from the medical imaging data set DeepLesion released by NIHCC, which contains more than 32,000 lesion annotations from more than 10,000 cases. In addition, we also used Liver Tumor Segmentation (LiTS) challenge data set. The LiTS data-set consists of 131 contrast-enhanced abdominal CT scans from various clinical sites around the world. The challenge provides reference annotations for the liver contours as well as for liver lesions. To facilitate training and testing, 12,150 images are selected as training samples and 8,800 images as testing samples.

B. PARAMETER SETTING AND EVALUATION CRITERIA

Since the scans vary in in-plane resolution and slice thickness, we aligned the directions of all images, but kept the different resolutions, so that the networks are able to process a range of resolutions and the results can be compared to the original labels. The image used in this paper is preprocessed and re-sized to 320 × 240 for training. The networks selected in this paper are all based on the TensorFlow framework, and their parameters are consistent with the literature [20].

The loss function λ_1, λ_2 are set to 0.15 and 0.1 respectively; the learning rate is initialized to 0.15, and then the learning rate is changed to 0.015 when training to the 50th Epoch; if 100 Epoch are reached, the loss function proposed in this paper does not change and then stops training. In practical applications, some parameters are usually done by cross-validation. Limited to GPU memory, the batch size is set to 8. Each network trained for 50 epochs, which equaled 60000 iterations. The weight is attenuated to 0.0005 and the probability of the Dropout layer is set to 0.5. For generator training, an Adam optimization algorithm with isotropic Gaussian weights is used. The experimental environment of this paper is: Xeon (Xeon) E7-8890 v2 @ 2.80GHz (X4), 128 GB (DDR3 1600MHz), Nvidia GeForce GTX 1080 Ti, Ubuntul6.04, 64-bit operating system.

According to the evaluation of 2017 LiTS challenge [31], we employed the mean values for Dice score, as well as Jaccard and volume overlap error (VOE), relative volume difference (RVD), average symmetric surface distance (ASSD), and maximum symmetric surface distance (MSSD) [32] to evaluate the liver and tumor segmentation performance respectively When applied to a binary segmentation task. Dice per case score refers to an average Dice score per volume while Dice global score is the Dice score evaluated by combining all datasets into one, which can evaluate the degree of overlap between the predicted segmentation mask and the reference segmentation mask. Given binary masks A and B, the Dice score can be described as:

$$\mathrm{Dice}(A, B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{14}$$

Its interval is [0,1] and a perfect segmentation yields a Dice score of 1. In order to evaluate the accuracy of image semantic segmentation results, this paper also uses Pixel Accuracy (PA), Mean Accuracy (MA) and Mean Intersection over Union (MIoU) as the evaluation criteria, whose formulas are written as follows:

$$\mathrm{PA} = \sum_i n_{ii} / \sum_i t_i \tag{15}$$

$$\mathrm{MA} = (1/n_{cl}) \sum_i n_{ii}/t_i \tag{16}$$

$$\mathrm{IoU} = (1/n_{cl}) \sum_i n_{ii}/(t_i + \sum_j n_{ji} - n_{ii}) \tag{17}$$

where, n_{ij} is the number of pixels in which class i is correctly classified as class j and t_i is the number of samples in class i. As can be seen from the IoU definition, this is equivalent to the result of dividing the overlap of the two regions by the set of the two regions. In general, a score greater than 0.5 can be considered to correctly detect and segment the object.

C. QUALITATIVE AND QUANTITATIVE RESULTS ANALYSIS

In order to qualitatively and quantitatively analyze the performance of the proposed semantic segmentation algorithm, the comparison algorithms selected in this paper are Deeplabv3 [12], DeconvNet [17], and SegNet [5].

TABLE 1. Average of the evaluating indicators UNDER LITS challenge data set.

Models	Dice	VOE	RVD	ASSD	MSSD
Deeplabv3	0.945	0.152	0.021	*3.028*	*41.598*
SegNet	0.938	0.115	0.015	4.860	48.305
DeconvNet	0.921	0.110	0.024	3.487	42.110
Proposed	**0.970**	**0.079**	**0.006**	**1.925**	**35.583**

1) ANALYSIS OF QUALITATIVE AND QUANTITATIVE RESULTS FOR LITS CHALLENGE DATA SET

In recent years, liver segmentation has been a subject of research in the medical image processing community due to the availability of the Liver Tumor Segmentation (LiTS) challenge data set. All top scoring automatic segmentation methods in the LiTS challenge used CNNs. In order to verify the performance of the proposed algorithm, the liver segmentation experiment was performed using the Liver Tumor Segmentation Challenge (LiTS) dataset.

Table I shows the results of different semantic models on Benchmark LiTS data set, where bold and *italics* are the best and second best results, respectively.

The Deeplabv3 model uses an improved 2D-FCN network for liver segmentation, which adjusts the receptive field of filter and controls the characteristic response resolution calculated by the convolutional neural network. It can be seen from Table I that the segmentation accuracy is only lower than the proposed algorithm. The reason is that the 2D FCN network cannot utilize the spatial information in the CT image, and our model uses the weighted adversarial-loss function. The DeconvNet method is also segmented by 2D FCN network, and the network structure and feature utilization are improved. The segmentation Dice coefficient reaches 0.921. The core idea of the DesNet model is to create a cross-layer connection to connect the front and back layers of the network, so that each layer in the network accepts the characteristics of all the layers as input. Since a large number of features are multiplexed, a large number of features can be generated using a small number of convolution kernels, and the size of the final model is also small. The algorithm has a segmentation Dice coefficient of 0.938. The proposed semantic segmentation algorithm has a Dice coefficient of 0.970 on the test data of 70 cases of LiTS which is higher than Deeplabv3, DeconvNet, and SegNet with a 5-fold cross-validation. The accuracy and loss change during training is shown in Fig.3 and 4. The experimental results show that the improved GAN segmentation algorithm based on weighted loss function can accurately segment organs. Our results differ greatly from the best results in ISBI 2017, which does not mean that our algorithm is better than all algorithms. The main reason is that we choose partial test data to test different comparison algorithms so as to illustrate the effectiveness of our algorithm.

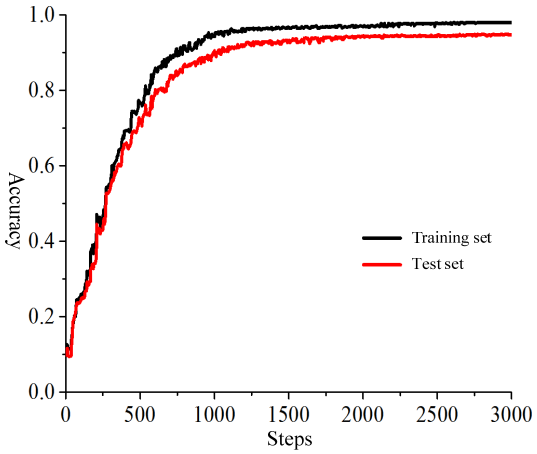


FIGURE 3. Accuracy change during training.

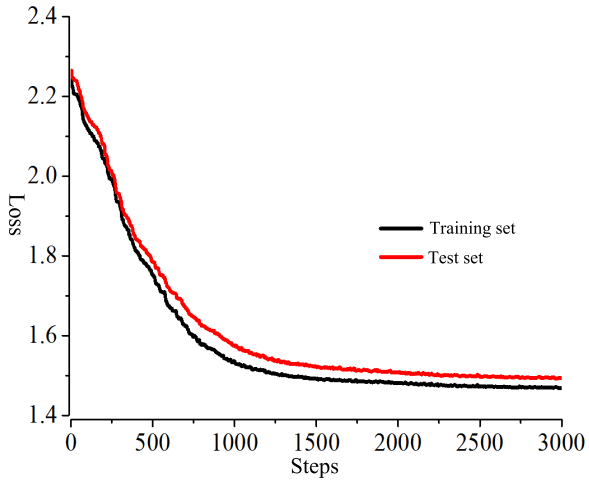


FIGURE 4. Loss change during training.

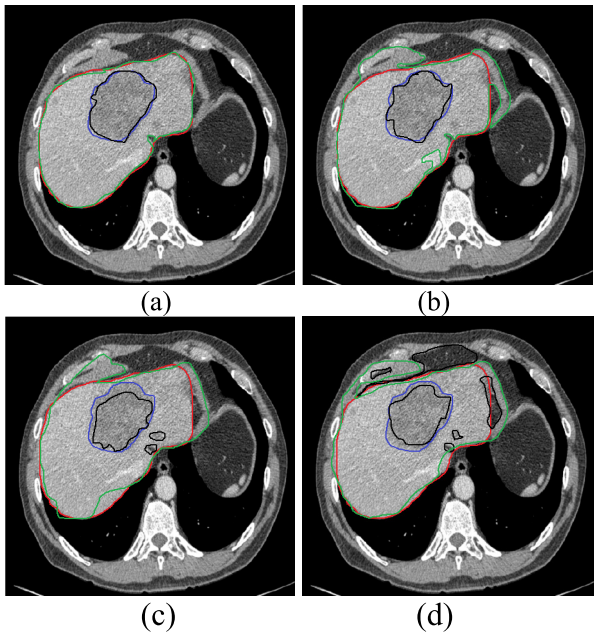


FIGURE 5. Segmentation results by different deep models. (a) Our proposed model; (b) Deconvnet; (c) Segnet; (d) Deeplabv3.

TABLE 2. Semantic segmentation indicator results under deeplesion data-set.

Models	PA(%)	MA(%)	IoU(%)
Deeplabv3	*86.13*	86.00	*81.47*
DeconvNet	84.09	85.91	80.17
SegNet	79.25	*87.03*	78.57
Proposed	**87.24**	**88.29**	**85.03**

In order to facilitate the analysis of the results of semantic segmentation by different models, we use liver contours to qualitatively analyze the accuracy of segmentation results, which can characterize spatial consistency very intuitively. Figure 4 shows segmentation results by different deep learning models. Ground truth is shown in blue and red curves, where the red curve is denoted as the contour of liver and blue is the edge of the tumor. Black curve and green curve represent the target contours segmented by different algorithms, respectively. It can be seen that the deformed contours through our proposed model are closer to the liver boundary, which shows that the deep network architecture proposed in this paper can stably improve the performance of the semantic segmentation model.

2) ANALYSIS OF QUALITATIVE AND QUANTITATIVE RESULTS FOR SELF-BUILT DATA SET

Table II shows the results of different semantic subsets of the Benchmark DeepLesion data set for different test subsets, where black and *italics* are the best and second best results, respectively. Our experiments are conducted using 5-fold cross-validation. It can be seen that the proposed algorithm is optimal for all algorithms, mainly due to the semantic segmentation model proposed in this paper. The Pix2pix network is used as Generative Adversarial Networks model, and the segmentation architecture with deep features and multi-scale semantic features is combined. The model introduced a weighted loss function to enhance the learning ability of the network to characterize the target. In the DeepLesion data set, we chose the most complicated Abdominal CT images. Our proposed algorithm in this test is better than Deeplabv3, which mainly due to the characteristics that the model has high spatial consistency and improves the accuracy of quantitative indexes.

Figure 4 shows the semantic segmentation results of different models for the DeepLesion data set, where 4(a) is the original CT image; 4(b) is the result of semantic segmentation of the SegNet model; 4(c) is the result of semantic segmentation of the DeconvNet model; 4(d) is the result of semantic segmentation of the Deeplabv3 model; 4(e) is the result of semantic segmentation of the proposed model in this paper; 4(f) is the benchmark result. The red region

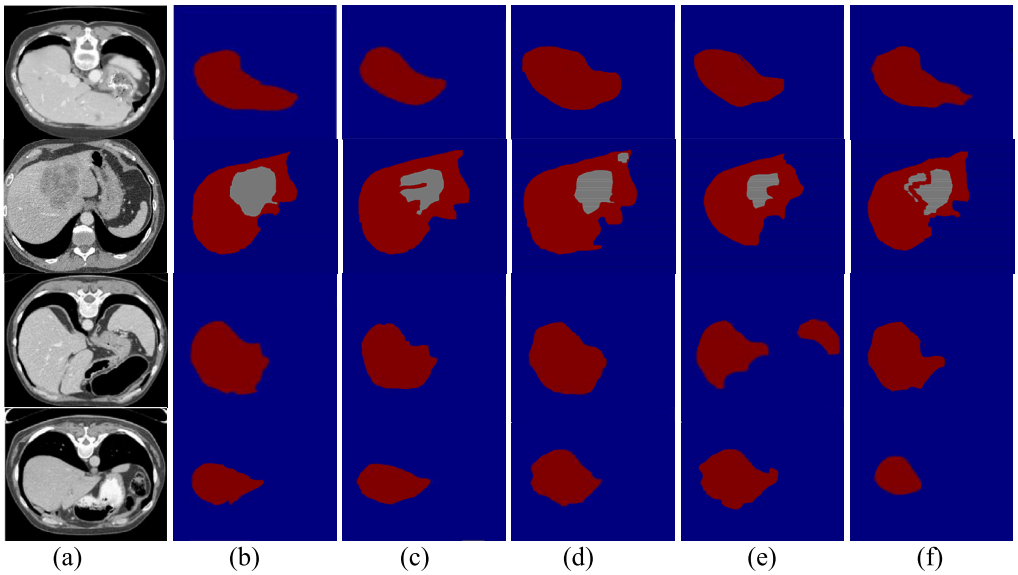


FIGURE 6. Comparison of results for different segmentation models. (a) Raw CT image; (b) Ground truth; (c) Proposed; (d) Deeplab v3; (e) SegNet; (f) DeconvNet.

represents the liver, the blue region is the background other than the liver, and the gray is the liver tumor region. As can be seen from the results, these deep learning architectures have achieved very good semantic segmentation results, but they also have their problems. As for Fig.6, the image is from the abdomen CT, whose boundary of the liver region is blurred. In the comparison result, Deeplabv3 is similar to the proposed model in this paper, and the liver can be stably segmented, but the pixel precision is lower than that of our model, mainly due to spatial inconsistency reduced segmentation accuracy. The results of SegNet and Deeplabv3 are rough, especially with a lot of sawtooth at the edge of the liver. This is mainly because SegNet generates semantic probability maps through convolutional layers and some skipped-connections, and then gradually refines the accuracy of semantic segmentation. However, SegNet directly uses the Softmax loss function to judge whether it is in the processing of boundaries or mall objects is rough and the spatial consistency is poor. The structure of DeconvNet is very similar to that of SegNet, but the network uses a fully connected layer as a relay between the encoder and the decoder; Deeplabv3 can adjust the filter field of view and control the powerful response of the convolutional neural network to calculate the characteristic response resolution and a BN layer has been added to the ASPP. The atrous convolution with different sampling rates can effectively capture multi-scale information, but the simulation results show that the effective weight of the filter becomes smaller as the sampling rate increases, which leads to the performance degradation.

D. COMPARATIVE ANALYSIS OF TRAINING CONVERGENCE FOR DIFFERENT MODEL

In this section, we conduct comprehensive experiments to analyze the effectiveness of our proposed model.

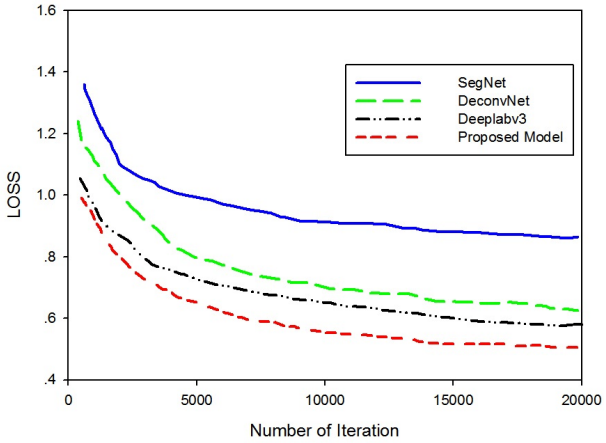


FIGURE 7. The training performance of all the algorithm models.

Figure 7 shows the training losses of different comparison models, which mainly represents the convergence performance of the whole algorithm. Note that SegNet model costs around 48 hours, nearly 1 time than Deeplab_v3. DeconvNet model costs nearly 42 hours, where 27 hours are spent for CNN training and 15 hours are to fine-tune the whole architecture in an end-to-end manner. It is worth mentioning that all of the models are run with the same hardware platform. It can be seen that the combination of Deeplabv3 and the Generative Adversarial Network can not only obtain a higher segmentation index, but also has more stable convergence performance. The performance of DeconvNet and SegNet is worse, which shows that the neural network trained by the algorithm framework has a better convergence effect.

The semantic segmentation model proposed in this paper first initializes the shared convolution layer with the pre-trained parameters of the residual network, and initializes

the GAN detection and semantic segmentation module with Xavier. In the early stage of model training, the strategy of alternating training is adopted: first input the object image and the forward propagation and back propagation parameters of the GAN module are updated; the semantic segmentation image is input, and the forward propagation and back propagation parameter update of the semantic segmentation are completed on the basis of the update parameters of the generator module in the previous step, and the two are alternately performed. Training of modules until both modules tend to converge. Once the alternation training is completed, the loss function of the two modules is proportionally weighted to obtain the total loss function. The total loss function is optimized by the Adam algorithm, and the appropriate weights are set for the two loss functions. Finally, the fusion network model can obtain the result of semantic segmentation in only one calculation.

V. CONCLUSION

Since the existing semantic segmentation algorithm has the problem of inconsistent segmentation result for segmentation of Liver, this paper proposed a multi-scale adversarial network semantic segmentation algorithm combined with a weighted loss function. This algorithm introduced Pix2pix network as a generative adversarial network model on the basis of the basic framework of DeepLab v3 so as to achieve multi-scale confrontation network semantic segmentation. In order to increase the generalization ability and training precision of the model, it proposed to combine the traditional multi-class cross entropy loss function with the content loss function of the generator output and the adversarial-loss function of the discriminator output to construct a weighted loss function. A large number of qualitative and quantitative experiments show that the deep network architecture proposed in this paper can stably improve the performance of the semantic segmentation model. In future work, we will optimize the algorithm and embed the module into the medical equipment to feedback the diagnosis results in real time and accurately so as to improve the automation level of Liver cancer examination.

REFERENCES

[1] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artif. Intell. Rev.*, vol. 50, no. 4, pp. 497–537, 2018.

[2] X. Lu, Q. Xie, Y. Zha, and D. Wang, "Fully automatic liver segmentation combining multi-dimensional graph cut with shape information in 3D CT images," *Sci. Rep.*, vol. 8, no. 3, pp. 10700–10715, 2018.

[3] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[4] J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T.-S. Chua, "Micro tells macro: Predicting the popularity of micro-videos via a transductive model," in *Proc. ACM Int. Conf. Multimedia*, 2016, pp. 898–907.

[5] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *Comput. Vis. Pattern Recognit.*, vol. 32, no. 3, pp. 1182–1199, 2015.

[6] F. Chaieb, T. B. Said, S. Mabrouk, and F. Ghorbel, "Accelerated liver tumor segmentation in four-phase computed tomography images," *J. Real-Time Image Process.*, vol. 13, no. 1, pp. 121–133, 2017.

[7] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," 2016, *arXiv:1611.08408*. [Online]. Available: https://arxiv.org/abs/1611.08408

[8] Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, and S.-H. Wang, "Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3613–3632, 2019.

[9] E. Vorontsov, A. Tang, D. Roy, C. J. Pal, and S. Kadoury, "Metastatic liver tumour segmentation with a neural network-guided 3D deformable model," *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 127–139, 2017.

[10] K.-J. Xia, H.-S. Yin, and Y.-D. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-flow algorithm," *J. Med. Syst.*, vol. 43, no. 1, p. 2, 2019.

[11] J. Wang, Z. Wang, D. Tao, S. See, and G. Wang, "Learning common and specific features for RGB-D semantic segmentation with deconvolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 664–679.

[12] 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. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[13] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2016, vol. 24, no. 3, pp. 1–16.

[14] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, vol. 32, no. 2, pp. 1–17.

[15] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," in *Proc. Neural Inf. Process. Syst. Conf. (NIPS)*, 2016. [Online]. Available: https://arxiv.org/abs/1701.00160

[16] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *Comput. Vis. Pattern Recognit.*, vol. 22, no. 7, pp. 1182–1189, 2017.

[17] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.

[18] D. Warde-Farley and Y. Bengio, "Improving generative adversarial networks with denoising feature matching," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–11.

[19] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann, "Stabilizing training of generative adversarial networks through regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 2018–2028.

[20] Tensorflow. (2018). *DeepLab: Deep Labelling for Semantic Image Segmentation*. [Online]. Available: http://github.com/tensorflow/models/tree/master/research/deeplab

[21] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Medical Image Computing and Computer Assisted Intervention*. Berlin, Germany: Springer, 2017, pp. 408–416.

[22] Y. Deng, Y. Shen, and H. Jin, "Disguise adversarial networks for click-through rate prediction," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 1589–1595.

[23] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process.*, vol. 35, no. 1, pp. 53–65, Jan. 2017.

[24] I. Gatos, S. Tsantis, M. Karamesini, S. Spiliopoulos, D. Karnabatidis, J. D. Hazle, and G. C. Kagadis, "Focal liver lesions segmentation and classification in nonenhanced T2-weighted MRI," *Med. Phys.*, vol. 44, no. 7, pp. 3695–3705, 2017.

[25] N. Jain and V. Kumar, "Liver ultrasound image segmentation using region-difference filters," *J. Digit. Imag.*, vol. 30, no. 3, pp. 376–390, 2017.

[26] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, pp. 58–66, Nov. 2017.

[27] P. Saiviroonporn, P. Korpraphong, V. Viprakasit, and R. Krittayaphong, "An automated segmentation of R2* iron-overloaded liver images using a fuzzy C-mean clustering scheme," *J. Comput. Assist. Tomogr.*, vol. 42, no. 3, pp. 387–398, 2018.

[28] A. Hoogi, C. F. Beaulieu, G. M. Cunha, E. Heba, C. B. Sirlin, S. Napel, and D. L. Rubin, "Adaptive local window for level set segmentation of CT and MRI liver lesions," *Med. Image Anal.*, vol. 37, pp. 46–55, Apr. 2017.

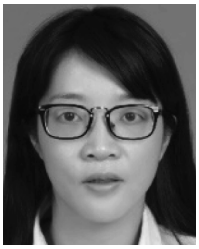
[29] D. Li, W. Zhong, K. M. Deh, T. D. Nguyen, M. R. Prince, Y. Wang, and P. Spincemaille, "Discontinuity preserving liver MR registration with three-dimensional active contour motion segmentation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 1884–1897, Jul. 2019.
[30] J. Y. Jang, H.-S. Han, Y. S. Yoon, J. Y. Cho, Y. Choi, W. Lee, H. K. Shin, and H. L. Choi, "Three-dimensional laparoscopic anatomical segment 8 liver resection with Glissonian approach," *Ann. Surgical Oncol.*, vol. 24, no. 3, pp. 1606–1609, 2017.
[31] A. S. Maklad, M. Matsuhiro, H. Suzuki, Y. Kawata, N. Niki, M. Satake, N. Moriyama, T. Utsunomiya, and M. Shimada "Blood vessel-based liver segmentation using the portal phase of an abdominal CT dataset," *Med. Phys.*, vol. 40, no. 11, 2013, Art. no. 113501.
[32] P. Bilic *et al.*, "The Liver Tumor Segmentation Benchmark (LiTS)," 2019, *arXiv:1901.04056*. [Online]. Available: https://arxiv.org/abs/1901.04056

KAIJIAN XIA was born in Jiangsu, China, in 1983. He received the master's degree from Jiangnan University. He is currently pursuing the Ph.D. degree with the China University of Mining and Technology. He is also with the Department of Computer, Changshu No. 1 People's Hospital. His research interests include medical information and medical image proceeding, and so on. He has published several papers in international journals, including the *Journal of Medical Systems*, and the *Journal of Medical Imaging and Health Informatics*.

HONGSHENG YIN was born in 1967. He was with the China University of Mining and Technology. He is currently a Professor. He has been engaged in the teaching and research work of mine monitoring and monitoring information processing, and computer network communication.

PENGJIANG QIAN (M'12) received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2011. He is currently an Associate Professor with the School of Digital Media, Jiangnan University, and the Case Western Reserve University, Cleveland, OH, USA, as a Research Scholar in medical image processing. His current research interests include data mining, pattern recognition, bioinformatics and their applications, such as analysis and processing for medical imaging, intelligent traffic dispatching, and advanced business intelligence in logistics. He has authored or coauthored over 30 papers in international or national journals and conferences.

YIZHANG JIANG (M'12) received the Ph.D. degree from the School of Digital Media, Jiangnan University, Wuxi, China, in 2016. He was also a Research Assistant with the Computing Department, The Hong Kong Polytechnic University, for almost two years. He has published several papers in international journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and the ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY. He is the author or coauthor of more than 20 research papers in international and national journals. He has served as the Reviewer for several international conferences and journals, including the ICDM, TKDE, TFS, TNNLS, TCYB, TSMCA, TII, and *Neurocomputing*. His research interests include pattern recognition, intelligent computation, and their applications.



SHUIHUA WANG received the B.S. degree from Southeast University, in 2008, the M.S. degree from the City University of New York, in 2012, and the Ph.D. degree from Nanjing University, in 2016. She is currently an Assistant Professor with Loughborough University, Loughborough, U.K.