Nucleus Image Segmentation Method Based on GAN Network and FCN Model

Kai Zhang

Changchun University of Finance and Economics

Yang Shi

Changchun University of Finance and Economics

Chengquan Hu

Jilin University

Hang Yu (✉ ccufe0123@163.com)

Jilin Agricultural University

Research Article

Keywords: Nucleus image, Image segmentation technology, GAN network, FCN model, Fully connected layer convolution, Pixel loss

Posted Date: September 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-858913/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Nucleus Image Segmentation Method Based on GAN Network and FCN Model

KaiZhang¹, Yang Shi¹, Chengquan Hu², Hang Yu^{3*}

¹ School of Information Engineering, Changchun University of Finance and Economics, Changchun, 130122 China

² School of Computing, Jilin University, Changchun, 130022 China

³School of Information Technology, Jilin Agricultural University, Changchun, 130022 China

*.Corresponding Author: HangYu

Email: ccufe0123@163.com

Abstract: Aiming at the problems of rough edges and low accuracy in processing cell nucleus image segmentation in existing image segmentation methods. A cell nucleus image segmentation technology based on generative adversarial network (GAN) network and fully convolutional network (FCN) model is proposed. First, the FCN model is used to perform preliminary segmentation of the cell nucleus image, in which the fully connected layer convolution and skip connection are used to improve the accuracy of image segmentation. Then, improve the GAN network, introduce splitting branches into the discriminator structure, and combine the GAN network and the splitting network into one. At the same time, pixel loss is introduced in the generator to obtain a nucleus image that is visually more similar to the real image. Finally, the segmented image output by the FCN model is used as the input of the GAN network to achieve high-precision segmentation of the nucleus image. The proposed method is experimentally demonstrated based on the 2018 data science bowl dataset. The results show that it can achieve rapid convergence, and the mean intersection over union (MIoU) is 85.34%, which is better than other comparison methods.

Keywords: Nucleus image; Image segmentation technology; GAN network; FCN model; Fully connected layer convolution; Pixel loss

1Introduction

With the accumulation of medical imaging images, experts and scholars began to perform detailed analysis and processing on massive medical imaging images. Through various algorithms, the association between some diseases and their medical imaging images is obtained ^[1]. At present, image generation technology has been widely used in medical institutions, and image processing-related technologies have become more and more important. It plays a very crucial role in improving the efficiency of medical diagnosis ^[2]. Therefore, fully mining the potential value of medical imaging images and improving the efficiency of clinical applications have become the key to the research of academia and medical institutions ^{[3][4]}. As the first step in the diagnosis and treatment of some diseases, medical image segmentation generally refers to the extraction of certain target areas in the entire image in some way, such as nucleus, certain organs or tissues ^[5]. The results of medical image segmentation usually have no intersection between each area, and each segmented area has a certain similarity in its interior. Therefore, the segmentation process needs to eliminate external interference factors to ensure the higher accuracy and reliability of the segmentation results ^[6].

At present, the segmentation methods for cell nucleus images at home and abroad generally include matched filter method, multi-threshold detection, image segmentation based on morphology, region growing method, neural network segmentation algorithm, multi-scale layer decomposition and local adaptive threshold segmentation Method, segmentation method based on active contour model and segmentation method based on fuzzy clustering ^{[7][8][9]}. The typical method is relatively simple and easy to implement, and has been widely used because of its high segmentation efficiency. Reference [10] proposed a multi-threshold differential evolution solution for image segmentation. The different efficiencies of the allocated difference algorithm can be evaluated by measuring the quality of the candidate solutions, so as to generate the optimal solution for the population. Reference [11] proposes a new type of active contour model that combines regional information and image edge information to achieve image segmentation. And the divergence operator is used to balance the information of all aspects of the image, which enhances the adaptability of the model. Reference [12] proposed an automatic method for detecting exudate in digital fundus images. It mainly includes four steps: color shift correction, disc elimination, exudate segmentation, and separation of exudate and background. It has high segmentation accuracy, but the segmentation efficiency needs to be improved. Reference [13] uses the space-restricted rough fuzzy C-means algorithm to segment medical images. This method combines the advantages of rough fuzzy clustering and non-global neighboring region information, and divides each cluster into possible lower approximations or core regions and probability boundary regions. It effectively improves the accuracy of segmentation, but requires higher classification standards, and the algorithm initialization process is more complicated.

With the rapid development of deep learning networks in recent years, it has begun to be applied in image segmentation and gradually promoted and improved ^{[14][15]}. Reference [16] proposed an image segmentation method based on the Convolutional Neural

Network (CNN) loss function, which directly reduces the Hausdorff distance and improves the segmentation accuracy. Reference[17] proposed an image super-resolution method using progressive generation of confrontation networks. Convert a lower-resolution image into a higher-resolution image that can be dynamically scaled. By using the triplet loss function, the output image quality of the next stage is gradually improved, but the algorithm segmentation efficiency needs to be improved. Reference[18] combines the advantages of residual blocks and evolutionary algorithms to design a segmentation network suitable for medical image processing. The network parameters automatically evolve and have high segmentation accuracy, but the algorithm does not consider the complexity of the nuclear image, and its feasibility and effectiveness need to be further verified. Reference[19] proposed an interactive segmentation method based on deep learning to improve the results obtained by automatic CNN. And reduce user interaction in the refining process, and achieve a more ideal segmentation effect.

There have been a lot of research results on medical image segmentation. But most of them are for larger targets such as organs. For images with relatively small targets such as nuclei, the edges are rough and fuzzy, and the segmentation effect of related segmentation techniques is poor [20]. For this reason, a cell nucleus image segmentation technology based on GAN network and FCN model is proposed. Its innovations are summarized as follows:

(1) Because traditional CNN has problems such as loss of image spatial information, easy over-fitting and difficult training, FCN is used for image segmentation. Among them, fully connected layer convolution and skip connection are used to improve the accuracy of image segmentation.

(2) In order to better ensure the accuracy of image segmentation, the proposed method optimizes the GAN network. Among them, the segmentation branch is introduced into the discriminator, and the segmentation loss fed back by the segmentation branch and the confrontation loss fed back by the discriminant branch are used to guide the generator to generate a more realistic image. At the same time, pixel loss is introduced in the generator, so that the generator can generate a nucleus image that is visually more similar to the real image.

This paper is divided into five sections: The first section of the introduction first briefly introduces the background and current research status. The second section introduces related theories and methods, and establishes the required technical model. The third section introduces the generator and discriminator network structure in detail. The experimental results and analysis in the fourth section show that the proposed method has good accuracy and reliability. Finally, the method of this paper is summarized.

2 Related theories and methods

2.1 VGGNet

VGGNet was proposed by researchers in the Visual Geometry Group (VGG) of the University of Oxford and named after it. The improvement of VGGNet lies in the use of several consecutive small convolution kernels instead of the larger ones in AlexNet, thereby including more activation function layers, providing higher efficiency and reducing the amount of trainable parameters. Under the premise of ensuring that the receptive field is not affected, the depth of the network is deepened, thereby improving the processing ability of the neural network [21]. The first five layers of the VGG16 network are used as the coding structure in the proposed segmentation network structure, including convolutional layers, activation functions and other parts. And replace the pooling layer with a convolutional layer with the same function. The convolutional layer is very important in the convolutional neural network, which is used to complete the feature extraction and transfer of the input image. The output size of any given convolutional layer is calculated as follows:

$$O = \left\lfloor \frac{I - K + 2P}{S} \right\rfloor + 1 \quad (1)$$

In the formula, O is the output feature size, I is the input feature size, K is the size of the convolution kernel, P is the "padding" that is the number of pixels to be filled, and S represents the stride. In the proposed network, the filling method in the ordinary convolutional layer is "SAME". That is, the size of the feature map after the convolutional layer processing remains unchanged, but there is no padding in the 2×2 convolutional layer that replaces the pooling layer, and the step size is 2, making the size of the feature map output by the convolutional layer It is 1/2 of the input feature map. This achieves the same function as the pooling layer downsampling.

The parameter of the convolutional layer can be regarded as the weight of the corresponding position pixel in the image. Assigning the weighted value of each pixel in the convolution kernel to the center pixel is the process of abstracting image features. The process can be expressed as:

$$y = \sigma(wx + b) \quad (2)$$

In the formula, x and y represent the input feature and output feature, respectively. w represents the convolutional layer parameter, b represents the offset, and σ is the activation function.

The function of the activation function is to transform the input features non-linearly. The Rectified Linear Unit (ReLU) activation function is used in VGGNet. Its mathematical description is:

$$\sigma(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

The ReLU activation function is non-linear, available in the backpropagation algorithm, and it does not activate all neurons at the same time. This makes the network sparse and improves computational efficiency.

2.2 Fully Convolutional Network

Although segmentation methods based on candidate regions have brought great progress to the development of semantic segmentation, the candidate regions lack spatial information. Especially for small objects, the information loss is serious, which directly affects the final segmentation effect. On this basis, a semantic segmentation model based on FCN came into being. It does not need to generate candidate regions, can input images of any size, and can directly realize end-to-end pixel-level prediction^{[22][23]}. The structure of the image segmentation model based on FCN is shown in Fig. 1.

Fig.1 Model structure of FCN

The FCN model does not need to modify the structure parameter scale in the network for pictures of various sizes, so that it can accept picture input of any size. And bilinear interpolation or transposed convolution is used to upsample the final convolutional layer. In this way, the output image will be the same size as the original input image. Finally, use Softmax to normalize the obtained output one by one. Find out the category corresponding to the maximum probability value, that is, get the category to which all pixels belong. So as to realize the end-to-end, pixel-to-pixel image segmentation^[24].

2.2.1 Fully connected layer convolution

Traditional CNN used for image classification usually needs to connect several fully connected layers at the end. The final output is converted into a vector, and normalized for classification^[25]. But for images of different sizes, the size of the fully connected layer parameter matrix must be redesigned, and the spatial information of the input image is lost. In addition, the fully connected layer makes the network too many parameters, easy to overfit and difficult to train. Therefore, it is necessary to remove the fully connected layer and add more convolutional layers after the original convolutional layer. In the optimized network, all the full connection layers are replaced by the convolutional layer to obtain the full convolutional network

As shown in Fig. 1, the fully connected layer is replaced by a convolutional layer with a 3-layer convolution kernel size of 1×1, so that the amount of network parameters can be minimized. Replace the last three fully connected layers with a three-layer convolution kernel with a size of 1×1, which can greatly reduce the parameter scale of the network. Because the convolution kernel is only connected to a certain part of the previous feature map and the convolution usually shares parameters during operation^[26]. On the other hand, when the fully connected layer is connected to the convolutional layer, the feature map of the convolutional layer will be flattened and converted into a vector before the operation is performed. In this way, the spatial information and relative position information contained in the feature map are lost. In the image classification task, since only the category in the image needs to be determined, this information can usually be ignored. However, it is very necessary for image segmentation that requires all pixel types to be output. All in all, the fully connected layer convolution is to replace the fully connected layer in a typical CNN with a convolutional layer with a convolution kernel size of 1×1.

2.2.2 Skip architecture

The segmentation results of the input images can be obtained by carrying out the full connection layer convolution of CNN and using a large magnification to upsample the end feature image. But every pooling operation will down-sample the feature map, and then lose some information. The feature map obtained after multiple pooling operations has lost too much information. After upsampling it, it will lose too much information relative to the original picture, which will result in lower accuracy of the image segmentation result. Especially the edge segmentation results are poor^[27]. The skip structure can better solve this problem and improve the accuracy and precision of the final output medical image segmentation results. The edge information is more complete than before. Its structure is shown in Fig. 2.

Fig. 2 Skip architecture

The skip structure combines the results of upsampling the feature maps corresponding to different levels of pooling layers. The model of adding each point one by one is used to fuse the higher level non-local features and the lower level local features. To a certain extent, it makes up for the loss of information. Make the final output image segmentation result higher accuracy and more perfect edge information.

2.3 Convolution

The convolutional layer is an important part of the FCN model, which limits the local receptive field of the node to the agreed category to obtain the non-global characteristics of the image and other information. Higher-level abstract features can be automatically extracted from the feature map through learning. The convolutional layer structure is shown in Fig. 3.

Fig.3 The structure of convolution layer

The shallow convolutional layer is extracted from pixel-level features, often extracting low-level image features. Such as the corners, textures, edges and lines of the image. These low-level image features will be non-linearly transformed by the deeper convolutional layer using the convolution kernel to extract higher-level and more complex features, such as shapes, contours, etc. [28]. When initializing network parameters, a small random number is mainly used to initialize the weight of each convolution kernel in the convolution layer. The input between layers is the original image pixel matrix and the feature map formed after convolution or pooling.

The calculation of the convolutional layer is to use a fixed-size convolution kernel with a preset size, and slide on the feature map output by the previous layer with a preset step size. After the inner product operation is performed, other nonlinear operations are performed. In this way, the feature image of the previous level is transformed into the feature image of the next level. The convolutional layer needs to manually set some hyperparameters in advance, and the size of the feature image is affected by multiple parameters such as the size of the convolution kernel and the sliding step length. The number of output channels of the feature image depends on the number of convolution kernels [29].

Different from typical neural networks, FCN has local receptive field and weight sharing. Each convolution kernel usually only performs local operations with the feature map of the upper layer. The number of output feature map channels is generally the same as the number of convolution kernels [30]. Usually the size of the convolution kernel determines the range that it can feel, and neurons usually share parameters. This significantly reduces the scale of the convolutional layer parameters.

2.4 Pooling layer

The pooling layer is also often referred to as the down-sampling layer. After the image has undergone sliding convolution of the convolutional layer, high-latitude features with local relevance are obtained. If all the extracted features are used for prediction, not

only is it easy to overfit, but also the amount of calculation will be relatively large. In order to cope with this difficulty and achieve the goal of representing large-size feature maps with small-size feature maps, down-sampling processing can be performed on the obtained feature maps. This process is called pooling, as shown in Fig. 4.

Fig. 4 Pooling operation

The specific implementation of the pooling operation is to use a convolution kernel with a size of 2×2 and two sliding steps for sliding, and select the maximum value of elements in any rectangle or the average value of all elements. The original 4×4 size feature map is transformed into a 2×2 size feature map through this operation, and the elements on the four feature maps become one element. Compared with the feature dimensionality obtained by convolution, the dimensionality of the feature after pooling will be much lower. And it can significantly suppress the over-fitting of the model, thereby enhancing the generalization ability of the model and reducing the computational complexity. Generally speaking, pooling includes two types: maximum pooling and average pooling. After the pooling process, the model can have a certain degree of translation invariance to a certain extent.

2.5 Deconvolution

After the network after the fully connected layer convolution transformation is input with a picture of a certain size, it will perform a number of processing. The pooling processing and even some convolution processing will down-sample the feature map of the upper layer. After each pooling process, the output feature map will be reduced, and the resolution will also decrease. In this way, compared to the original input image, the final network output image will be much smaller. However, image segmentation tasks need to be classified at the pixel level, and end-to-end, pixel-to-pixel. The final output image size needs to be consistent with the input image. Therefore, it is necessary to increase the up-sampling operation to enlarge the characteristic image to the size of the original input image.

The most commonly used method of upsampling is deconvolution, which is usually called upconvolution or transposed convolution. Deconvolution and convolution are similar. The difference is that the former is a value operation on a feature map to generate multiple values, and the latter is a value operation on multiple feature maps to generate a value. Deconvolution was proposed by Zeiler et al. in 2010, and its arithmetic operation is the inverse operation of the convolution operation. Therefore, the multiplicative convolution of the deconvolution can also be regarded as the permutation matrix of the convolution [31][32]. The deconvolution process is shown in Fig. 5.

Fig.5 The structure of deconvolution

Deconvolution can increase the resolution of the feature map and expand the receptive field. When initializing the deconvolution kernel, bilinear interpolation is usually used to facilitate the rapid and effective convergence of the deconvolution kernel parameters. The well-trained deconvolution kernel can well complete the up-sampling of the feature map.

2.6 Generative Adversarial Network

Generative Adversarial Networks (GAN) includes generators and discriminators. As a reverse CNN model, the generator implements up-sampling through the corresponding deconvolution layer, and uses the excitation layer to convert the low-latitude metric into a vector that is the same as the actual image dimension for output [33]. Among them, the randomly generated Gaussian white noise is the input of the generator, through its network decoding, the output vector is consistent with the actual image size. The discriminator is a simplified CNN model whose input is the actual image and the image generated by the generator. After correspondingly extracting the characteristics of the input image, it finally outputs the probability value in the range of [0,1]. And through the confidence value predicted by the discriminator, the error between the tags corresponding to the true type is calculated. Directly treat this error as a back-propagation error, dynamically adjust the parameters and the original input. GAN uses the

cross-entropy loss function to implement parameter optimization, and the mathematical description is as follows:

$$\min_G \max_D L(D, G) = E_{x \sim \varphi_{data}(x)} \{1b[D(x)]\} + E_{x \sim \varphi_{G(z)}} \{1b\{1 - D[G(z)]\}\} \quad (4)$$

Where, $\min_G \max_D V(D, G)$ is the loss function, L is the loss function, E is the mean operation, D is the discriminator model, G is the generator model, φ is the vector distribution type, x is the real image, z is the random noise of input.

The function of the discriminator is to determine whether the input image is the actual image or the generated image of the generator. When the input is a real image, the value of $D(x)$ approaches 1; When the input is an image generated by the generator, the value of $D(x)$ approaches 0. The generator needs to adjust the distribution of z as much as possible to minimize the difference between the distribution of $G(z)$ and x , and make $D[G(z)]$ approach 1. Therefore, GAN is composed of two models: a generator and a discriminator, and the proposed training mode can be gradually crossed. The training loss function of the discriminator is:

$$\max_D L(D, G) = E_{x \sim \varphi_{data}(x)} \{1b[D(x)]\} + E_{x \sim \varphi_{G(z)}} \{1b\{1 - D[G(z)]\}\} \quad (5)$$

If the above formula is satisfied, it is considered that the performance of the generator has reached the optimal situation, and only the classification performance of the discriminator needs to be trained. When training the loss function of the generator, it is considered that the performance of the discriminator has reached the optimal situation, that is, $E_{x \sim \varphi_{data}(x)} \{1b[D(x)]\}$ is always a fixed constant and no training is required. Therefore, the loss function when training the generator is:

$$\max_G L(D, G) = E_{x \sim \varphi_{G(z)}} \{1b\{1 - D[G(z)]\}\} \quad (6)$$

3Method

3.1Generator network structure

In dealing with the problem of cell nucleus image generation, the complexity of the model is often increased by adding a residual structure to the generation network, while avoiding the problem of poor generalization ability caused by over-fitting. Improve the generation ability and generalization ability of the network, so that the learning ability of the model is greatly improved. In order to further improve the generator's performance in generating cell nucleus images, the deep convolution can be fitted to obtain the output characteristics of the shallow convolution. It also combines low-level features and high-level features to form a multi-layer residual fusion model while preventing the problem of gradient disappearance in the network. The proposed method expands the 16 residual blocks of the generator network into 32. The comparison diagram of the generator network structure is shown in Fig. 6. The network consists of a convolution layer, an activation layer, a normalization layer, and a deconvolution layer. The gray structure in the figure is the residual block, and the dark gray square corresponds to the internal structure of a single residual block.

Fig.6Generator network structure comparison chart

3.2Discriminator network structure

Different from the traditional discriminator, the discriminator model consists of two branches, namely the discriminant branch and the image segmentation branch. The structure of the discriminator is shown in Fig. 7. It takes the feature map corresponding to the real cell nucleus image and the feature map generated by the generator as input, after feature extraction and classification, and

finally output.

Fig. 7 Discriminator structure

Among them, the discrimination branch has the same function as the traditional discriminator, which is used to discriminate whether the input is a real image or an image generated by a generator. It is mainly composed of a convolutional layer, an activation layer, and a normalization layer. The general structure of the image segmentation branch consists of a dense block structure, a transition down structure, a convolutional layer, and a deconvolutional layer. It has two functions, namely, generating the image segmentation result of the cell nucleus and obtaining the segmentation loss. The arrow of the segmentation branch indicates that the output of the layer where the arrow start point is located and the output of the layer where the arrow end point is located are concatenated. After splicing, a feature map with a thicker structure is formed, and the combined feature map is used as the input of the next layer.

3.2.1 Differential discriminant branch

In the traditional discriminator structure, the discriminator model separately learns the real cell nucleus image and the generated cell nucleus image. And according to the corresponding label to correct the learning results of the two images. In this way, the internal connection between the two images is ignored, and the function of the discriminator cannot be maximized. In order to better explore the potential correlation between the real image distribution and the generated image distribution, the network is guided to generate more realistic images, and the difference discrimination branch is introduced, so that the discriminator can distinguish between the two images without splitting. It is to distinguish the difference between the two. The difference between the real cell nucleus image and the generated cell nucleus image can be expressed as:

$$\psi[x, G(z)] = \left| \phi(x) - E\{\phi[G(z)]\} \right|_{L_1} \quad (7)$$

Where, $\phi(x)$ is the feature extracted from the real cell nucleus image after passing through the discriminator, $\phi[G(z)]$ is the feature extracted from the generated cell nucleus image after passing through the discriminator.

The output of the traditional discriminator is mapped to the probability value of the $[0, 1]$ interval by the activation function $D = \sigma[\phi(\cdot)]$, where σ is the activation function. For a real nucleus image discriminator, its output is $D\{\psi[x, G(z)]\}$, and the corresponding generated nuclear image discriminator output is $D\{\psi[G(z), x]\}$. After activation function mapping, the discriminator output at this time is still the probability value of $[0, 1]$. Consistent with the typical discriminator, the corresponding label of the true image relative to the generated image is 1, otherwise it is 0. At this time, the loss L_a of the GAN network is defined as:

$$L_a = E_{x \sim \mathcal{P}_{data}(x)} \left\{ 1b \left\{ D\{\psi[x, G(z)]\} \right\} \right\} + E_{x \sim \mathcal{P}_{G(z)}} \left\{ 1b \left\{ 1 - D\{\psi[G(z), x]\} \right\} \right\} \quad (8)$$

3.2.2 Image segmentation branch

First, the cell nucleus image is generated by the generative model, and the image and its corresponding segmentation label map are sent to the segmentation branch together. The corresponding image is then generated. The process of calculating the loss through a given segmentation label map can be expressed as:

$$L_b = \sum_{ij} [\beta_{ij} \text{lb} \alpha_{ij} + (1 - \beta_{ij}) \text{lb} (1 - \alpha_{ij})] \quad (9)$$

Where, L_b is the segmentation loss, α_{ij} is the category of the pixels corresponding to positions i and j in the network prediction map. β_{ij} is the category of the pixel corresponding to the i and j positions of the label image. The loss is passed back to the generator, and the generator adjusts the model hyperparameters according to the returned gradient information and refines the generated image. It causes the generator to generate detailed information that is easier to implement segmentation branch detection.

Only use adversarial loss and segmentation loss to adjust GAN, there will often be a problem that although the network loss is very low, the result is not good. Because both the adversarial loss and the segmentation loss are obtained from the perspective of image feature analysis to obtain the optimized gradient information used to generate the image. For the model, the equivalent pixel distribution can be obtained. Due to the different values of different positions, the effect of image presentation varies greatly. Therefore, in order to make up for the visual imperfections caused by the difference in pixel distribution, pixel loss is added when optimizing the image generated by the generator, which can be expressed as:

$$L_c = |G(z) - x| \quad (10)$$

In the formula, L_c is the pixel loss. Pixel loss is combined with counter loss and segmentation loss to jointly control the distribution of cell nucleus images. The loss function L_{Θ} of the GAN model is:

$$L_{\Theta} = L_a + \lambda L_b + \mu L_c \quad (11)$$

In the formula, λ is the segmentation parameter, and μ is the pixel parameter.

4 Experimental results and analysis

The experimental model is based on the TensorFlow1.9 framework and uses the cuDNN7.5 kernel for calculation. The workstation is configured with Intel Core TMi7-6800K CPU@3.4 GHz, GTX 1080Ti graphics card, and memory 128 GB.

4.1 Network parameter setting

During network training, the optimization method uses random gradient descent (SGD), the momentum parameter is set to 0.99, and the weight decay rate is 1×10^{-4} . The learning rate is a reduction strategy of the initial learning rate $learning\ rate = 7 \times 10^{-3}$ multiplied by $(1 - current\_iter / max\_iter)^{\tau}$. Among them, $\tau = 0.99$ and $current\_iter$ are the current number of iterations, and max_iter is the maximum number of iterations in the training process. The number of trainings in the experiment is 10,000, the image size in the experiment is set to 512×512 pixels, and the training batch size is 3×6=18. At the same time, during training, data enhancement measures such as random cropping, horizontal flipping, vertical flipping, and random sample scrambling are adopted. In the evaluation, the image is scaled in multiple scales, and the zoom ratio is 0.65-1.65.

4.2 Experimental data

The experimental data comes from the 2018 data science bowl, which is manually labeled by professional doctors. Contains 670 pairs of original images with 9 resolutions and annotated segmented images of each cell nucleus, as shown in Fig. 8. The original pictures in the data set can be divided into grayscale images and color images, as shown in Fig. 8(a). Each original picture corresponds to multiple labeled segmented images of cell nuclei, that is, an original picture usually contains multiple cell nuclei. The annotated images of multiple nuclei are merged as shown in Fig. 8(b).

Fig.8 Some examples of nucleus segmentation image

Because when collecting raw data, different acquisition methods, different magnification magnifications, and different cell presentation methods are used, and the collected cell types are inconsistent. This leads to different cell images in the data set with distinct light and dark, which requires a strong generalization ability of the model to be able to adapt to various situations.

4.2.1 Image preprocessing

Due to the large imaging differences in the cell images in the data set, the image segmentation will be affected. Therefore, image enhancement method is used to preprocess the input image to obtain high contrast image. In the proposed technique, Adaptive HE (AHE) is used to enhance the local contrast to enhance the edge sharpness of each image module. So as to improve the quality of the original CT image, improve the shape, texture and other information of the medical image. The quality of the original CT image is improved, and the shape, texture and boundary information of the organ are improved. Before entering the network, the normalized image gray value is [0~1], and the image size is 512×512.

In addition, in the process of acquiring images, various noises are often interfered and contaminated, resulting in a reduction in the signal-to-noise ratio of the image and blurring of the edges of the cells and the background. In order to improve the signal-to-noise ratio of the image, it is generally necessary to perform a filtering operation on the image. Among them, Gaussian smoothing filter is used to complete the image preprocessing, and the mathematical description of the filtering process is:

$$G(x, y) = \frac{1}{2\pi\epsilon^2} e^{-\frac{(x-\kappa-1)^2 + (y-\kappa-1)^2}{2\epsilon^2}} \quad (12)$$

In the formula, ϵ is the standard deviation of the gray value, and κ is the dimension of the Gaussian convolution kernel.

4.2.2 Data enhancement

In order to overcome the over-fitting phenomenon generated by the neural network, enhancement techniques such as random shearing, flipping, gray-scale perturbation and shape perturbation are adopted for the data image. Gray perturbation can transform each pixel in a small range, multiplying the gray value of the cell nucleus image by a random number [0.82~1.12], plus a random number [-0.12~0.12]. The gray level disturbance in the training set can improve the stability of the network.

The affine transformation is used to deform the nucleus image and the contour image to form a shape disturbance. The deformation method is first to get the coordinates of 3 vertices (upper left, upper right and lower left), and then move each point randomly. The range of random movement is the image length. Finally, affine transformation is performed on the entire image.

4.3 Evaluation index

Pixel accuracy (PA), mean pixel accuracy (MPA), and Mean Intersection over Union (MIOU) are used as indicators to evaluate the performance of the proposed technology. And take MIOU as the main evaluation criterion of the experiment. Among them, PA is the simplest accuracy measure for semantic segmentation, which represents the proportion of pixels that are correctly labeled to the total pixels. MPA calculates the average value of all types based on the proportion of pixels that are correctly segmented within each type. MIOU calculates the pixel intersection ratio within each pixel category, and then calculates the average value. The calculation of PA, MPA and MIOU is as follows:

$$PA = \frac{\sum_{m=0}^g Q_{mm}}{\sum_{m=0}^g \sum_{n=0}^g Q_{mn}} \quad (13)$$

$$MPA = \frac{1}{g+1} \sum_{m=0}^g \frac{Q_{mm}}{\sum_{n=0}^g Q_{mn}} \quad (14)$$

$$MIOU = \frac{1}{g+1} \sum_{m=0}^g \frac{Q_{mm}}{\sum_{n=0}^g Q_{mn} + \sum_{n=0}^g Q_{nm} - Q_{mm}} \quad (15)$$

In the formula, Q_{mm} represents the correct number of divisions; Q_{mn} indicates that the original pixel belongs to the m category, but is divided into the number of n categories; Q_{nm} represents the number of pixels that originally belonged to n categories, but were divided into m categories. There are $g+1$ categories (including g category and an empty category or background category).

4.4 Training process

When training the network, in order to enhance the generalization ability of the network, the input image undergoes local response normalization before the first layer of convolution, $\alpha = 0.001$, and $\beta = 0.075$. The objective loss function is optimized by using the Adam algorithm with a learning rate of 0.0001 and iterated until the loss function converges. The weight decay is 0.0001, and the number of iterations is set to 10000. In the training process, the training data set is randomly shuffled, and then the batchsize is set to 20. Due to the large deviation of the number of pixels in each category in the data set, the median frequency equalization method is used to balance between classes.

In the experiment, the proposed method was iteratively trained on the data set for 50 epochs. The changes in MIOU, MPA, PA

and loss of the validation set during the training process are shown in Fig. 9.

Fig.9 Curves of MIOU, MPA, PA and Loss

It can be seen from Figure 9 that the MIOU value reached 80% when the network was trained to the 25th epoch. After iterating to 40 epochs, the value of MIOU stabilized and reached 85%. The MPA value and the PA value loss function curve are similar, and reach a stable value after 30 epochs, which are close to 92% and 90%, respectively. Similarly, the loss value is reduced to about 7% when the network iteration reaches 30 epochs. After 50 epoch iterations, the loss value basically converged, and it was reduced to about 4%.

4.5 Comparison of the method effects

In order to demonstrate the segmentation effect of the proposed method, it is compared with the methods in reference [13] and [19]. The segmentation result is shown in Fig. 10.

Fig.10 Different methods of nucleus image segmentation

It can be seen from Fig. 10 that the reference [13] uses the rough fuzzy C-means algorithm to segment the central part of the nucleus image very well, but the rough fuzzy C-means algorithm has a general learning ability. The segmentation of edge details and smaller cell nuclei is not very good, and there are certain over-segmentation and under-segmentation. Reference [19] completes image segmentation by automatically optimizing the CNN network for interactive segmentation learning. Similarly, the central part of the nucleus image is segmented very well, and the phenomenon of over-segmentation and under-segmentation is relatively reduced. However, the result of manual annotation is relatively rough, and the ability to segment smaller nucleus is poor. The proposed method improves the FCN model and combines it with the GAN network. For edge details and small nuclei segmentation, the effect is better than reference [19], and the phenomenon of over-segmentation and under-segmentation is relatively reduced, which is closer to the results of manual annotation. It can be argued that the proposed method has ideal segmentation capabilities.

In addition, three indicators of PA, MPA and MIoU are used to evaluate the segmentation results of the proposed method and reference [13] and [19]. Use all the test set data in the experiment to calculate the difference between the results of each technology segmentation and the manual segmentation standard. The evaluation results are shown in Table 1.

Tab.1 Performance comparison of three image segmentation methods

	PA/%	MPA/%	MIoU/%
Ref.[13]	84.83	85.17	77.52
Ref.[19]	87.19	88.72	81.61
The proposed method	90.08	92.56	85.34

It can be seen from Table 1 that compared with other methods, the proposed method has the best segmentation performance, and its MIoU reaches 85.34%. Because the proposed method combines the image processing advantages of the FCN model and the GAN network, the segmentation accuracy is further improved. Reference [13] uses rough fuzzy C-means to effectively improve the segmentation accuracy of medical images. However, the processing power of traditional algorithms is not good, so the overall performance is not ideal, MPA is only 77.52%. Reference [21] proposed an interactive subdivision and segmentation method based on deep learning, which achieved a better segmentation effect. However, there are still some shortcomings in the image segmentation of complex nuclei. Compared with the proposed method, its MIoU is reduced by 3.73%.

The experimental results prove that the method in this paper can improve the accuracy of image segmentation and at the same time generate more realistic nucleus images.

5 Conclusion

The correct segmentation of cell nucleus is of great practical significance to assist doctors in medical image analysis. For this reason, a cell nucleus image segmentation technology based on GAN network and FCN model is proposed. The preliminary segmentation result of the nuclear image obtained by the FCN model is used as the input of the GAN network. Optimize the structure of GAN network, introduce segmentation branch in the discriminator, and introduce pixel loss in the generator. In this way, an image of the cell nucleus that is visually more similar to the real image is obtained. Finally, the segmented image output by the FCN model is used as the input of the GAN network to achieve high-precision segmentation of the nucleus image. The proposed method is experimentally demonstrated based on the 2018 data science bowl dataset. The results show that it has better edge details and smaller cell nucleus segmentation capabilities. PA, MPA and MIoU are respectively 90.08%, 92.56% and 85.34%, which are better than

---

other comparison methods and have good application prospects.

Since the target of medical image segmentation usually has a certain shape, in the following research, we can try to set some rules to judge the result of deep learning network segmentation. In this way, segmentation results that do not belong to the target are filtered out to further improve the accuracy of segmentation.

Declarations

Funding: No funding available.

Conflicts of interest/Competing interests: No conflicts to declare.

Availability of data and material: No data available.

Code availability: No code available.

Authors' contributions: Zhang, Shi, Hu, and Yu designed the network model, analysed the segments for extraction, implemented the proposed method, validated results and wrote the paper.

References

- [1] Ramu S M , Rajappa M , Krithivasan K , et al. A novel fast medical image segmentation scheme for anatomical scans[J]. Multimedia Tools and Applications, 2019, 78(15):1-32.
- [2] Mahmoud A A , El-Shafai W , Taha T E , et al. An Efficient Segmentation Technique for Different Medical Image Modalities[J]. Menoufia Journal of Electronic Engineering Research, 2021, 30(1):22-28.
- [3] Moirangthem M , Singh T R . Brain tumor detection through content-based medical image retrieval using roi segmentation with Harmony search optimization[J]. Journal of Green Engineering, 2020, 10(10):8939-8969.
- [4] Fadzil A F A , Khalid N E A , Ibrahim S . Amplification of pixels in medical image data for segmentation via deep learning object-oriented approach[J]. International Journal of Advanced Technology and Engineering Exploration, 2021, 8(74):82-90.
- [5] Reddy U , Dhanalakshmi P , Reddy P . Image Segmentation Technique Using SVM Classifier for Detection of Medical Disorders[J]. Ingénierie des Systèmes D Information, 2019, 24(2):173-176.
- [6] Veeramuthu A , Meenakshi S , Kumar K A . A neural network based deep learning approach for efficient segmentation of brain tumor medical image data[J]. Journal of Intelligent and Fuzzy Systems, 2019, 36(1):1-8.
- [7] Kadoya N . Deformable Image Registration and Auto-Segmentation for Various Medical Imaging Types[J]. Igakubutsuri: Nihon IgakuButsuri Gakkai kikanishi = Japanese journal of medical physics: an official journal of Japan Society of Medical Physics, 2019, 39(1):12-19.
- [8] Taheri M , Rastgarpour M , Koochari A . A Novel Method for Medical Image Segmentation Based on Convolutional Neural Networks with SGD Optimization[J]. Iranian Journal of Electrical and Electronic Engineering, 2021, 9(6):37-46.
- [9] Fatmatülzehra USLU, Bass C , Bharath A A . PERI-Net: a parameter efficient residual inception network for medical image segmentation[J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2020, 28(4):2261-2277.
- [10] Omid, Tarkhaneh, Haifeng, et al. An adaptive differential evolution algorithm to optimal multi-level thresholding for MRI brain image segmentation[J]. Expert systems with applications, 2019, 138(8):112820-112820.
- [11] Wang X , Li W , Zhang C , et al. An adaptable active contour model for medical image segmentation based on region and edge information[J]. Multimedia Tools and Applications, 2019, 78(23):33921-33937.
- [12] Elbalaoui A , Fakir M . Exudates detection in fundus images using mean-shift segmentation and adaptive thresholding[J]. Computer methods in biomechanics and bio, 2019, 7(1-2):147-155.
- [13] Roy S , Maji P . Medical Image Segmentation by Partitioning Spatially Constrained Fuzzy Approximation Spaces[J]. IEEE Transactions on Fuzzy Systems, 2020, 28(5):965-977.
- [14] Zheng H , Zhang Y , Yang L , et al. An Annotation Sparsification Strategy for 3D Medical Image Segmentation via Representative Selection and Self-Training[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4):6925-6932.

-
- [15] Shang H , Zhao S , Du H , et al. A new solution model for cardiac medical image segmentation[J]. Journal of Thoracic Disease, 2020, 12(12):7298-7312.
- [16] Karimi D ,Salcudean S E . Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks[J]. IEEE Transactions on Medical Imaging, 2020, 39(2):499-513.
- [17] A D M , B BB , A R G . Image super-resolution using progressive generative adversarial networks for medical image analysis[J]. Computerized Medical Imaging and Graphics, 2019, 71(8):30-39.
- [18] HassanzadehT ,Essam D , Sarker R . An Evolutionary DenseRes Deep Convolutional Neural Network for Medical Image Segmentation[J]. IEEE Access, 2020, 8(3):212298-212314.
- [19] GuotaiW ,Zuluaga M A , Wenqi L , et al. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7):1559-1572.
- [20] KhinT , Raju K S , Sinha G R , et al. Review of Optimization Methods of Medical Image Segmentation[J]. Advances in Intelligent Systems and Computing, 2020, 1090(1):213-218.
- [21] Fang L , Wang X , Wang L . Multi-modal medical image segmentation based on vector-valued active contour models[J]. Information Sciences, 2020, 513(4):504-518.
- [22] Li K , Yu L , Wang S , et al. Towards Cross-Modality Medical Image Segmentation with Online Mutual Knowledge Distillation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):775-783.
- [23] Cheng J , Tian S , Yu L , et al. Multi-Attention Mechanism Medical Image Segmentation Combined with Word Embedding Technology[J]. Automatic Control and Computer Sciences, 2020, 54(6):560-571.
- [24] Wang E K , Chen C M , Hassan M M , et al. A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain[J]. Future Generation Computer Systems, 2020, 108(3):135-144.
- [25] Li D , Deng N , Chen X . Level set medical image segmentation aided by cooperative quantum particle optimization with Lévy flights[J]. Vibroengineering PROCEDIA, 2019, 28(10):93-98.
- [26] NiuY , Cao J . Local difference-based active contour model for medical image segmentation and bias correction[J]. IET Image Processing, 2019, 13(10):1755-1762.
- [27] NieD , Wang L , Xiang L , et al. Difficulty-Aware Attention Network with Confidence Learning for Medical Image Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(7):1085-1092.
- [28] NazeeraS ,Vinodkumar R S , Ramya R R . A Survey on Methodology based Medical Image Segmentation Techniques and their Applications[J]. Journal of Advanced Research in Dynamical and Control Systems, 2019, 11(9):612-622.
- [29] Yang T , Yoshimura Y , Morita A , et al. Pyramid Predictive Attention Network for Medical Image Segmentation[J]. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, 2019, 102 (9):1225-1234.
- [30] Dar A S ,Padha D . Medical Image Segmentation A Review of Recent Techniques, Advancements and a Comprehensive Comparison[J]. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, 2019, 7(7):114-124.
- [31] Chen C , Dou Q , Chen H , et al. Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(8):865-872.
- [32] Wang G , Li W , Aertsen M , et al. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks[J]. Neurocomputing, 2019, 338(04):34-45.
- [33] Han B , Han Y , Gao X , et al. Boundary constraint factor embedded localizing active contour model for medical image segmentation[J]. Journal of ambient intelligence and humanized computing, 2019, 10(10):3853-3862.

Kai Zhang, Master of Computer Science, Associate Professor. Graduated from Changchun University of Science and Technology in 2006. Worked in Changchun University of Finance and Economics. His research interests include machine learning and Internet of Things.

Shi Yang, Master of Computer Science, Associate Professor. Graduated from Changchun University of Science and Technology in 2010. Worked in Changchun University of Finance and Economics. His research interests include machine learning and cloud computing.

Chengquan Hu, Master of Computer Science, Professor. Graduated from Jilin University in 1985. His research interests include machine learning and Internet of Things.

Hang Yu, Master of Computer Science, Associate Professor. Graduated from Jilin University in 2008. Worked in Jilin Agricultural University. Her research interests include machine learning and big data.