Gastric Cancer Pathological Image Segmentation based on Convolutional Neural Network

Wenyue Zhang[1,a], Ziyan Jia[1*], Qing Li[2,b], Dachuan Zhang[2,c], Lingjiao Pan[1,d], Dawei Shen[1,e]

[1]School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou Jiangsu, China
[2]Department of Pathology, Changzhou First People's Hospital, Changzhou, Jiangsu, China

Abstract: The difficulty in pathological image diagnosis of gastric cancer lies in the accurate segmentation of cancerous tissue in the picture. To increase gastric cancer pathological images segmentation accuracy, we optimized the basic UNet model and proposed the DCU-Net model. First, add a direct channel module to each layer of the encoding part to obtain more detailed information. In addition, considering that the image may cause a loss of information during the transmission process, a CA module is added before the up-sampling and down-sampling of each layer so that the model can obtain more channel information. After conducting segmentation experiments on our own gastric cancer data set and comparing it with several current classic segmentation models. The experiment proved that the segmentation accuracy using this article's model can be enhanced on our own gastric cancer image segmentation data set, achieving an accuracy of 91.30% and an IoU of 79.88%.

1. Introduction

Cancer, also known as malignant tumors, is the leading cause of death for the global population. Millions of people die from cancer every year, which has greatly hindered the improvement of overall human life expectancy [1]. According to the malignant tumor registration data from my country's Cancer Registration Center in 2023, the incidence of gastric cancer ranks third among the top ten malignant tumors in my country, accounting for 12.37% of men and 9.26% of women. At the same time, among patients with gastric cancer, the mortality rate for male gastric cancer patients is 13.08%, ranking third among the top ten cancer types, and the mortality rate for female gastric cancer patients is 10.01%, ranking second among the top ten cancer types [2].

In recent years, with deep learning technology developing at a rapid pace, assisting doctors in diagnosing pathology photos based on deep learning technology has progressively gained a popular research direction [3]. Research on gastric cancer pathological image processing technology using deep learning technology is currently in the exploratory stage. More and more scholars are studying how to increase the accuracy of the detection and the efficacy of pathology pictures for gastric cancer detection by improving the algorithm model structure based on deep learning [4].

The current mainstream technology of medical image segmentation relies on the fully convolutional neural network with U-shaped structure, especially the classic model U-Net [5]. Based on, Li et al. [6] proposed a fresh model of deep learning that can be used for accurate segmentation of histopathological images. The author added a layer of noise reduction to the U-Net basic model's output layer. This design not only reduces the stringent requirements for the accuracy of training data, but also achieves efficient performance by combining the noise segmentation results of traditional image segmentation algorithms, and achieve efficient unsupervised segmentation. Experimental results show that this enhanced version of U-Net model significantly surpasses traditional methods such as K-Means and Otsu in performance, and shows superior performance than the original U-Net in noise-free situations. Yi et al. [7] proposed to use fully convolutional neural networks for microvessel prediction on hematoxylin-eosin staining pathological images and to assist doctors in treating malignant tumors more effectively. The author used a fully convolutional network based on the Caffe framework for microvessel detection, and optimized the loss function through a stochastic gradient descent algorithm, achieving remarkable results. This study not only demonstrates the superior performance of this method in microvascular detection, but also proves its potential for clinical prediction. Azad et al. [8] proposed the BCDU-Net network by integrating bidirectional convolutional LSTM modules into dense skip connections. This enables the network to nonlinearly fuse two sets of feature maps to generate a set of features that contain rich

[a]zhangwenyue97wren@163.com
[*]Corresponding author: jiaziyan@jsut.edu.cn
[b]liqblk@163.com, [c]zhangdachuan@suda.edu.cn
[d]jsjshedy@jsut.edu.cn, [e]sdw@jsut.edu.cn

https://doi.org/10.1051/bioconf/202411103020

local semantics, significantly improving the accuracy of image segmentation. WU et al. [9] proposed the CGNet network structure and designed a context information guidance module that can learn global features and local features at the same time. ZHANG et al. [10] proposed to fuse the improved ASPP module with the U-Net model to lessen the loss of information about the edges in the U-Net segmentation operate brought on by several up- and down-sampling procedures. To replace the original convolution structure and lower the model's complexity, utilize depth-wise separable convolution.

2. Method

This article's proposed segmentation model framework is based on the U-Net foundational model. The network structure of the model is shown in Figure 1. Based on the U-Net model, consider adding Direct connection channel and attention modules. We will introduce these two parts in detail in the following sections. The optimized structure is displayed in Figure 2.

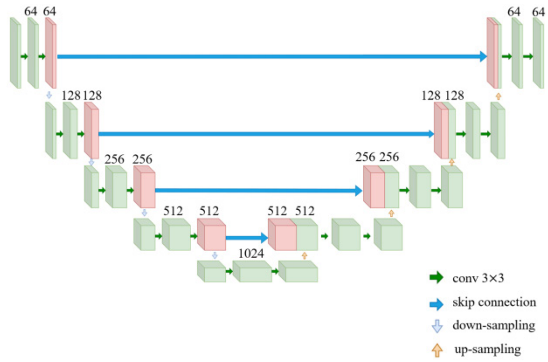

Fig. 1. Diagram of the U-Net network structure.

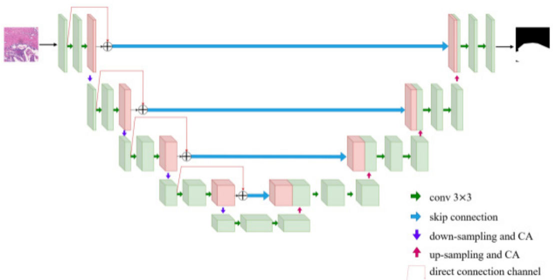

Fig. 2. Model structure of this article.

2.1. Directly connected channel module

The structure of the ResNet network model is mainly composed of residual modules. The residual module is designed to solve the problems of vanishing gradients and limited expression ability in deep neural networks. In traditional convolutional neural networks, as the number of network layers increases, the gradient signal will gradually weaken, making the network difficult to train. The ResNet network model proposes to solve the above problems through skip connections.

The residual module structure is to add the input features to the output features, rather than simply converting the input features into output features through a convolution operation. Such a design allows the network to learn identity mapping or approximate identity mapping more easily, and to propagate gradient signals more efficiently. Due to the existence of skip connections, the network can gradually optimize from shallow layers without End-to-end training from start to finish. This feature makes the network easier to train and helps speed up the training process.

This article introduces a direct connection channel module in the double convolution process of each layer in the encoding part. A direct connection channel structure diagram is displayed in Figure 3.

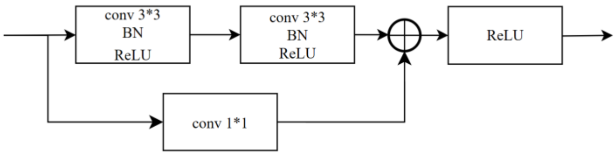

Fig. 3. Directly connected channel structure diagram.

2.2. Module for Coordinated Attention

Attention Mechanism is inspired by the brain's ability to solve information overload. The Coordinate Attention (CA) module is to solve the problem of image spatial information loss caused by the existing attention mechanism during the pooling process. The structure of the CA module is displayed in Figure 4.

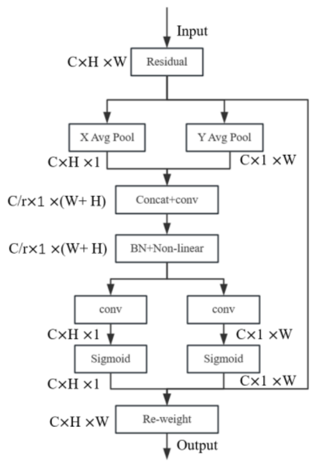

Fig. 4. CA module structure diagram.

The attention module's input is the feature tensor that the preceding layer produced. To obtain feature maps in the width and height dimensions, respectively, the input feature tensor is globally averaged and pooled in those directions. Suppose that the input feature layer's shape is [C, H, W]. Following width-wise average pooling, the resulting feature layer form is [C, H, 1]. Currently, we map the features to high dimensions. Following height-directional average pooling, the resulting feature layer form is [C, 1, W], and currently we map the features to the wide dimension.

https://doi.org/10.1051/bioconf/202411103020

The output in the height direction can be expressed as:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \tag{1}$$

The output in the width direction can be expressed as:

$$z_c^w(h) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, W) \tag{2}$$

After that, convert the width and height to the identical dimension and combine the features from the two dimensions to create the feature layer that is [C, 1, H+W], and then. the features are obtained after calculating convolution, normalization, and activation functions. Then it is divided into two parallel channels again, and then the width and height are divided into: [C, 1, H] and [C, 1, W], and then transposed to obtain two feature layers [C, H, 1] and [C, 1, W]. Then use convolution to adjust the number of channels and use the sigmoid activation function to obtain the attention in the width and height dimensions, and multiply by the original features to obtain the output tensor processed by the Coordinate Attention module.

3. Datasets and Evaluation

3.1. Datasets

The data set used in this article is 12 panoramic gastric cancer digital pathology images provided by Changzhou First People's Hospital. After slicing the entire image through a sliding window, the data set used for the segmentation task has a total of 961 cancer pathological images. Utilizing an 8:1:1 ratio, divide the training, test, and validation sets.

3.1.1 Image preprocessing

By transforming an image, the gray histogram distribution can be made more uniform by using the histogram equalization algorithm. However, parts of the image may also generate noise or lose detail information due to transitional equalization. To solve this shortcoming, a local histogram equalization method is proposed, but this method only considers pixels in local areas and ignores other areas. For similar areas in the image, the noise will be excessively amplified. The emergence of CLAHE algorithm can better solve the shortcomings of the above methods.

The CLAHE algorithm was proposed. By introducing the concept of limiting contrast, the height of the local histogram is limited in the process of image enhancement to control the degree of local contrast enhancement, thereby effectively avoiding excessive amplification of noise and excessive enhancement of local contrast. Can effectively improve image quality. The schematic diagram of CLAHE algorithm is shown in Figure 5.

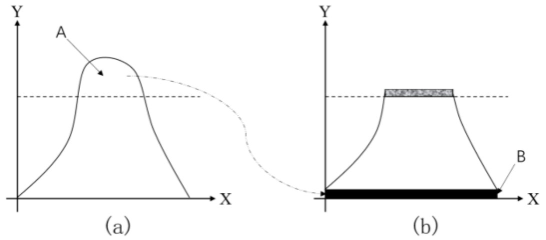

Fig. 5. The schematic diagram of CLAHE algorithm. (a). Distribution of pixel values not processed by CLAHE algorithm. (b). Pixel distribution processed using CLAHE algorithm.

However, the CLAHE approach is exceedingly time-consuming to process because each pixel in the image requires the calculation of its domain histogram and related transformation function. Therefore, interpolation operations are applied on this foundation to increase the algorithm's computational efficiency. After the gastric cancer pathological images are processed by the CLAHE algorithm, the comparison findings are displayed in Figure 6. It is visible from fig. 6. The processed image boundary information and other features are clearer, and there is no obvious noise impact.

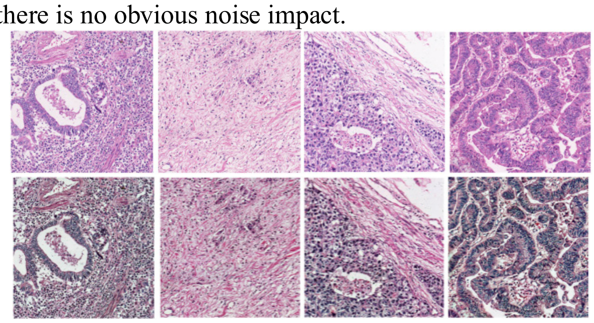

Fig. 6. Comparison of effects before and after CLAHE algorithm processing. The processed image is shown in the second row, while the original image is shown in the first.

3.1.2 Data enhancement

Data enhancement refers to generating more and richer training data through a series of transformations and expansions on the original data. Thus, the model's robustness and capacity for generalization are enhanced. Common data enhancement techniques include randomly rotating images, flipping images, Scale image size, crop image area, adjust image brightness, etc. The effect of image and label data enhancement in this article is shown in Figure 7.

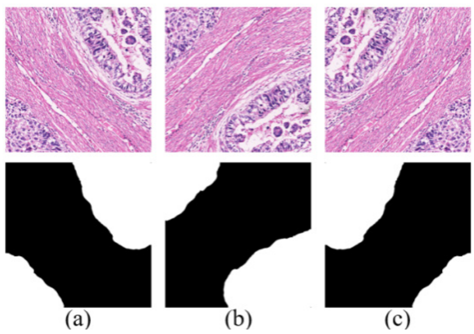

Fig. 7. Data enhancement of images and labels. (a). Images and labels without data enhancement. (b) Rotated images and labels. (c). Flipped images and labels.

3.2. Evaluation indicators

The evaluation indicators used in this article are the IoU and accuracy rate often employed in the segmentation of medical images. These evaluation indicators have a value range of 0 to 1. The model's segmentation performance improves with increasing value. The calculation of the IoU and accuracy needs to be obtained by calculating the confusion matrix, which is shown in Table 1.

Table 1. Confusion matrix.

	P(Positives)	N(Negatives)
T(Ture)	TP	TN
F(False)	FP	FN

The calculation formula for IoU is:

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

The calculation formula for accuracy is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

4. Experimental Results

4.1. Experimental setup

The model is implemented using Python 3.8 and Pytorch. Model training and testing are performed on a server. The classification model and segmentation model use Tesla T4 graphics card. The optimizer used is RMSprop, the learning rate is 0.00001, the epochs used for training are 100, and the batch size is 4.

4.2. Ablation study

By adding different modules, it is verified that the segmentation performance of this model is optimal, which will be covered in more detail in the sections that follow.

4.2.1 Add different attention modules

This article proposes to add an attention mechanism before up-sampling and down-sampling of the U-Net basic model. Currently, the more commonly used attention modules include SE, CBAM and CA. The following will verify the effects of different attention modules added to the basic model. The results are shown in Table 2.

Table 2. Ablation study of different attention modules.

Different attention modules	Accuracy (%)
SE	86.70
CBAM	85.33
CA	90.51

As Table 2 illustrates, compared to the other two attention modules, this model with added CA module has higher segmentation accuracy, with the Accuracy being 3.81% and 5.18% higher respectively.

4.2.2 Add different optimization modules

Ablation tests were performed on each module in the U-Net basic model to confirm the effectiveness of each module in this model. The specific results are displayed in Table 3.

Table 3. Ablation experiments of different optimization modules.

Directly connected channel module	Coordinate Attention module	IoU (%)	Accuracy (%)
-	-	73.98	89.54
√	-	75.21	90.48
-	√	75.28	90.51
√	√	79.88	91.30

As Table 3 illustrates, when the direct channel module and the CA module are added, respectively, the IoU and accuracy of the segmentation task are better than the basic U-Net model. When two optimization modules are added at the same time, its IoU and accuracy are improved by 5.9% and 1.76%, respectively, compared with the U-Net base model.

4.3. Comparison with segmentation model

In addition to testing the model proposed in this article on a private data set, several classic segmentation models were also selected for verification on this data set, including TransUNet, Deeplab-V3, and CGNet. The comparison results are shown in Table 4.

Table 4. Comparison of different segmentation models.

Models	IoU (%)	Accuracy (%)
TransUNet	71.68	89.59
Deeplab-V3	72.26	88.99
CGNet	76.59	90.11
Article model	79.88	91.30

According to Table 4, it can be seen that the proposed model has the best segmentation performance compared to traditional segmentation models and recent segmentation models. In order to compare the segmentation results more intuitively, the segmentation effects of each segmentation algorithm on pathological images are shown in Figure 8.

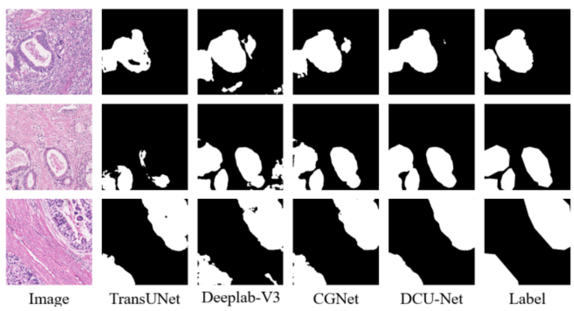

Fig. 8. Visualize segmentation results.

5. Conclusions

In the task of gastric cancer pathological image segmentation, we initially proposed two optimization structures. One is to add direct channels to each convolution layer of the encoding part to maximize the retention of detailed information in the image. Later, we tried to add an attention mechanism to the U-Net basic model, in order to ensure that the calculation process pays attention to image information to the greatest extent. After trying, it was determined to add the Coordinate Attention module before the up-sampling and down-sampling of each layer. Finally, after comparing the two optimization structures, it was proved through experiments that only when both optimizations are integrated into the basic model can the highest accuracy and IoU be obtained. After comparing the classic models, it was found that the model proposed in this question has the best performance. Excellent, the results can be obtained with an accuracy of 91.30% and IoU of 79.88%. In subsequent research, the number of pathological image samples in the dataset can be expanded, while further optimizing the existing structure to achieve more accurate segmentation results.

Acknowledgments

National Natural Science Foundation of China (62001196);
Jiangsu Province “333 High-level Talent Training Project” project (No.2022-3-4-107);
Changzhou Science and Technology Planning Project (CM20223015);
Changzhou Applied Basic Research Project (CJ20220064, CJ20220059).

References

1. Bray F, Laversanne M, Weiderpass E, et al. (2021) The ever-increasing importance of cancer as a leading cause of premature death worldwide. Cancer, 127(16):3029-3030.
2. Zheng R, Zhang S, Sun K, et al. (2023) Cancer statistics in China, 2016 [J]. Chinese Journal of Oncology, 45(3): 212-220.
3. Zhang G, Chen X, Zhao Y, et al. (2022) Lightweight YOLOv3 Algorithm for Small Object Detection. Laser & Optoelectronics Progress, 59(16)152-160.
4. Song Y. (2023) Application Progress of Artificial Intelligence Deep Learning in Gastric Cancer. Guide of China Medicine, 21(28)63-66.
5. Ronneberger O, Fischer P, Brox T. (2015) U-Net: convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer Assisted Intervention (MICCAI), 3:234-241.
6. Li W, Qian X, Ji J. (2017) Noise-tolerant deep learning for histopathological image segmentation. In 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 3075-3079.
7. Yi F, Yang L, Wang S, et al. (2018) Microvessel prediction in H&E Stained Pathology Images using fully convolutional neural networks. BMC bioinformatics, 19(1):1-9.
8. Azad R, Asadi-aghbolaghi M, Fathy M, et al. (2019) Bi-Directional ConvLSTM U-Net with densley connected convolutions. arXiv: image and video processing.
9. Wu T, Tang S, Zhang R, et al. (2020) CGnet: A Light-Weight Context Guided Network for Semantic Segmentation. IEEE Transactions on lmage Processing, 30:1169-1179.
10. Zhang Y, Li Y, HE D. (2023) U-shaped network with multi-scale feature extraction for lesion region segmentation in gastric cancer images. In 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), 234-238.