

Research on the nucleus segmentation method based on improved U-Net network

Ruohuan Tang^{1,a,*}, Zhiyong Liu^{1,b}, Jianqiang Jia^{1,c}

¹School of Intelligent Engineering and Intelligent Manufacturing, Hunan University of Technology and Business, Changsha, China

^atruohuan@163.com, ^blzy18075858568@163.com, ^c369601423@qq.com

*Corresponding author

Abstract: Medical image segmentation is a key technology in the field of medical image processing and analysis, which aims to segment special parts of medical images and extract relevant features to provide a reliable basis for clinical diagnosis and pathology research, and to assist doctors in making accurate judgments. Accurate nuclear segmentation is the basis of cell detection, cell classification and tumor grading, and has gradually attracted great attention from researchers in recent years. Due to the low contrast of nucleus images, large differences in the spatial distribution of nuclei, and easy formation of adhesions between cells, the accurate segmentation of nucleus images has become one of the difficult problems in medical image analysis. Based on the deep learning image segmentation algorithm and the U-Net network model, this paper designs a nucleus segmentation model based on ResNet residual blocks, and optimizes its network structure to achieve accurate segmentation of nucleus images.

Keywords: nucleus segmentation; U-Net; ResNet; pathological images

1. Introduction

Cancer, also known as malignant tumor, usually refers to the abnormal proliferation of cells, and the abnormal mechanism of cell division and proliferation can lead to certain diseases, and the proliferative cells can infiltrate into the normal tissues of the body or transfer to other parts of the body from the internal circulatory systems and lymphatic systems. At present, cancer maintains a very high growth rate and mortality rate every year. In addition to cardiovascular and cerebrovascular diseases, cancer causes the largest number of deaths every year. According to a survey by the World Health Organization (WHO), cancer is the second leading cause of death in the world, accounting for nearly one-sixth of all mortality. At present, the overall global cancer prevalence is on the rise, which not only has a huge impact on the economic development of various countries, but also brings a heavy burden on patients and their families. In recent years, the statistics of China's cancer centers show that the overall cancer incidence rate in China has maintained an annual increase of about 3.9%, and the number of cases is the highest in the world. The statistical results also show that the mortality rate of cancer in China has increased by about 2.5% per year. The problem of cancer has become a common problem faced by mankind. Therefore, it is of great significance to the whole world and China to strengthen the research on the diagnosis, treatment and prognosis of cancer.

In recent years, with the rapid development of medical imaging technology, microscopic imaging, X-ray, magnetic resonance imaging and ultrasound imaging have become indispensable in various medical diagnoses, which are of great value for improving the diagnostic efficiency and accuracy of doctors^[1]. Compared with natural scene images such as houses, cars, buildings, etc., medical images have great particularities. Usually in natural scenes, the differences between different targets are large and the contours of the targets are obvious, while medical images often have problems such as blurred imaging, low contrast, uneven spatial distribution of nuclei, and different imaging equipment and image storage methods will lead to differences between images, which makes nucleus segmentation in medical images a challenging task. Although there are many algorithms for medical nucleus image segmentation, there is still a lot of room for improvement in the performance of nucleus segmentation.

2. Related work

With the continuous maturity and development of digital image processing technology, the use of

computer graphics for medical image processing has gradually become a hot research and application hotspot at home and abroad. Many scholars are committed to improving its accuracy and efficiency to help medical diagnosis and treatment.

In the early days, traditional nuclei segmentation mostly used methods such as threshold, watershed, activity profile, clustering, and graph theory. Among them, the threshold segmentation method is the earliest method used in nucleus segmentation. Phoulady et al.^[2] proposed an iterative threshold method to segment the nucleus and overlapping cytoplasm in cervical cell images, which is more resistant to brightness changes in the image, but has lower overall segmentation accuracy. Rajyalakshmi et al.^[3] proposed a modified marker-controlled watershed method (MMCWA) for nuclear segmentation of preprocessed Haematoxylin & Eosin stained breast cancer histopathological images. Small fixed structural element (Squeeze-Excitation module) sizes remove the corresponding light and dark details during opening and closing modalities, while large SE sizes remove large contour details from the input image. By using the weighted variance method, the adaptive structural element size of the SE image is obtained to protect all the details in the image. Kass et al.^[4] proposed a profile-based segmentation method in which the segmentation boundary is represented as a deformable contour line, and the gradient information is used to construct an energy function and minimize the energy function to determine the shape of the profile. Cui et al.^[5] proposed a network to predict kernel instances, nucleus profiles and backgrounds. The network utilizes a weighted loss function based on the relative position of pixels in the image to improve and stabilize contour predictions. Kothari et al.^[6] used K-means clustering to achieve segmentation of nuclei and background. The accuracy of the depression detection is better, but the results depend on the color segmentation effect during the K-means clustering process. Oyeboode et al.^[7] proposed an adaptive parameter selection method based on graph theory to solve the problem of manual parameter selection, in which the grayscale image of the nucleus was logarithmically transformed, and the nucleus boundary was extracted by narrowing the dynamic range of the foreground pixels.

The traditional method as described above has simple principles and is easy to understand and implement, but it has limitations and poor robustness, and requires a lot of pre-processing or post-processing. Therefore, traditional image segmentation methods have not been widely used in medical nucleus segmentation, and image segmentation methods based on deep learning have attracted the attention of scientific researchers.

Li et al.^[8] combined the FCN network with ResNet to solve the problem of gradient vanishing in the deep networks and achieve accurate segmentation of cancer cells. In addition to the structural improvement of the FCN network, different loss functions can also be adopted. Naylor et al.^[9] proposed a deep learning method that does not consider contours, modeling the segmentation task as a supervised regression problem, thereby solving the segmentation problem of overlapping nuclei through adaptive moment estimation. The network also avoids making predictions on areas with unclear contours. Yang et al.^[10] constructed a CNN model consisting of two convolutional-pooling layer pairs and two fully connected layers for histopathological nuclei images, implemented end-to-end model training, and then applied the level set model according to the output results to complete the segmentation of the overlapping nuclei of the test images by minimizing the energy function. Fakhry et al.^[11] further improved the segmentation framework based on CNN network, and applied watershed merging trees to the improved network framework to improve the accuracy of neuron segmentation.

In summary, whether it is based on traditional methods or deep learning-based nucleus segmentation in pathological images, it is necessary to comprehensively consider factors such as the amount of training data, data quality, and image quality to select the most suitable algorithm for segmentation.

3. Methods of Nucleus Segmentation

3.1. Introduction to the U-Net network model

The U-Net network^[12] was first proposed at the MICCAI conference in 2015. The U-Net network structure consists of a contraction path that captures context and an extended path that supports precise localization, which can be trained end-to-end with very few images and achieves state-of-the-art results on the segmentation of neural structures for electron microscopy in the ISBI challenge.

In the coding stage of U-Net, two consecutive convolution operations with a kernel size of 3×3 are performed on the input image, each convolution operation is followed by the ReLU activation function, and then the feature dimension is reduced by the maximum pooling layer operation. In the decoding stage, an upsampling operation is required to restore the feature information compressed during the

downsampling process in the encoding stage. In this process, not only can the feature information compressed in the downsampling stage be restored, but also the output information of the downsampled and upsampled output information of the corresponding network layer can be fused through skip connection in the corresponding network layer. This enables the splicing of cross-level feature maps. Then, the fused information is also subjected to two consecutive convolution operations with a kernel size of 3×3 . Each convolution operation will be followed by the ReLU activation function, and the obtained output result will be used as the input value of the next upsampling. Finally, a convolution operation with a kernel size of 1×1 is used to map the output result to an image to achieve the binary classification problem. The output results only include the segmentation results and the background. Overall, the entire U-Net network includes 23 convolutional network layers, 4 maximum pooling layers that implement downsampling operations, 4 upsampling operation layers, and 4 skip connection operations between the network layers at the corresponding locations.

The focus of image segmentation tasks based on deep learning is how to make efficient use of spatial information. In the encoding stage, continuous convolution operations are followed by maximum pooling operations, so that the resolution of the feature map becomes smaller and smaller. If only continuous upsampling operations are used to restore the lost feature detail information in the decoding stage, the segmentation effect will be very poor, so it is necessary to add the compressed information in the encoding stage. Through the short-circuit connection of skip connection, the network output information in the encoding process can be fused into the decoding process, and the fusion method effectively uses the cross-level feature information. In addition, this fusion of multi-scale feature information will produce more accurate segmentation results.

3.2. Nucleus segmentation model based on ResNet residual blocks

In this paper, we propose to add ResNet residual blocks^[13] on the basis of the U-Net network model, which can solve the overfitting problem when the network depth is very large. The combination of residual blocks and convolution operations is used to integrate residual blocks into all layers of the whole network, and the add operations between different network convolutional layers are used to integrate multi-scale and cross-level feature information. The output information obtained by the input information of the network after two consecutive convolution combination operations is added to the original input information to obtain the output result, and the result is processed by the activation function to obtain the final output result. The formation process of the whole residual block is shown in Figure 1. The process of forming the residual block is shown in Equations (1) and (2):

$$y_1 = h(x_1) + F(x_1, W_1) \tag{1}$$

$$x_{l+1} = f(y_1) \tag{2}$$

$F(x_i, W_i)$ is the residual function, which represents the learned residuals. f is the ReLU activation function. From the above equations, the learning features from the layer l to the layer L can be calculated, as shown in Equation (3):

$$x_L = x_1 + \sum_{i=1}^{L-1} F(x_i, W_i) \tag{3}$$

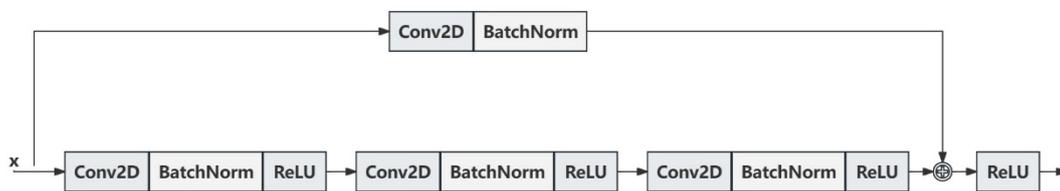


Figure 1: ResNet network residual block.

This paper plans to design a nucleus segmentation model based on ResNet. In the encoder part, the input image information of the network can be used to extract the image feature information after the residual block processing, and then the downsampling operation is carried out to reduce the feature dimension of the image. A total of four downsampling operations are performed in this process. The output information of the encoder is processed by the residual block as the input information of the decoder. In the decoder part, the input information is upsampled, and the input residual block is processed to obtain the output information. In order to restore the feature information of the image compressed by the model in the encoding stage, a total of four upsampling operations are performed. At the same time,

the model adopts the skip connection operation in the U-Net network structure to integrate cross-level image feature information.

The loss function used in the training process is the cross-entropy loss function^[14] commonly used for binary classification problems, and its Equation is shown in (4):

$$L = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})] \quad (4)$$

Where y is the real data value and \hat{y} is the predicted data value. When $y = 1$, $L = -\log \hat{y}$; When $y = 0$, $L = -\log(1 - \hat{y})$.

4. Experimental results and analysis

4.1. Introduction to the dataset

In this paper, we used the Colorectal Nuclear Segmentation and Phenotypes (CoNSeP) dataset, which consists of 41 Haematoxylin & Eosin stained images, each of size 1,000×1,000 pixels at 40× objective magnification. Images were extracted from 16 colorectal adenocarcinoma (CRA) WSIs, each belonging to an individual patient, and the dataset focused on a single cancer type in order to be able to show true changes in tissues within colorectal cancer WSIs. In addition to delineating the nuclear boundaries, every nucleus was labeled as either: normal epithelial, malignant/dysplastic epithelial, fibroblast, muscle, inflammatory, endothelial or miscellaneous.

4.2. Processing of Datasets

In response to the problem of less training data in the nucleus image segmentation task, in order to prevent the model from overfitting, this paper will use methods such as flipping, scaling, cropping, shifting, and Gaussian noise^[16] to enhance the dataset during the training process.

- (1) Flipping: Flip the picture vertically or horizontally;
- (2) Scaling: Scaling is mainly divided into inward scaling and outward scaling, outward scaling is to expand the original picture according to a certain ratio, and inward scaling is to shrink the original image according to a certain ratio;
- (3) Cropping: Mostly, a part of the original image is randomly selected, and then enlarge the image to the same size as the original image;
- (4) Shifting: Move the picture on the X-axis or Y-axis;
- (5) Gaussian noise: Gaussian noise with zero mean is usually used to weaken the impact of high-frequency features on the model, and the neural network will have better learning ability by adding noise to the picture appropriately.

When training a neural network, the training set includes the original image and the corresponding annotated image (ground truth), and the data augmentation should not only enhance the original image of the training set, but also enhance the corresponding annotated image. If only the original image is augmented with data, there will be a problem that the original image and the labeled image cannot correspond. The dataset is incorrect, and the segmentation network cannot be trained.

4.3. Experimental setting

This paper uses the Keras library to build commonly used layers such as the convolutional layer, pooling layer, and regularization layer. In order to build more flexible deep learning network models and related computing requirements, Tensorflow is used for gradient computing and related function development. The tool versions and hardware information are shown in Table 1.

To ensure the effectiveness of the comparison experiment, the hyperparameter settings are the same for all model training. Specifically, cross-entropy is used as the loss function, Adam is used as the optimizer, and the learning rate is 0.0001. The specific parameter indicators are shown in Table 2.

Table 1: Experimental operating environment.

Name	Attribute
Deep learning framework	Tensorflow 1.15.0、Keras2.3.1
Python version	3.5.6
Numpy version	1.17.2
Matplotlib	3.1.1
Opencv	3.4.2
GPU video memory	11019MiB (GeForce RTX 2080 Ti)
Memory	64GB

Table 2: Model hyperparameter settings.

The type of the hyperparameter	The value of the hyperparameter
Loss function	cross-entropy
Learning rate	0.0001
Optimizer	Adam
Epoch	100
Batchsize	4

4.4. Performance metrics

In order to verify the segmentation accuracy of the nuclear model based on ResNet residual blocks, this paper uses the Dice coefficient as the evaluation index, which is the most frequently used metric in medical image processing. It is a metric function of set similarity, which is usually used to calculate the similarity between two samples.

The DICE coefficient evaluation index equation is shown in (5):

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \tag{5}$$

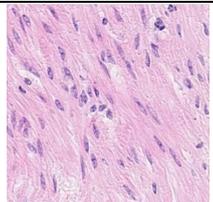
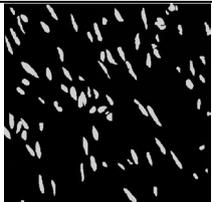
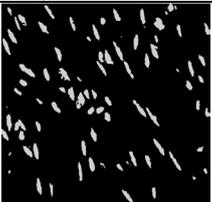
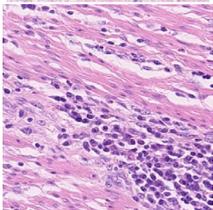
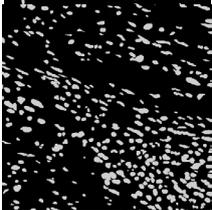
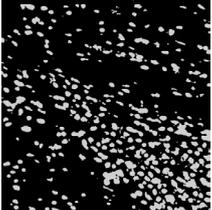
where X is the segmentation result value output after the image is input to the trained neural network, and Y is the target image standardly outlined by experienced pathologists. In the field of medical image processing, it is generally referred to as the ground truth.

$|X \cap Y|$ is the intersection between X and Y, $|X|$ and $|Y|$ represent the number of element X and element Y, respectively. The value range of equation (5) is within the interval [0,1], the larger the value, the higher the coincidence rate between the segmentation result and the ground truth, indicating that the trained network model has a good fitting effect and achieves the ideal segmentation effect.

4.5. Experimental results

The experimental results of the nuclear segmentation model based on ResNet residual blocks are shown in Table 3.

Table 3: Results of nuclear segmentation model based on ResNet residual blocks

Test pictures	ground truth	The result of segmentation	DICE
			0.8207
			0.8032



In order to verify the performance of the proposed method on nucleus segmentation, it was compared with two other commonly used image segmentation algorithms on the same dataset. Table 4 lists the experimental results of different models on the nucleus. In addition to the comparison of the U-Net network, the CNN is also compared, and it can be found from Table 4 that the nuclear segmentation algorithm proposed in this paper has achieved better results. The average Dice coefficient on the test set reached 81.37%, which is higher than that of CNN and U-Net network, and has important application value.

Table 4: Segmentation results of each network model.

Network model	Average Dice coefficient
CNN ^[17]	0.7822
U-Net ^[12]	0.8056
Methods of this paper	0.8137

5. Conclusions

With the rapid development of medical imaging technology and digital pathology, the efficiency and accuracy of medical diagnosis have been greatly improved. The accurate segmentation of the nucleus is the basis of pathological analysis of medical images, and it is also the premise of accurate analysis of tumor diagnosis by computer. This paper designs and implements a nuclear segmentation model based on ResNet residual blocks. ResNet residual blocks are combined in the encoding and decoding stages of the U-Net network model. On the same nuclear pathology dataset, the model was compared with the basic U-Net network model. In order to solve the problem that it is difficult to identify the boundaries of pathological nuclei images with sparse staining and overlapping boundaries, combined with the ResNet residual block, the local feature information obtained in the coding stage is helpful to improve the segmentation accuracy of nuclei. The experimental results show that the average Dice coefficient of the nucleus segmentation model based on the ResNet residual block is 0.8137, which is higher than that of the U-Net model, and has important clinical application value.

Although the method proposed in this paper has achieved good results, there are still shortcomings that require further research and improvement. In future research, it is necessary to increase the amount of training data and further improve the generalization ability of the model.

Acknowledgments

This project was sponsored by the Postgraduate Scientific Research Innovation Project of Hunan Province (No. CX20221180).

References

- [1] Dytrych T, Launey KD, Draayer J P, et al. Physics of nuclei: Key role of an emergent symmetry[J]. *Physical Review Letters*, 2020, 124(4):1-6.
- [2] Phoulady H A, Goldgof D, Hall L O, et al. A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images[J]. *Computerized Medical Imaging and Graphics*, 2017, 59: 38-49.
- [3] Rajyalakshmi U, Rao S K, Prasad K S. Supervised classification of breast cancer malignancy using integrated modified marker controlled watershed approach[C]//2017 IEEE 7th International Advance Computing Conference (IACC). IEEE, 2017: 584-589.
- [4] Kass M, Witkin A, Terzopoulos D. Snake: active contour models [C]//Proceedings of the 1st International Conference on Computer Vision London: IF Computer Society Press, 1987:259-268.
- [5] Cui Y, Zhang G, Liu Z, et al. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images[J]. *Medical & biological engineering & computing*, 2019, 57:

2027-2043.

[6] Kothari S, Chaudry Q, Wang M D. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques[C]//Boston, USA, IEEE International Symposium on Biomedical Imaging, 2009: 795-798.

[7] Oyeboode K O, Tapamo J R. Adaptive parameter selection for graph cut-based segmentation on cell images[J]. *Image Analysis Stereology*, 2016, 35(1): 29-37.

[8] Li Y, Shen L, Yu S. HEP-2 specimen image segmentation and classification using very deep fully convolutional network[J]. *IEEE Transactions on Medical Imaging*, 2017, 36(7):1561-1572.

[9] Naylor, P., Laé, M., Reyal, F., Walter, T. Segmentation of nuclei in histopathology images by deep regression of the distance map[J]. *IEEE Trans. Med. Imaging*, 2018.

[10] Yang L, Xing FY, Xie YP. An automatic learning-based framework for robust nucleus segmentation[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(2): 550-566.

[11] Fakhry A, Peng HC, Ji S. Deep models for brain EM image segmentation: Novel insights and improved performance. *Bioinformatics*, 2016, 32(15): 2352-2358.

[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation: MICCAI 2015, Cham, 2015[C]. Springer International Publishing, 2015.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[14] Ho Y, Wookey S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling[J]. *IEEE access*, 2019, 8: 4806-4813.

[15] Deng M, Brägelmann J, Schultze J L, et al. Web-TCGA: an online platform for integrated analysis of molecular cancer data sets[J]. *BMC bioinformatics*, 2016, 17: 1-7.

[16] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. *Journal of big data*, 2019, 6(1): 1-48.

[17] O'Shea K, Nash R. An introduction to convolutional neural networks[J]. *arXiv preprint arXiv:1511.08458*, 2015.