IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 4, 2020, accepted June 13, 2020, date of publication June 29, 2020, date of current version July 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005384

Hybrid Segmentation Algorithm for Medical Image Segmentation Based on Generating Adversarial Networks, Mutual Information and Multi-Scale Information

YI SUN[1], PEISEN YUAN[2], (Member, IEEE), YUMING SUN[1], AND ZHAOYU ZHAI[3]

[1]School of Computer Science, Fudan University, Shanghai 210043, China
[2]College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China
[3]Departamento de Ingeniería Telemática y Electrónica (DTE), Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación (ETSIST), Universidad Politécnica de Madrid (UPM), 28031 Madrid, Spain

Corresponding author: Peisen Yuan (peiseny@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806097, in part by the Shanghai Science and Technology Committee under Grant 17511104200, Grant 18411952100, and Grant 17411953500, and in part by the College Students Entrepreneurship Training Program 2019 under Grant S20190025.

ABSTRACT This paper proposes 3D-MedGAN, MLU-Net and Info-Max-Net models for overcoming the lack of labeled data and extracting the multi-level feature of images in medical image segmentation. 3D-MedGAN is aimed at dealing with the lack of labeled data in medical images. It uses a generative adversarial network to simulate data and then draws newly generated samples from the distribution learned by the model. Training the segmentation model by mixing generated samples with real samples can effectively improve the effect of the segmentation model. MLU-Net uses multiple layers of different levels of convolutional angles to extract feature information from multiple angles in medical images. By adopting the attention mechanism to fuse the multi-level feature information, MLU-Net is able to improve the feature expressions and segmentation effect. Info-Max-Net is aimed at handling the noise problem in medical images. When the information in the images is complex, it is difficult to extract features. Using mutual information to measure the dependency between the image and the extracted features can effectively reduce noise in the image and improve the effect of segmentation. At the same time, for solving the problem that the high dimension of the image makes it difficult to measure mutual information, this paper uses a lower bound BL-estimator to measure the mutual information between the optimized image and the extracted features. Therefore, the model can maintain a high convergence speed as it approaches the true value of mutual information. Considering that the quality of images generated by 3D-MedGAN are not as good as the original images, we combine the 3D-MedGAN, MLU-Net, and Info-Max-Net to improve the sensitivity and the power of feature extraction of the hybird model. The effectiveness of our model is verified through experiments of the 3D-MedGAN, MLU-Net, Info-Max-Net, and the hybrid model over the LIVER100 dataset.

INDEX TERMS Medical image segmentation, generative adversarial network, mutual information maximization, multi-scale information.

I. INTRODUCTION

Medical image segmentation is a key and complex step in the field of medical image processing and analysis. Its key point emphasizes on the basic step of pathologic localization and anatomic structure research. Accurate image segmentation can provide reliable evidence for clinical diagnosis and pathology research, assist doctors to establish more accurate diagnosis and make a better treatment plan. At the same time, medical image segmentation is complicated. In particular, automatic segmentation from medical images is a difficult task, because medical images have high complexity and

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

lack of simple linear features. The accuracy of segmentation results is also affected by volume effect, gray inhomogeneity, artifacts, the proximity of different soft tissue gray levels, etc. [1] As a consequence, medical image segmentation remains to be explored and can be further improved.

Medical image segmentation methods can be broadly divided into the following categories: (1) threshold-based segmentation [2]: one or more gray thresholds are calculated based on the gray characteristics of the image. The gray values of each pixel in the image are compared with the threshold values, and these pixels are classified into appropriate categories according to the comparison results; (2) edge-based segmentation [3]: the gray values of pixels on the boundaries of different regions usually vary greatly. If the image is transformed from spatial domain to frequency domain through Fourier transform, the edges correspond to the high frequency part. According to this characteristic, the edge pixels can be determined first and then connected together to form the boundary between regions; (3) region based segmentation [4], [5]: This method makes use of the feature of smooth and uniforms the surface of objects to achieve the segmentation. Because the smooth surface corresponds to the region with constant intensity or slowly-changing intensity in the image, the region with the uniform property can be separated through the region growth method or the splitting and merging method; (4) segmentation based on fuzzy theory [6]: this method improves the method of ''one-size-fits-all'' approach, introduces the concept of ''membership degree'' in the fuzzy theory, and divides pixel points into regions with a high membership degree; (5) segmentation based on deep learning [7]–[10]: by simulating the abstract and iterative process in the human visual system through the convolution neural network, the deep features perceived by the brain can be extracted. Then the deconvolution layer samples the deep features and classifies pixels one by one through the sampling process to complete image segmentation. Nowadays, deep learning methods have achieved a great success in medical image segmentation. Despite the success of deep learning methods, two major problems in medical image segmentation have been identified:

- There is a serious lack of labeled data in medical images: Because labeling medical images needs professional annotation, it requires a considerable level of professional medical literacy, time and cost to obtain a labeled medical image data. At the same time, medical images also involve the privacy of patients, so labeled medical image data are sometimes unavailable, leading to great diffculties to develop the medical image segmentation model;
- The noise and dimensional problem in medical images: Because of the high dimension and the serious noise in medical images, it is very diffcult to extract the features from image data.

To address those two problems, we proposed three models: 3D-Medical GAN (3D-MedGAN), Multi-Level U-Net (MLU-Net) and mutual Infomation Maximization Net (Info-Max-Net). Then, we proposed a hybrid model to aggregate the advantage of these models. Our contributions are listed as follows.

- In view of the above two problems in medical image segmentation, we proposed three models, which can effectively solve the problems of medical image segmentation with insufficient labeled data, multi-scale feature extraction of medical image and noise in medical image.
- In order to solve the problem that the image generated by 3D-MedGAN contains more noise and it is difficult to extract features, we mixed the three models. By using Info-Max-Net and MLU-Net, we can effectively extract the original image and extract features of the generated image, so as to achieve better segmentation results
- Experimental results of medical image segmentation based on LIVER 100 dataset demonstrated the effectiveness of the proposed model and the hybrid model

The brief introduction of these three models can be found in the next sub-sections.

A. INTRODUCTION TO 3D-MedGAN

The rapid development of deep learning technologies in the field of medical image segmentation is based on the acquisition of a large amount of high-quality data. However, the amount of available medical image data is very limited, resulting in a major obstaclein applying deep learning technologies to the medical field. The lack of available medical image data is embodied in two aspects, one is the lack of images, while the other is the lack of professional annotations, which directly affects the training model. Aiming at overcoming the challenge of acquiring training data of medical images, the proposed solution is to enlarge the training data set by generating the synthetic simulation samples through the Generative Adversarial Networks (GAN) [11], so as to alleviate the problem of insufficient labeled data.

Based on 3D Unet, we propose an end-to-end network architecture, which can synthesize labelled 3D Nuclear Magnetic Resonance (NMR) image data by using the conditional generative adversarial networks. Users can modify the dimension, shape and position of the real data, and then input the modified annotation into the trained model to get the corresponding Magnetic Resonance (MR) image. We referred to this model as 3D-Medical GAN.

We used the 3D Unet [12] network as our generator *G*. The batch normalization layer [13] in the lower sample block and the upper sample block is replaced by the instance normalization layer [14]. The batch size of our experiment was set to 1 due to GPU memory limitations. The activation function in the down-sampling and up-sampling blocks is LeakyReLU. The activation function for the final output module uses tanh for the image generation task. Besides, we added the spectral normalization operation [15] to each convolution layer to stabilize the training of the entire generative adversarial networks. The same scheme is applied to the discriminator *D* [16] as well. Spectral normalization constrains the

Lipschitz constant of the whole network by constraining the spectral norm of the weight matrix at each level. It does not require additional hyper-parameter adjustments, and the additional computational cost is low. SimGAN [17] had been proved as a reliable and effective way to use discriminators to help generators simulate local features in 2D images. By limiting the size of the receptive field of the convolution kernel, D identifies the local region of the input image. This method not only reduces the parameters of the discriminator network but also enriches the training samples of the discriminator.

B. INTRODUCTION TO MLU-Net

In the field of computer vision, extracting image features is a crucial issue, which is related to the feasibility of subsequent models, and the effectiveness of the classification and recognition system. In the medical image, there are many features such as the organ and the lesion in the image. At the same time, the edge of the medical image is often blurred and the precision of segmentation is required. To address this issue, we proposed a model, called MLU-Net, which used multi-level information and attention for medical image segmentation. The overall framework adopted by MLU-Net is based on U-Net networks. It has a symmetrical set of encoders and decoders on both sides of the network and can be directly connected by skipping the connection between the corresponding encoders and decoders. In Figure 1, the first line is the structure of the encoder and the second line is the structure of the decoder. The U-shaped structure of the network and its characteristics of skip connections are the core of the network. This structure can help the encoder to extract the deep spatial features with multiple complexities, and enable the decoder to receive the feature information with multiple complexities.

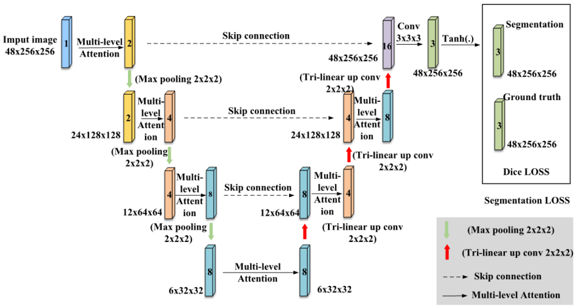


FIGURE 1. **The framework of MLU-Net.**

Our network was improved on the basis of the traditional U-Net [12] by introducing a special multi-layer attention [18] unit. Figure 2 shows this multi-layer structural unit. This structure contains two key technologies, namely, multi-layer extraction mechanism and attention mechanism. Multi-layer paths provide encoders with rich semantic information. Each cell contains several convolution filters, which can improve the performance of the convolution layer and the efficiency of the network.

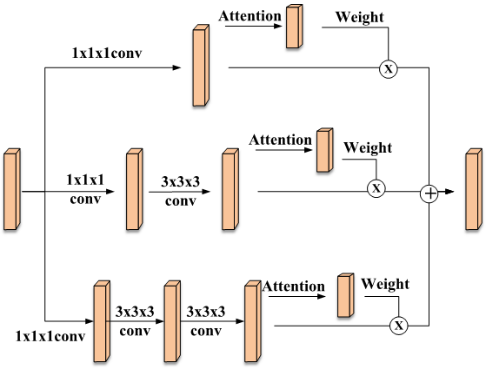


FIGURE 2. **The framework of Multi-Level block.**

C. INTRODUCTION TO Info-Max-Net

Because of medical equipment and operation, medical images often contain a lot of noise, resulting in a huge obstacle to the image feature extraction. Aiming at tackling this problem, we proposed a model, called Info-Max-Net, which captures useful features and filter noise unsupervised by maximizing the mutual information [19] between the original image and the feature coding.

Info-Max-Net consists of two subnets and a bilinear interpolation function discriminator. One subnet performs feature extraction and image segmentation, while the other subnet and discriminator simultaneously estimates and maximizes the mutual information between the image and the feature coding to improve the quality of image deep features. The Info-Max-Net model can filter the noise of medical images without supervision, capture the information which is helpful for classifying the images, and improve the accuracy of image segmentation.

Info-Max-Net utilizes a typical U-shaped network for feature extraction and image segmentation. The subnetwork U-Net consists of a contraction path (to the left) and an expansion path (to the right). The contraction path follows the typical architecture of convolution networks, consisting of three down-sampled modules, each of which contains two 3×3 convolutions followed by a Rectified Linear Unit (ReLU), and a 2×2 largest pooled layer. With each down-sampling step, the number of feature channels is doubled, and finally the feature coding is obtained. In each step of the expansion path, the feature spectrum is sampled first, then 2×2 deconvolved, and then the number of feature channels is halved, connected with the feature spectrum obtained from the corresponding levels in the contraction path, followed by a linear rectification function through two 3×3 convolutions. The boundary pixels should be cropped, since they are lost during each convolution. Finally, the 1×1 convolution maps each pixel to its own category.

The number of categories (including the background) we split is L, the number of pixels is N, the p_{ic} indicates the probability of predicting that the i-th pixel belong to the c category, and the g_{ic} represents the 0-1 value of whether the i-th pixel actually belong to the c-th category, then the split loss function [20] for the first subnet can be expressed as

$$\text{DiceLoss} = \frac{1}{L}\sum_{c=1}^{L}\left(1 - \frac{2\sum_{i=1}^{N} p_{ic} g_{ic}}{\sum_{i=1}^{N} p_{ic}^{2} + \sum_{i=1}^{N} g_{ic}^{2} + \epsilon}\right), \quad (1)$$

where, ϵ is a small positive number to ensure that the top denominator is not 0.

Another subnetwork of Info-Max-Net is the mutual-information encoder, which still uses U-Net, similar to the first subnetwork. Input images from the first subnetwork and other random images are input into the mutual information encoder, and the last layer of the mutual information encoder is used as their characteristic spectra.

We linked the feature coding of the matched input image with the feature spectrum as the positive sample of the training discriminator and link the feature coding of the input image extracted from the first subnet with the feature spectrum of the other images computed from the second subnet as the negative sample of the training data. The positive and negative samples are input into the bilinear function of the discriminator, and the discriminator calculates the score.

From the specification of the BL-mutual information estimator [21], mutual information is estimated as follows,

$$\mathbb{E}_{\mathbb{P}}[\log\sigma(T_\theta(x, E_\psi(x)))] + \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[\log(1-\sigma(T_\theta(x', E_\psi(x))))] \quad (2)$$

The essence of parameter θ is to train a discriminator that needs to distinguish the matching between the feature map and the feature code in the sample. If the discriminator achieves a high score of $(x, E_\psi(x))$ for a matched image from a distribution of $\mathbb{P}$, and reaches a low score of $(x', E_\psi(x))$ for a mismatched image and encoding (x'), then the value of the BL-estimator will approximate to its supremum, thus approaching to the true value of mutual information.

It is required to find the discriminant function parameter θ which maximizes the BL-estimator to ensure that the value of the estimator is close to the true value of mutual information. It is also necessary to identify the base encoder parameter ψ which maximizes the BL estimator to ensure that the learned image encoding is optimal. The parameters θ and ψ can be determined by the following formula

$$(\theta^*, \psi^*) = \arg\max_{\theta,\psi} \mathbb{E}_{\mathbb{P}}[\log\sigma(T_\theta(x, E_\psi(x)))] + \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[\log(1-\sigma(T_\theta(x', E_\psi(x))))].$$

If the discriminator T_θ can easily determine whether an image matches the encoding, then the encoding has a strong distinction between other images, meaning that the model captures useful information from the original image and filters out noise, so it is the encoding we want.

In addition, in order to train both the subnetworks and the discriminator parameters at the same time, we maximized the discriminator's target function, the BL mutual information estimator, and equivalent to minimize the discriminant loss function. Let the number of training samples be M, $T_\theta(x^{(i)}, y^{(i)})$ denotes the discriminator's score on the i, and h_i represents whether or not the i sample is actually a value of 0-1 for a positive sample, 1 for a positive sample, and 0 for a negative sample, then the loss function of the Info-Max-Net discriminator can be expressed as a loss function

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{M}\sum_{i=1}^{M}\Big[h_i \log\sigma(T_\theta(x^{(i)}, y^{(i)})) + (1-h_i)\log(1-\sigma(T_\theta(x^{(i)}, y^{(i)})))\Big].$$

Thus, the loss function of the Info-Max-Net is an addition of the split loss function and the discriminant loss function. When its gradient drops to 0, the two subnetworks and the discriminator are well trained.

We use the bilinear interpolation function as the discriminator of Info-Max-Net. The result of bilinear interpolation is not linear, but the product of two linear functions. Assuming that the number of pixels in the feature map is N, the number of components in the feature code is K, the i-th pixel of the feature map is x_i, and the j-th of the feature code is y_j, then the bilinear function is expressed as follows.

$$T_\theta(x, y) = \sum_{i=1}^{N}\sum_{j=1}^{K}\theta_{ij}x_i y_j + \sum_{i=1}^{N}\theta_{i,0}x_i + \sum_{j=1}^{K}\theta_{0,j}y_j + \theta_{00},$$

where $\theta_{ij}(i \geq 1, j \geq 1)$ is a quadratic parameter, $\theta_{i,0}$ and $\theta_{0,j}$ are a single parameter, and θ_{00} is a constant parameter.

D. EXPERIMENTS OF THREE MODELS

In this subsection, we show the experimental results of 3D-MedGAN, MLU-Net, and Info-Max-Net.

1) EXPERIMENTS OF 3D-MedGAN

We applied the 3D-MedGAN model to the BraTS15 and BraTS17 datasets. We trained four 3D Unet models on BraTS datasets for different modes (T1, T1ce, T2 and FLAIR). Then these models were used to segment the synthesized MR images. After that, we compared the differences between the segmentation results and the labeled images used in the synthesized images. Compared with the commonly used GAN metrics, such as inception score (IS) and Fréchet inception distance (FID), the above method can directly evaluate the effect of data enhancement of these synthetic data in the segmentation task. These models were still trained using the Adam optimizer, with a learning rate of 2.0×10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. Based on the BraTS competition criteria, the segmentation results should be compared with the original label map on three sub-tasks: (1) Intact tumor (necrotic and unenhanced tumor, enhanced tumor and edema) (2) Tumor core (necrotic and unenhanced and enhanced tumor), and (3) Enhanced tumor region. We used dice score as a metric.

The evaluation results are displayed in Figure 3. It is shown that the quality of the composite image is acceptable and reliable. Four trained 3D Unet models segmented the "entire tumor" region of the real image with a dice score, more than 0.9. For the composite image this index did not drop significantly, and the four modes on the dice score were

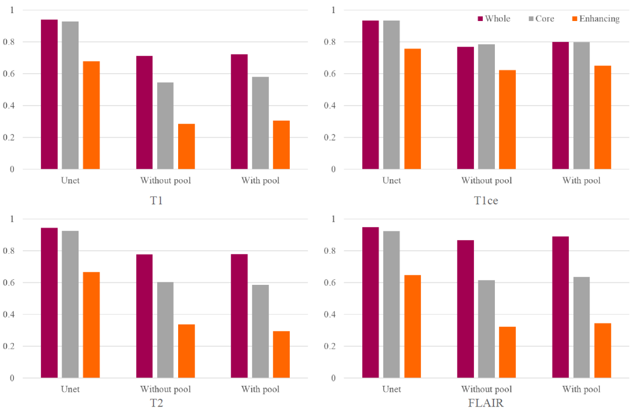


FIGURE 3. Using 3D Unet to evaluate the quality of the composite image. The purple, gray, and orange columns represent the 'whole tumor', the 'tumor core', and the 'enhanced tumor' task dice score, respectively. Four histograms correspond to four modal 'T1s', 'T1ce ', 'T2s' and 'FLAIRs' respectively. The three groups of columns in the diagrams, 'Unet', 'Without pool' and 'With pool', represent the results of 3D Unet model versus real data, data synthesized by networks without image buffer pooling and data partitioning by networks with that mechanism respectively.

close to 0.8. However, due to class imbalance and insufficient data in the training-generated model, the relatively small regions of "tumor core" and "enhanced tumor" can not be well restored in the synthetic image. In T1, T2 and FLAIR modalities, the "enhanced tumor" scores of MR images were all lower than 0.4, while in real images the scores were all higher than 0.6. Compared with the score of 0.92, 0.92 and 0.76 on the real image, the composite image achieved good scores of 0.78,0.79 and 0.62 on the three regions, respectively.

In the following experiments, we do extensive experiments to verify whether the performance of the segmentation model is improved by training false samples and real samples. The experimental results is shown in Table 1. The experimental data show that these false samples significantly improve the performance of brain tumor segmentation network. Besides, our bogus samples also provide good protection for personal health information. Even if the sample is synthesized with a real annotation, it will show a significant change compared to the real image corresponding to the annotation. It is almost impossible to identify the owner (patient) of a synthetic image if it further erases information on the synthetic image that does not affect tumor segmentation, such as the DICOM primordial data or the removal of the skull. This means that the agency that collects the original MR image can train the 3D-MedGAN with the original image, and then share the sample data from the 3D-MedGAN with researchers outside the agency without worrying about any disclosure of patient privacy. To a large extent, it breaks the restriction of the ethics committee and promotes the sharing of medical data.

2) EXPERIMENTS OF MLU-Net AND Info-Max-Net

We experimented with MLU-Net on the liver image segmentation dataset LIVER 100. The dataset had a total of 100 samples, of which we used 80 as a training set and 20 as a test set. Because the volume of the tumor region in each sample is very small, preprocessing is required to balance the computational resources and the data volume. Firstly, we used the interpolation algorithm to sample the training set, and extract the small pieces which contain the tumor region in the sampled area. Because the deep neural network needs a large number of training sets, we used random methods to meet the data requirements. For each set of data, we first extracted a slice of 48 × 256 × 256, and then input it to the training network, setting the number of batches to 1. During the training process, the data were sampled down and cut into chunks of 48 × 256 × 256. Eventually, the results of the network prediction were combined together and interpolated to the size of the initial state of the previous sample.

TABLE 1. Performance (Dice Score) of the segmentation networks trained by the different data components gained from *BRATS17*. The first part are the results of model based on 3D U-Net [22], and the second are those of Triple-Cascaded-Net [23]. Here, the F is the abbreviation of fake data, and R is the abbreviation of real data. An obvious increase is presented in 3D U-Net.

Data Component	WT	TC	ET
100% R	0.8722±0.14	0.7525±0.24	0.6638±0.33
100% F	0.6539±0.25	0.5170±0.28	0.4117±0.34
100% F + 100% R	0.8720±0.14	0.7621±0.21	0.6632±0.33
300% F + 100% R	0.8875±0.09	**0.7685±0.20**	0.6618±0.33
500% F + 100% R	**0.8891±0.10**	0.7674±0.20	**0.6662±0.32**
100% R	0.8989±0.07	**0.8356±0.16**	0.7349±0.29
100% F	0.5227±0.30	0.3763±0.30	0.3380±0.31
100% F + 100% R	0.8981±0.07	0.8249±0.15	0.7321±0.28
300% F + 100% R	0.8963±0.08	0.8298±0.16	0.7346±0.28
500% F + 100% R	**0.8990±0.08**	0.8305±0.15	**0.7379±0.27**

We compared the proposed MLU-Net with some representing models from this year in the task of medical image segmentation and extraction. The results were listed in Table 2. From this table, we know that our model scored 4.6%, 1.9%, 1.1%, 3.0%, and 5.3% higher than U-Net, U-Net++, DialResNet, AgNet, and RA-UNet for liver extraction tasks, respectively. Our model scored 23.7%, 28.9%, 10.5%, 27.4%, and 23.6% higher than above methods respectively, in the tumor extraction task. Therefore, our model achieved better performance than all other approaches in the segmentation and extraction tasks.

TABLE 2. Performance on LIVER 100 Liver and Tumor data.

Methods	Liver dice score	Tumor dice score
U-Net	0.7003	0.2899
U-Net++	0.7309	0.2782
DialResNet	0.7247	0.3245
AgNet	0.7112	0.2814
RA-UNet	0.6955	0.2902
MLU-Net	0.7323	0.3587
Info-Max-Net	0.6522	0.3499

We also applied Info-Max-Net to the liver tumor segmentation data set LIVER 100. Data set preprocessing and model evaluation methods are consistent with those in MLU-Net.

We trained five previously proposed models, U-Net [24], U-Net++ [25], DialResNet [20], AgNet [26], and RA-Unet [27], on the LIVER100 dataset with 80% of the training data, and 20% of the test data. The parameters for each model are optimized, and the results are shown in Table 2. It can be seen that the proposed Info-Max-Net model had a gap of about 0.05 dice score in liver segmentation task compared with other models, but it performed better than any other model in the more difficult tumor location task. The tumor dice scores of the Info-Max-Net model increased by 20.7%, 25.8%, 7.5%, 24.3%, 20.6%, compared with U-Net, U-Net++, DialResNet, AgNet, and RA-Unet, respectively.The liver dice score of the Info-Max-Net model is lower than that of the U-Net, U-Net++, DialResNet, AgNet and RA-UNet, respectively, by 6.9%, 10.8%, 10.0%, 8.3%, and 6.2%. It is concluded that Info-Max-Net is able to identify tumors more accurately because, by maximizing mutual information, Info-Max-Net can more accurately capture the underlying features that help to classify tumors in different images, and determine their location and size, compared with the traditional U-Net and its variants.

II. HYBRID OF GENERATIVE ADVERSARIAL NETWORK AND MUTUAL INFORMATION

Despite the improvement of Info-Max-Net, we also found that although we used a bilinear discriminator to improve the generalization ability of the model, the loss function still oscillated repeatedly during the training process due to the small amount of training data, and the segmentation effect of the model depended on the input sequence of training samples and other phenomena. If we can use the 3D-MedGAN model to generate more simulation training images, we can overcome the above shortcomings, stabilize the training process, and enhance the generalization ability of the model. This section introduces the framework of IM-MedGAN taking into annount the combination of 3D-MedGAN and Info-Max-Net.

This paper uses a two-stage model to integrate 3D-MedGAN and Info-Max-Net, as shown in Figure 4. In the first stage, we use 3D-MedGAN to augment the training set. 3D-MedGAN is a generative adversarial network composed of a generator and a discriminator. The generator is a U-shaped network with an activation function LeakyReLU. The discriminator is a multi-layer convolutional neural network with a large to small size and a small to large flux. When training 3D-MedGAN, we input the artificial annotation map of the medical image into the generator, and the generator outputs the simulated medical image map that conforms with the annotation. The generated simulation images and the real images in the training set are input to the discriminator together, and the discriminator makes a distinction. During the game between the generator and the discriminator, the performance of both sides is improved. With the enhancement of the discriminator's ability to distinguish between real images and generating images, the generator also gradually has the ability to generate images that are highly similar to real medical images. After training, we mixed the simulation images output by the 3D-MedGAN generator with the real images in the original training set to obtain a larger training data set.

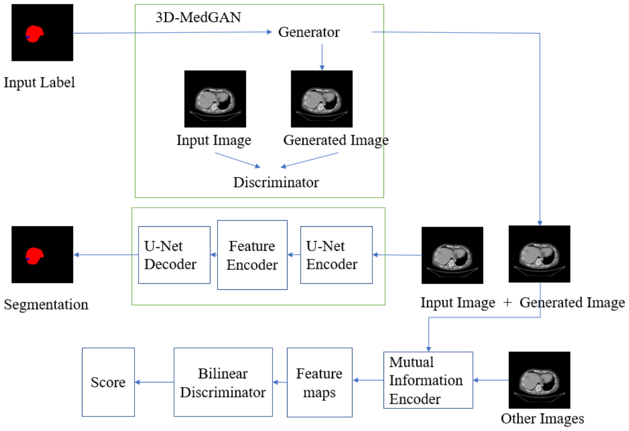


FIGURE 4. The framework of IM-MedGAN.

In the second stage, we apply the expanded training set obtained in the first stage train the Info-Max-Net image segmentation network, and use the trained Info-Max-Net encoder to extract the deep-level features of the medical image. The decoder up-samples the feature code to obtain the medical image segmentation result. The Info-Max-Net model consists of two sub-networks, one of which performs feature extraction and image segmentation, and the other sub-network estimates the mutual information between the input image and the extracted features. We add the loss functions of the two sub-networks and train the entire Info-Max-Net model by Adam gradient descent method, that is, to simultaneously estimate and maximize the mutual information between the input image and its corresponding feature code to obtain the data set. The optimal spatial encoding method and the parameters of the deconvolution layer that upsamples the feature encoding. After training, we input the medical images to be segmented into Info-Max-Net, and through the steps of feature extraction and up-sampling, we obtain the predicted segmentation maps corresponding to the input images.

The IM-MedGAN hybrid segmentation model takes the advantages of the 3D-MedGAN and the Info-Max-Net models to effectively alleviate the problem of insufficient labeled medical data. At the same time, by maximizing mutual information, the quality of image deep features and the accuracy of image segmentation can be improved.

III. HYBRID OF GENERATIVE ADVERSARIAL NETWORK AND MULTI-SCALE INFORMATION

In the experiment, we found that even if we add an attention mechanism to MLU-Net, it adaptively selects locations that need to be processed with high resolution, but because MLU-Net has more convolutional layers, the amount of parameters to be determined is large, and there are few medical images with manual annotations. We still need to use the 3D-MedGAN model to generate simulation images to enrich the original data set. This section introduces a hybrid segmentation model MLU-MedGAN based on generating adversarial networks and multi-scale information.

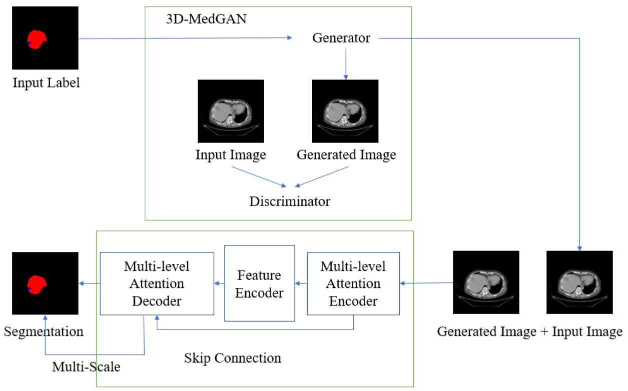


FIGURE 5. The framework of MLU-MedGAN.

This section also uses a two-stage model to combine 3D-MedGAN and MLU-Net, as shown in Figure 5. In the first stage of MLU-MedGAN, we use the 3D-MedGAN model to synthesize simulation samples to expand the training data set, to alleviate the problem of insufficient labeled data. The 3D-MedGAN model restores the continuous distribution of random data from the discrete distribution composed of limited training data by making the generator and discriminator in the adversarial network compete with each other. Sampling is performed on the continuous distribution to obtain countless labeled data that are distinct or similar to the real data. In the second stage of MLU-MedGAN, we input the expanded training set to MLU-Net. MLU-Net extracts the multi-scale features in the training image and makes predictions on the categories to which the image pixels belong at multiple levels. MLU-Net follows a U-shaped structure. The encoder and decoder are symmetrically distributed on both sides of the U-shaped chain, and the corresponding convolution layers contain skip connections. We set up several multi-level units in the encoder part of MLU-Net, and then use the attention mechanism to combine the branches of the multi-level units. This attention mechanism enables the encoder to extract the features of each level of the medical image with high efficiency. The decoder of MLU-Net contains multiple deconvolution layers of different scales, which perform the multi-scale prediction on the category to which pixels belong at different levels.

The MLU-MedGAN hybrid segmentation model can not only greatly expand the training data set, improve the training effect of deep neural networks, but also make multi-scale predictions by combining features at different levels of the image to improve the accuracy of image segmentation.

IV. HYBRID SEGMENTATION MODEL BASED ON MUTUAL INFORMATION AND MULTI-SCALE INFORMATION (IM-MLU-Net)

The accuracy of the Info-Max-Net model for tumor classification is much higher than other models, but its performance on liver segmentation tasks does not reach expectations. After analysis, we found that this is because Info-Max-Net only optimizes the deep-level features, and fails to combine the shallow-level features of the image for multi-scale prediction. Medical images often have blurred borders and complex gradients. The high-resolution information is required to provide details for accurate segmentation. At the same time, the human body structure is relatively fixed, and the distribution of segmentation targets in medical images has a strong law. The deep low-resolution information can capture the characteristics of the target and its surrounding environment, which can be used for segmentation target detection and recognition. Therefore, a single-scale model with only deep or shallow features is often incapable of segmenting medical images. We need to use the MLU-Net model to extract image information of different scales and different levels to complete this task. Below we introduce the IM-MLU-Net, a hybrid segmentation model based on mutual information and multi-scale.

The fusion method used in this section is different from the two-stage model. We trained MLU-Net and Info-Max-Net synchronously, as shown in Figure 6. Specifically, after the input image passes through the three multi-level attention modules of the MLU-Net encoder, low-resolution deep-level information is obtained to provide the relationship between the segmentation target and its environment. The high-resolution information of the shallow information of the image is directly transferred from the encoder to the decoder of the same height through a jump link, providing more fine-grained features for segmentation. We input the input image and other images in MLU-Net to the mutual information encoder to extract the feature map. The feature map of the input image and the matching feature code extracted by MLU-Net are used as a positive sample for training the discriminator, while the feature map of other images and the feature code of the input image are stitched as a negative sample.

The segmentation loss function of MLU-Net is added to the discriminative loss function of the mutual information discriminator. As the overall loss function of IM-MLU-Net, the entire IM-MLU-Net model is trained by the Adam gradient descent method. The intuitive explanation of IM-MLU-Net is that if the mutual information discriminator can easily determine whether the input image of MLU-Net matches the feature code, it means that the code has strong discrimination to other images. Meanwhile, the code can capture the useful information of the original image and filter out the noise. We combined the deep features that maximize the mutual information with other shallow features extracted from the encoder to predict the category to which the pixel belongs at each scale.

The IM-MLU-Net model combines the advantages of the Info-Max-Net model and the MLU-Net model. By maximizing mutual information, the encoding of the image can capture the deep features in the image via skip connection and multi-scale prediction, leading to good classification performance.

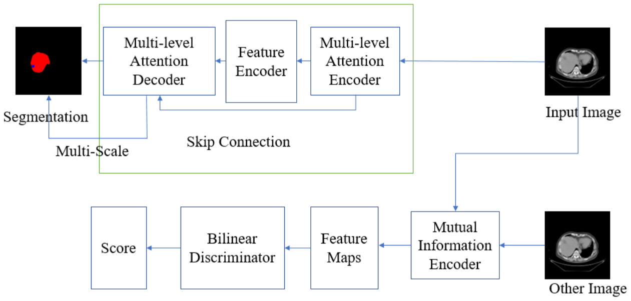


FIGURE 6. The framework of IM-MLU-Net.

V. HYBRID SEGMENTATION MODEL BASED ON GENERATING ADVERSARIAL NETWORKS, MUTUAL INFORMATION, AND MULTI-SCALE INFORMATION

In order to solve the problem that medical image segmentation is lack of medical image data, medical image contains multi-scale information and medical image contains a lot of noise, we propose a hybrid model, called IM-MLU-MedGAN, which combines the advantage of MLU-Net, Info-Max-Net and 3D-MedGAN.

The structure of the IM-MU-MedGAN model is shown in Figure 7. Firstly, in order to solve the shortage of labeled medical data, we used 3D-MedGAN to expand the training set. 3D-MedGAN synthesizes 3D image data with annotation by conditional generative adversarial networks. The essence of 3D-MedGAN training process is the game process between generator and discriminator. In this process, the performance of both generators and discriminators has been improved. With the enhancement of discriminator's ability to distinguish real images from generated ones, generators gradually have the ability to generate images that are highly similar to real ones. After training, we can modify the existing real data such as tumor size, and input the modified label map into the trained generator. The output of the generator accords with the label. We mixed the simulated image generated by 3D-MedGAN generator with the real image of the original training set, and therefore achieve the goal of expanding the training set.

After getting the expanded training set, we use this data set to train IM-MLU-Net and output the segmentation results of the test image through the trained model. During the training phase of the IM-MLU-Net model, we input both real and synthetic medical images into the hybrid segmentation network IM-MLU-Net. The MLU-Net's multi-level attention mechanism module extracts the features and images of the input image. The feature is expanded to the original image size through the deconvolution layer, and the prediction result of the segmentation is finally output.

We compare the segmentation results output by MLU-Net with the manually labeled maps to calculate the segmentation loss function value. The bilinear function discriminator in the IM-MLU-Net network judges that if the images and codes in the input samples match with each other, then high scores are assigned to the corresponding images and codes, otherwise, low scores are assigned to the mismatched pairs. The loss function is added to the segmentation loss of MLU-Net. The Adam gradient descent method is used to train the segmentation network as a whole so that the BL-estimator continuously approaches to the true value of the mutual information. The mutual information between the feature code extracted by MLU-Net and the interaction between the input images reaches the upper bound, and at the same time, the segmentation map predicted by the training image gradually approaches to the actual artificial labeling result. After training, we input the image to be segmented into MLU-Net to get the segmentation result of the medical image.

The loss function used in the experimental training model in this section is explained as follows. Both the generator and discriminator of the 3D-MedGAN model use a least squares loss function. We recorded the input label map of the generator G as z, where z is obeying the distribution $\mathbb{P}_z$, and the simulation image generated by the generator as $G(z)$, where $\mathbb{P}_z$ is initially set as a Gaussian distribution. For obtaining the true distribution of the liver tumor image x that the generator finally learns to be recorded as $\mathbb{P}_{data}$. The following function is adopted.

$$\mathcal{L}_{\text{3D-MedGAN}}(D) = \frac{1}{2}\mathbb{E}_{x\sim\mathbb{P}_{data}(x)}[(D(x)-b)^2] + \frac{1}{2}\mathbb{E}_{z\sim\mathbb{P}_z(z)}[(D(G(z))-a)^2],$$

Among them, a and b are used to mark generated images and real images, respectively. In the experiment, we set a as a matrix with dimensions $w \times h \times d$, where all elements are 1, and set b as a zero matrix with the same size. The loss function of generator G in the 3D-MedGAN model can be expressed as

$$\mathcal{L}_{\text{3D-MedGAN}}(G) = \frac{1}{2}\mathbb{E}_{z\sim\mathbb{P}_z(z)}[(D(G(z))-c)^2],$$

Among them, c is the threshold matrix for G to make D believe that the generated image is a real one. In the experiments in this section, it is set to be the same as a, with dimensions of $w \times h \times d$, and all elements are 1.

MLU-Net uses the Dice loss function. We recorded the number of divided categories (including the background) as L, the number of pixels is N, and p_{ic} means that the i th pixel is predicted to belong to c. Probability of each category g_{ic} is a 0-1 value representing whether the i th pixel actually belongs to the c th category, then the segmentation loss function of MLU-Net can be expressed as

$$\mathcal{L}_{\text{MLU-Net}} = \frac{1}{L}\sum_{c=1}^{L}\left(1 - \frac{2\sum_{i=1}^{N} p_{ic}g_{ic}}{\sum_{i=1}^{N} p_{ic}^2 + \sum_{i=1}^{N} g_{ic}^2 + \epsilon}\right),$$

where ϵ is a small positive number to ensure that the denominator is not zero.

The bilinear function discriminator used to estimate mutual information uses a binary classification entropy loss function with logical sterics ($\mathcal{L}_{\text{BCE}}$), and the number of training samples is M, $T_\theta(x^{(i)}, y^{(i)})$ represents the discriminator's score on

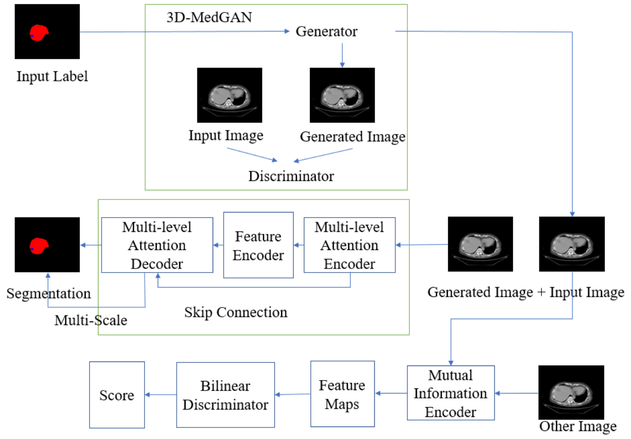


FIGURE 7. The framework of IM-MLU-MedGAN.

the i th sample, and h_i is a 0-1 value, representing whether the i th sample is actually a positive sample. For positive samples it is 1, while for negative samples it is 0. Then the loss function of Info-Max-Net discriminator can be expressed as [28]

$$\mathcal{L}_{\text{Info-Max-Net}}(D) = -\frac{1}{M}\sum_{i=1}^{M}\Big[h_i \log \sigma(T_\theta(x^{(i)}, y^{(i)})) \\ +(1-h_i)\log(1-\sigma(T_\theta(x^{(i)}, y^{(i)})))\Big],$$

Among them, $\delta(z)$ is a Sigmoid function, which can be expressed as

$$\delta(z) = \frac{1}{1+\exp(-z)},$$

The above equation converts the score $T_\theta(x, y)$ from the discriminator to a 0-1 value.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL DATASETS AND PREPROCESSING

To show the effect of the hybrid segmentation model, the dataset of the liver tumor image segmentation on the LIVER100 are used. To ensure that the experiment runs smoothly and the test results are true and accurate, we first perform the following preprocessing steps on each 3D image in the LIVER100 dataset:

1) Adjusting the size of the liver tumor slice to 256*256 to ensure that the pooling operation of the MLU-Net model can be applied to layers with the same size on the horizontal and vertical axes, thereby seamlessly stitching the segmentation result map;
2) Counting the gray values of all images and truncating the part of the gray range outside 0.5%-99.5% for the purpose of eliminating outliers in the image;
3) Interpolating the labeled map of the liver tumor by nearest-neighbor algorithm, that is, the gray value of the pixel closest to the position of the pixel to be valued is taken as its gray value;
4) Finding the start and end slices of the liver region, expand outward in two directions, and randomly extracting from the 48 slices which contain the liver as the input to the hybrid segmentation model. If the number of slices containing liver is less than 48, then these data are directly discarded. In fact, it rarely happens that three-dimensional images with less than 48 slices of liver are encounted in the experiments.

B. EXPERIMENTAL SETTINGS

The liver tumor images in the LIVER100 dataset are used and they are divided into three categories: background, liver, and tumor. Similar to the previous experiments of the MLU-Net model and the Info-Max-Net model, the experimental environment of the experiments in this section is conducted in Linux with TensorFlow 1.14.0 and Python 3.7.4.

The GPU model is GeForce GTX 1080Ti, and the test speed is 6.29 seconds per image. To maximize the use of GPU memory, we tried to use larger input tiles, so set the batch size to 1, which was a single image. We used the Adam optimization algorithm to iteratively update the Info-Max-Net network weights. The initial learning step size is 10^{-4}. After training 1500 batches, the loss function value of the test set decreases slowly and the step size is reduced. Thus, the learning steps size is adjusted to one-tenth of the original, which is 10^{-5}.

C. EXPERIMENTAL RESULTS

In this section, we will show the results of image segmentation on the LIVER100 dataset based on a hybrid segmentation model IM-MLU-MedGAN.

First, we augment the training data set with 3D-MedGAN. Then, we mixed the generated simulation images with 80% images in the LIVER100 dataset as the training set for segmentation network; the remaining 20% images of liver tumors in the LIVER100 dataset were used as the test set for this experiment. The training curve of the IM-MLU-MedGAN model's loss function during training is shown in Figure 8. It can be seen from the Figure 8 that the loss function value had dropped to less than 2 when completing a few batches of training, and the amplitude of the loss function curve is stable after 1500 batches. In the end, after a total of 3,000 batches of training, the value of the loss function stopped falling, and the model converged.

We used the dice score to quantitatively evaluate the segmentation accuracy of IM-MLU-MedGAN. The possible value of the dice score of the liver or tumor is between 0-1. The higher the degree of coincidence between the predicted result and the actual label, the larger the dice score will be. The average liver dice score of IM-MLU-MedGAN on the entire test set was 0.7913, and the average tumor dice score was 0.3902.

In order to verify the effectiveness of the hybrid segmentation model, we compared the segmentation results predicted by the IM-MLU-MedGAN model with some other recently proposed models. Each model is based on the

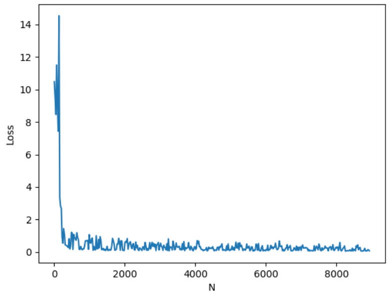


FIGURE 8. MLU-MedGAN model training curve.

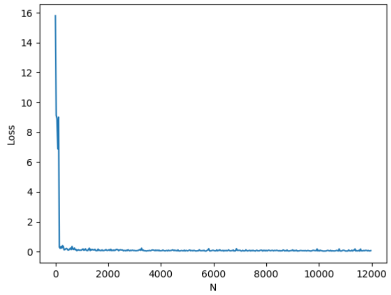


FIGURE 9. IM-MLU-Net model training curve.

TABLE 3. IM-MLU-MedGAN hybrid model and other models for liver segmentation and dice score of tumor detection on LIVER100 dataset.

Model Name	Tumor dice score	Liver dice score
U-Net	0.2899	0.7003
U-Net++	0.2782	0.7309
DialResNet	0.3245	0.7247
AgNet	0.2814	0.7112
RA-UNet	0.2902	0.6955
Info-Max-Net	0.3499	0.6522
MLU-Net	0.3587	0.7323
IM-MLU-MedGAN	0.3902	0.7913

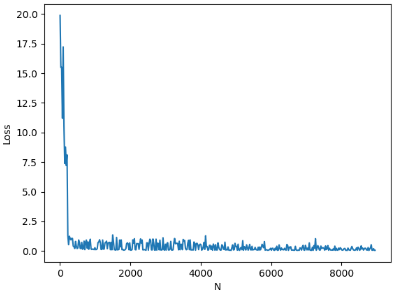


FIGURE 10. IM-MedGAN model training curve.

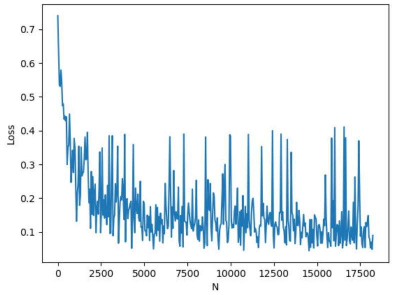


FIGURE 11. MLU-MedGAN model training curve.

LIVER100 dataset. The tested dice scores are shown in the Table 3, where the parameters of each model have been adjusted to the optimum through grid search. The tumor dice score of the IM-MLU-MedGAN model is better than U-Net, U-Net++, DialResNet, AgNet, RA-UNet, Info-Max-Net, MLU-Net, increased by 34.6%, 40.3%, 20.2%, 38.7%, 34.5%, 11.5%, 8.8% respectively, an average improvement of 33.7% over the current popular image segmentation models. IM-MLU-MedGAN model's liver dice score is higher than these seven models by 13.0%, 8.3%, 9.2%, 11.3%, 13.8%, 21.3%, 8.1%, which is 11.12% on average. From this we can see that the usage of 3D-MedGAN to supplement the data, combined with using multi-scale information, attention mechanism and mutual information maximization can effectively improve the quality of feature extraction and the medical image segmentation.

D. ABLATION EXPERIMENT

The IM-MLU-MedGAN hybrid segmentation model is mainly composed of 3D-MedGAN, Info-Max-Net, and MLU-Net. In order to explore the impact of these three parts on the hybrid model, we combine these three parts in pairs to get the IM-MLU-Net model, IM-MedGAN model and MLU-MedGAN model. Segmentation results of medical images by combining the Info-Max-Net model and MLU-Net model, as well as three mixed models of IM-MLU-Net, IM-MedGAN and MLU-MedGAN, with our final IM-MLU-MedGAN model are compared. Through comparison, we can evaluate the impact of each part of the IM-MLU-MedGAN hybrid segmentation model on the segmentation effect.

Figure 8, 9, 10 and 11 show the loss curve of IM-MLU-MedGAN, IM-MLU-Net model, IM-MedGAN model and MLU-MedGAN model during training. It can be seen from these figures that the loss function of the IM-MLU-Net model and the IM-MedGAN model are relatively similar. The loss is higher at the beginning, then decreases rapidly, and gradually stabilizes. In the MLU-MedGAN model, the value of the loss function is not high at the beginning, but the loss oscillation is more serious afterward.

The dice scores of each independent model and their mixed models for liver segmentation and tumor detection on the LIVER100 dataset are shown in Table 4. Based on the independent model, the 3D-MedGAN model is introduced to expand the training data set, which can improve the performance of the model on both liver segmentation and tumor detection tasks. The IM-MedGAN hybrid model improves the tumor dice score compared to the Info-Max-Net model alone. It was increased by 8.9%, the liver dice score was increased by 19.8%, and the MLU-MedGAN model was increased by 3.6% compared with the tumor and liver dice scores of the MLU-Net model alone. It can be seen that the improvement effect of 3D-MedGAN on Info-Max-Net is more significant than that on MLU-Net. We analyze that the reason is that Info-Max-Net uses U-Net [24] for feature

TABLE 4. The independent models and their mixed models proposed in this paper perform liver segmentation and dice score of tumor detection on the LIVER100 dataset.

Model name	Tumor dice score	Liver dice score
Info-Max-Net	0.3499	0.6522
MLU-Net	0.3587	0.7323
IM-MLU-Net	0.3818	0.7849
IM-MedGAN	0.3811	0.7811
MLU-MedGAN	0.3717	0.7590
IM-MLU-MedGAN	0.3902	0.7913

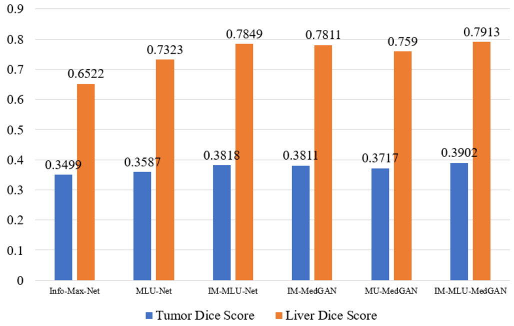


FIGURE 12. Dice scores for liver segmentation and tumor detection on the LIVER100 dataset with independent models and their hybrid models.

extraction and image segmentation, while MLU-Net Then, a multi-level attention mechanism is introduced based on the classic U-Net, so the features of different levels of the image can be more efficiently learned, and thus the sensitivity to the size of the training data set is not as strong as the Info-Max-Net model.

Among the two-part hybrid segmentation model, the IM-MLU-Net model has the highest segmentation accuracy. It is very close to the dice score of the segmentation result of the IM-MLU-MedGAN model. This shows that Info-Max-Net optimizes deep-level features. It is very suitable for combining with MLU-Net's utilization of multi-scale features. MLU-MedGAN has the lowest dice score in the hybrid segmentation model. From the loss function curve of our model, we can observe that our model have been trained well and stable. Combined with the analysis of the loss function curve, we believe that the main reason is that the loss oscillation of the MLU-MedGAN model is serious in the later period, which makes the model's convergence effect poor, so the model's effect cannot be achieved and cannot outperforms other hybrid models.

The dice score of the IM-MLU-MedGAN model is the highest of all the models we have tested, and it's tumor dice score is higher than the IM-MLU-Net model, IM-MedGAN model, and MLU-MedGAN model by 2.2%, 2.4% and 5.0%, with liver dice score increased by 0.8%, 1.3%, 4.3%. The above data shows that combining any of our proposed 3D-MedGAN, Info-Max-Net, and MLU-Net models with other models has a positive effect on improving the accuracy of medical image segmentation. The most obvious improvement is the introduction of Info-Max-Net to estimate and maximize the mutual information between the image and the feature code. The IM-MLU-MedGAN model has a tumor dice score of 11.5% higher than the Info-Max-Net model alone, and a liver dice score of 21.3% higher than the tumor dice score of the MLU-Net model alone. The score is increased by 8.0%. Experimental data show that the improvement of image segmentation models cannot be limited to only one aspect. It can only alleviate the lack of training data at the same time, improve the feature extraction efficiency of the convolution layer, use the multi-scale features of medical images, and consider the statistical significance of feature coding, to improve the effect of image segmentation.

VII. CONCLUSIONS

In this paper, we proposed 3D-MedGAN, MLU-Net and Info-Max-Net. The 3D-MedGAN can enrich the training data set by generating fake images via adversarial learning, improving the segmentation performance greatly. Meanwhile we use the discriminator Semi-GAN to make the training process more stable. MLU-Net introduces a multi-level block to extraction different level features of the medical image via attention mechanism. Info-Max-Net enables the encoder of image to capture the deep features in medical images that are beneficial to classification and ignores the noises in the medical images. Then, we proposed a hybrid model, called IM-MLU-MedGAN, to effectively alleviate the lack of labeled data and feature extraction in the field of medical image processing by integrating the 3D-MedGAN, the Info-Max-Net and the MLU-Net models. We conduct experiments on the LIVER 100 dataset, and from experimental results we can see that IM-MLU-MedGAN model can achieve better performance than Info-Max-Net, MLU-Net, and other popular deep learning models on medical image segmentation tasks, which verified the effectiveness of IM-MLU-MedGAN.

REFERENCES

[1] T. Zuva and O. O. Olugbara, "Image segmentation, available techniques, developments and open issues," *Can. J. Image Process. Comput. Vis.*, vol. 2, no. 3, pp. 20–29, 2011.

[2] R. Frank, T. Grabowski, and H. Damasio, "Voxelvise percentage tissue segmentation of human brain magnetic resonance images," in *Proc. Abstr., 25th Annu. Meeting, Soc. Neuro-Sci.*, 1995, p. 694.

[3] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[4] A. Bert, I. Dmitriev, S. Agliozzo, N. Pietrosemoli, M. Mandelkern, T. Gallo, and D. Regge, "An automatic method for colon segmentation in CT colonography," *Comput. Med. Imag. Graph.*, vol. 33, no. 4, pp. 325–331, Jun. 2009.

[5] N. Sharma and A. K. Ray, "Computer aided segmentation of medical images based on hybridized approach of edge and region based techniques," in *Proc. Int. Conf. Math. Biol.*, 2006, pp. 150–155.

[6] L.-K. Huang and M.-J.-J. Wang, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recognit.*, vol. 28, no. 1, pp. 41–51, Jan. 1995.

[7] R. Korez, B. Likar, F. Pernuš, and T. Vrtovec, "Model-based segmentation of vertebral bodies from MR images with 3D CNNs," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2016, pp. 433–441.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[9] N.-Q. Nguyen and S.-W. Lee, "Robust boundary segmentation in medical images using a consecutive deep encoder-decoder network," *IEEE Access*, vol. 7, pp. 33795–33808, 2019.

[10] R. Zhao, W. Chen, and G. Cao, "Edge-boosted U-Net for 2D medical image segmentation," *IEEE Access*, vol. 7, pp. 171214–171222, 2019.

[11] 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. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[12] Ö. A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 9901, Oct. 2016, pp. 424–432. [Online]. Available: http://arxiv.org/abs/1606.06650

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[14] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*. [Online]. Available: http://arxiv.org/abs/1607.08022

[15] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[16] X. Zhang, Y. Wei, G. Kang, Y. Yang, and T. Huang, "Self-produced guidance for weakly-supervised object localization," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 597–613.

[17] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2107–2116.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[19] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, May 2019, pp. 1–24.

[20] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA, Oct. 2016, pp. 565–571.

[21] I. R. S. Belghazi, A. Baratin, R. D. Hjelm, and A. C. Courville, "MINE: Mutual information neural estimation," *CoRR*, vol. abs/1801.04062, pp. 1–18, Jun. 2018.

[22] Ö. A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Proc. MICCAI*, 2016, pp. 424–432.

[23] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*. Cham, Switzerland: Springer, 2018, pp. 178–190.

[24] 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.*, 2015, pp. 234–241.

[25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, 2018, pp. 3–11.

[26] O. Oktay, J. Schlemper, L. Le 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

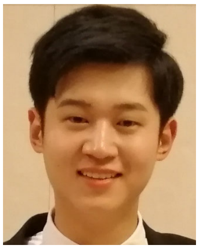
[27] Q. Jin, Z. Meng, C. Sun, L. Wei, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," 2018, *arXiv:1811.01328*. [Online]. Available: http://arxiv.org/abs/1811.01328

[28] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*. [Online]. Available: http://arxiv.org/abs/1807.03748

[29] G. Zhang, S. Dong, H. Xu, H. Zhang, Y. Wu, Y. Zhang, X. Xi, and Y. Yin, "Correction learning for medical image segmentation," *IEEE Access*, vol. 7, pp. 143597–143607, 2019.

YI SUN received the master's degree in software engineering from Fudan University, in 2009, where he is currently pursuing the Ph.D. degree. He is also a Senior Engineer at the Information Office of Fudan University. His research interests include medical image analysis and adversarial attack in deep learning.

PEISEN YUAN (Member, IEEE) received the Ph.D. degree from Fudan University, in 2011, and the M.S. degree from the Nanjing University of Aeronautics and Astronautics, in 2007. He is currently an Assistant Professor at the College of Information Science and Technology, Nanjing Agricultural University. His research interests include machine learning, optimization, and big data processing technologies.



YUMING SUN is currently a Graduate Student at the School of Data Science, Fudan University. His research interests include medical image analysis and adversarial attack in deep learning.

ZHAOYU ZHAI received the M.S. degree in systems and services engineering from the Information Society, Technical University of Madrid, in 2018, where he is currently pursuing the Ph.D. degree in systems and services engineering. His research interests include decision support system and big data analysis.