EURASIP Journal on Image
and Video Processing

RESEARCH Open Access

Large-scale tissue histopathology image segmentation based on feature pyramid

Pinle Qin, Jun Chen, Jianchao Zeng[*], Rui Chai and Lifang Wang

Abstract

Histopathology image analysis is a gold standard for cancer recognition and diagnosis. But typical problems with histopathology images that hamper automatic analysis include complex clinical features, insufficient training data, and large size of a single image (always up to gigapixels). In this paper, an image semantic segmentation algorithm based on feature Pyramid (ResNet50-GICN-GPP) is proposed. Firstly, the patch sampling method is adopted to resample the image, reduce the size of a single sample, and expand the number of training samples. Secondly, design the whole convolution network based on ResNet50 learning feature location information, then use GICN structure and deconvolution network to integrate multi-level features. Finally, in order to solve the problem that the GICN structure may lose the small object, the GPP structure should be joined to explore the multi-scale semantic information. The proposed method achieves 63% of the average segmentation accuracy (Dice coefficient) on Camelyon16 and Gastric WSIs Data, compared with U-Net, FCN and SegNet which has 10~20% improvement, and fully demonstrates the effectiveness of this method in different types of cancer. By experimentally comparing the segmentation accuracy of various scales of cancerous tissues, the performance of ResNet50-GICN-GPP is balanced and the multi-scale information localization is more accurate.

Keywords: Feature pyramid, Semantic segmentation, Convolution neural network, Multi-scale features, Multi-level features, Depth learning

1 Introduction

Medical image segmentation is the basis of various medical image applications, especially in computer-aided diagnosis (CAD), image-guided surgery (IGS) and oncology radiation therapy (ORT) [1]. In recent years, Whole-Slide Images (WSIs) [2] have driven the shift of pathological section to high resolution, where in all-digital workflows, the huge amount of information it contains provides a big data backdrop for quantitative analysis tasks (classification and segmentation). Combined with the rapid development in the field of computer vision, the quantitative analysis of pathological sections can not only save the doctors free from looking for candidate lesions one by one in a boring environment but also improve the accuracy of pathological diagnosis.

Segmentation of digital histopathology images presents three challenges: complex clinical features, inadequate training data, and the huge size of a single histopathological image.

The first one is the complex clinical features. Histopathology of different types of cancers show different morphologies, sizes, textures, and color distributions, making it difficult to find a general pattern for tumor image segmentation. Ilea et al. [3] studied an image segmentation algorithm that only extract texture features, and Tashk et al. [4] studied one that only extract texture features. Belsare et al. [5] proposed a hyper-pixel generation method based on similarity, combined with the text representation to form the space-texture-color map of breast histology images. Xu et al. [6] proposed an unsupervised SNMF algorithm, which is divided into two steps: color unmixing and spatial segmentation. However, the particularity of these designs limits the application of them in other types of tumor image segmentation.

* Correspondence: jianchaozeng@163.com
School of Data Science and Technology, North University of China, Taiyuan 030051, Shanxi, China

The second problem is the lack of large-scale medical image data. The number of images depends on the number of occurrences of the disease, which can make the collection process more difficult if the frequency of diseases studied is low. In addition, manual annotation of data requires a lot of workforce, but some clinical diagnosis is difficult to quantify, manual annotation is also not clear essentially.

The last problem is that a single histopathological image contains a huge amount of information. A typical WSI scan will produce an image with the size of more than 100,000 × 100,000 pixels that contain over one million descriptive objects. Due to the large-scale nature of the data, the feature extraction model is required both for efficiency and accuracy, and the learning algorithm should be designed to extract as much information as possible from these large images.

With the advent of deep convolutional neural networks (CNNs), CNN activation features have achieved great success in computer vision [7, 8]. There are many visual databases, such as ImageNet, with more than 10 million images and more than 20,000 categories [9], which enable CNN to learn a rich and varied feature description from these images. Xu et al. [10] studied the image abstraction provided by ImageNet in different responses of CNN hidden layer, transferred these features to histopathological images, then solved the problem of limited training data of medical images. Jia et al. [11] proposed an algorithm called DWS-MIL that takes multi-instance learning framework of full convolutional neural network (FCN) as the baseline and conducts multi-scale learning under weak supervision. The annotation process only requires a little extra work; then, the segmentation accuracy will be improved significantly. Although CNN itself is capable of image segmentation, it is unwise to segment the histopathology image with CNN directly. On the one hand, the size of a single histopathological image is extremely large, and it is impractical to construct CNNs with huge input size; on the other hand, scaling up the entire histopathological image to the acceptable input size of CNN will lose a great deal of details. Based on this fact, this paper uses the patch sampling technique [12] to sample the initial data and obtain the acceptable input data of CNN.

The work in the literature [7, 8, 10, 11] is mostly based on the classification network of the existed framework (AlexNet, VGGNet, GoogleNet, etc.). Peng et al. [13] found that the need of classification and segmentation tasks is contradictory. For classification tasks, the model needs to be invariant to various transformations, such as translation and rotation, but for the task of segmentation, the model needs to be sensitive to the transformation, that is, to locate the semantic category of each pixel precisely. For these reasons, this article attempts to design a new architecture to overcome these problems. Firstly, it should be designed on the principle of positioning (location) and should abandon the full connection layer and the pooling layer, because these layers will lose the spatial information. Then, use the large-size convolution kernel from the view of classification, to ensure the structure connect densely. The experiments in Section 4.2 show that when the convolution kernel size increased to the same size as characteristics (features), the classification effect is more obvious. Based on these two principles, this paper presents Global Inception Convolution Networks (GICN), as is shown in Fig. 2. To reduce the loss of contextual information in different sub-regions, this paper used the global average pooling [14] and built a global pyramid pooling (GPP) structure. Different size outputs in GPP contain information on different scales and different sub-regions.

In this paper, we proposed a simple and effective combination of feature pyramid for segmentation of histopathological images and verified the validity of the method on two data sets. The contribution of the framework has the following points:

(1) Sampling the initial data using patch sampling technology, solve the problem of the extremely large size of training data; it also added training samples.

(2) The GICN structure at different feature levels constitutes a multi-level feature pyramid that solves the contradiction between classification and location.

(3) The Global average pooling with different sizes constitutes a multi-scale feature pyramid that facilitates the integration of contextual information of different sizes and regions.

This paper is organized as follows: The second section briefly reviews the semantic segmentation algorithm in the field of natural images. The third section details the ideas and specific methods mentioned in this article. Finally, the fourth section provides the experimental results.

2 Segmentation algorithm of natural image domain

Before deep learning was applied to computer vision, researchers typically used methods such as TextonForest and Random Forest to build semantic-partitioned classifiers. Later, with the extensive application of deep learning, image block classification techniques utilize the image blocks surrounding each pixel to classify each pixel into a corresponding category. In 2014, Long et al. proposed FCN [15], which promoted the original CNN structure and spatially densely predicted without full connectivity layer. Compared with the image block classification, FCN can generate any size of the split map and improve the processing speed.

In addition to the full connected layer, another problem with image segmentation using the CNN network is the pooling layer structure. The pooling layer also discards part of the location information while increasing the upper layer convolution core and aggregated background information. However, the semantic segmentation method is sensitive to the position information of the feature map. To keep this information as far as possible, the researchers propose two structures to solve this problem.

2.1 Encoder-decoder structure

The encoder uses the pooling layer to reduce the spatial dimension of the feature map. The decoder gradually restores the target details and the corresponding spatial dimension through the deconvolution layer. For example, Noh et al. [16] use deconvolution to up-sample low-resolution feature responses. U-Net [17] connects the encoder features to the corresponding decoder though the jump layer connection to help the decoder recover the target details better. SegNet [18] recorded the pooling index of the feature map in the encoder. The decoder extracted the information and mapped it to its original location. LRR [19] uses Laplace Pyramid to reconstruct the network and multiplicative gating integrates effectively the underlying position information and higher-level semantic information. Recently, RefineNet [20] demonstrated the validity of the semantic segmentation problem based on the coder-decoder structure, which has also been practiced in the target detection.

2.2 Dilated convolutions structure

The use of dilated convolutions [21] instead of the pooling layer ensures that the original network, such as FCN, retains its original receptive filed and the size of the feature map that have not lose the information. PSPNet [22] uses dilated convolutions to improve the ResNet network and connect the feature graph of the ResNet to the upper sample output of the Pyramid parallel pool layer. DeepLab-V2 [23] imported different size images into extended convolution layers with different sampling rates to achieve semantic segmentation of multi-scale images.

In addition to the improvements in both of the above network architectures, some researchers used conditional random field (CRF) to improve segmentation in post-processing. The CRF method is a graph model of “smooth” segmentation based on the pixel strength of the underlying image, and the points of the pixel intensity are marked in the same category at runtime. Addition of conditional random field method can improve the final score of 1–2% [24].

3 Methods

Due to the lack of data sources and difficulties in data collection, medical image databases are often much less than the natural scene image data sets. It is not appropriate to apply the previous machine learning algorithms to the medical image data sets directly. Drawing on the successful semantic segmentation algorithm in natural scene images, we put forward a general solution to histopathological image segmentation in this paper. Firstly, we resample the data using patch sampling method. Secondly, we analyzed the contradiction between classification and segmentation, introduced new GICN structure using ResNet50 as the baseline method, and discussed the importance of spatial information to semantic segmentation. Finally, we proposed to use GPP structure to explore multi-scale semantic information aiming at the problem that the large convolution kernels may loss the small targets.

3.1 CNN architecture

ResNet showed good performance in image classification tasks once proposed. Experiments show that the residual network is easier to optimize and can improve the accuracy by increasing the depth of the network. Due to the limitation of the computing power of the experimental equipment, this paper selects ResNet50 as the basic method; uses cascading Res-2, Res-3, Res-4, and Res-5 for multi-level feature integration; and adds additional networks to the Res-5 for multi-scale feature integration. And finally, the two features are fused to get the prediction result. The overall network structure is shown in Fig. 1.

3.1.1 Resampling

For the large size of the WSI data, it is necessary for local feature extraction. This paper selected × 40 magnification data (about 151,872 × 151,872 pixels, the size of the data collected by different machines may be different), designed a rectangular grid with the size of 512 × 512 and the 512 strides, and obtained the patches traversing each WSI, as shown in Fig. 1a,b. In order to filter “bad data”, the RGB value of all the pixels in patch generation cannot be larger than 220(the experience value); otherwise, the patch data will be considered to contain only white background which should be discarded.

3.1.2 The multi-level feature pyramid

The ResNet50 input is a 3-channel image with the size of 512 × 512, shown in Fig. 4 in Appendix for more details. The full connection layer and the pooling layer in the original ResNet50 network will lose spatial information. Based on the principle of “positioning priority” in semantic segmentation, the network down-sample using a CNN with a stride size of 2 to replace the pooling layer. In the new Resnet50 structure, Res-2, Res-3, Res-4, and Res-5 form the encoder part of the entire structure. Then, in the point of classification, the

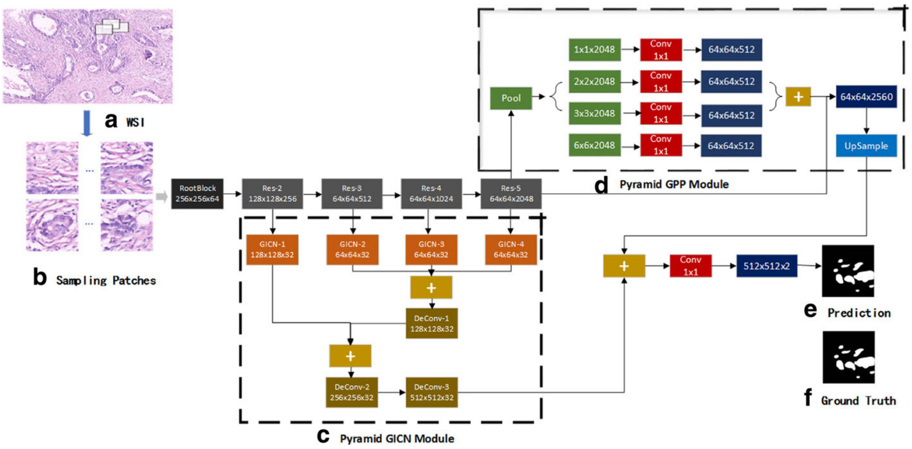


Fig. 1 The semantic segmentation network structure. **a** Digital histopathology images. **b** Sampling data for patch sampling (size 512 × 512). **c** Multi-level feature pyramid to extract spatial information. **d** Multiscale features pyramids focus on smaller targets. **e** Segmentation results, based on pixel-level prediction. **f** Ground Truth

encoder features need to be strongly connected to each pixel classifier to enhance the network's ability to handle various conversions. Figure 1c shows the Pyramid GICN Module structure in which the GICN network and the deconvolution layer together form the decoder section.

As mentioned before, the large CNN can establish dense connection; however, the large core structure brings the problems of high computational cost and too many structural parameters. In order to optimize the model and relieve the high computation, we designed

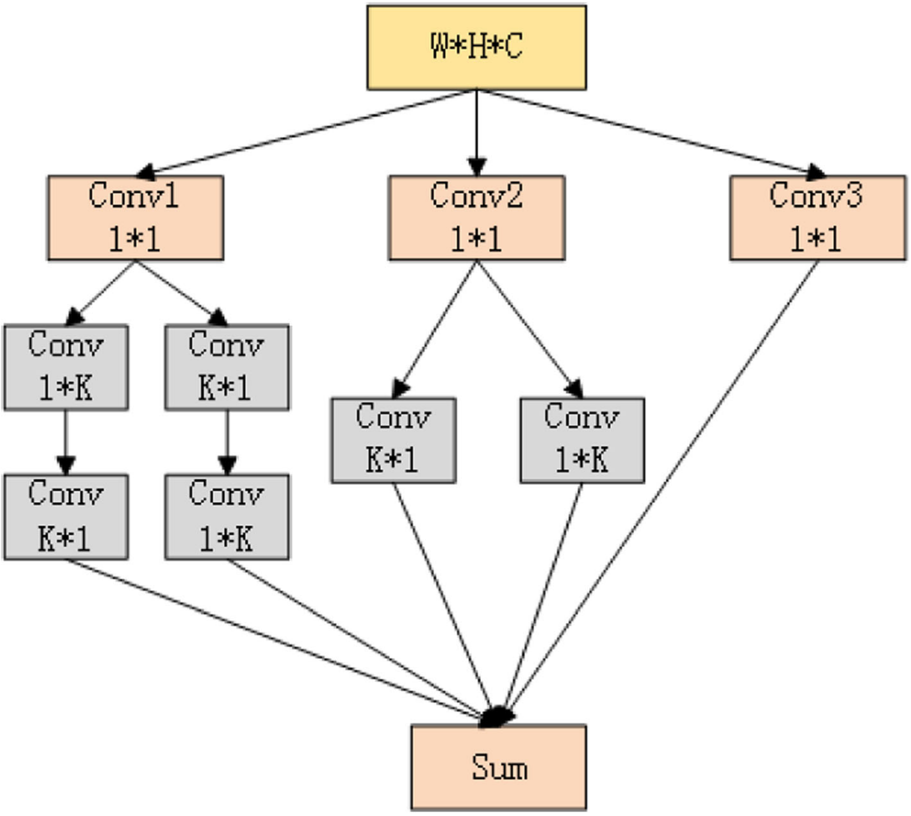


Fig. 2 GICN structure

the GICN structure with reference to the GoogleNet network model (as shown in Fig. 2) which replaces the large-core CNN network. The GICN structure is connected to the k^*k region on the encoder feature using $(1^*k + k^*1) + (k^*1 + 1^*k)$ in which compared with the single k^*k, the network does not need any nonlinear behind the convolutional layer with the just $O(\frac{3}{k})$computing cost and the count of parameters.

In the decoder part, four GICN structures are respectively connected to the corresponding decoder features, where the GICN-1 mapping with Res-2, GICN-2 mapping with Res-3, and so on. The GICN-1 feature layer is composed of 32 128 × 128 feature maps, and the GICN-2, GICN-3, and GICN-4 are all composed of 32 64 × 64 feature maps, in which it used the transpose convolution to up-sample with the kernel size of 3 × 3 and the strides 2. Deconv-1 outputs 32 128 × 128 feature maps. The features consisting of the DeConv-1 outputs and GICN-1 used transpose convolution to up-sample with convolution kernel size of 3 × 3 and stride size of 2. In the Deconv-2 output 32 256 × 256 features, the network generate the feature map with the same size of original image by using transpose convolution to the DeConv-2 output with the size of 2 × 2, and stride of 2. At this point, we completed the multi-level feature pyramid feature extraction.

3.1.3 Multi-scale features pyramid

This part of the work is shown in Fig. 1d Pyramid GPP Module. We took the 2048 64 × 64 features of the Res-5 module as input. Due to that the Pyramid GICN module is difficult to identify small and insignificant cancerous tissues, we proposed the GPP to make up for the semantic information deficiency of Pyramid GICN in small-scale target.

In Fig. 1, the green pool structure is actually a global average pooling layer. We covered the entire area of the Res-5, 1/4 zone, 1/9 zone, and 1/36 zone using pooling kernels with the size of four different scale (10 × 10, 20 × 20, 30 × 30, 60 × 60). The structure can not only obtain the features of different positions by sampling different sub-regions, at the same time, the output of different Pooling cores contains different size features. Next, the 1 × 1 CNN is used to reduce the dimensionality of the context features generated by each Pooling kernel, in which the uniform dimension is 512 and the sizes are 6 × 6, 3 × 3, 2 × 2, and 1 × 1, respectively. Finally, we linearly interpolated these dimensionality reduction features and output 512 64 × 64 feature maps on all four channels. Due to that the original image size is 512 × 512, we need to up-sample the GPP output. Firstly, we implement fast connection between the feature maps of Res-5 and GPP four-channel feature maps; Secondly, as shown in Fig. 1, the blue up-sampling structure up-samples in many times in the quick connection use transpose convolution with size of 3 × 3, stride of 2 and until feature size becomes 512 × 512.

Here we completed the feature extraction work of the multi-scale feature pyramid. The GPP output 64 512 × 512 feature map at last. We connected the GPP output with the output of the Pyramid GICN and use the 1 × 1 CNN dimension reduction to obtain two 512 × 512 probability maps as the segmentation results in this paper.

3.2 Training

The loss of the network training consists of two parts: one part is the loss of the regression location and the other is the classification loss. The total loss function can be expressed as:

$$L(y, c, P) = L_{conf}(y, c) + \partial L_{loc}(y, P) \tag{1}$$

Where a is the weight coefficient, set to 0.5, c is the confidence of each classification, y is the true value, and P is the predicted value. $L_{conf}(y, c)$is the loss of categorize confidence which used Softmax loss with multiple categories. $L_{loc}(y, P)$Describes the degree of similarity between the two models

$$L_{loc}(y, P) = 0.5 - \frac{|y \cap P| + k}{|y| + |P| + 2k} \tag{2}$$

Due to that some medical images do not have cancerous tissues, there will be the phenomenon of empty map, and the smoothing value k is introduced to correct the function. In this paper, we use k = 5e-4.

In addition, we used data augmentation to adjust the positive and negative sample ratios and to increase the number of training samples. The optimization function momentum = 0.9, weight decay 0.0001.

3.3 Evaluation

The experiment used the Dice coefficient to evaluate the model. The Dice coefficient is a set of similarity functions that evaluate the degree of similarity between two samples. Dice coefficient as (3) shows:

$$\text{Score} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3}$$

Where X is Ground Truth and Y is the predicted value. The $|X \cap Y|$ represents the count of intersection pixels in two samples. The $|X| + |Y|$ represents the sum of the pixels of Ground Truth and the predicted value. When X and Y are equal, the Dice coefficient is 1; when X and Y do not intersect at all, the Dice coefficient is 0. Therefore, the larger the Dice coefficient, the closer X and Y are, and the more accurate the segmentation.

Table 1 Camelyon16 segmentation results comparison

	FCN-VGG16	U-Net	SegNet	Resnet50-GICN (k = 7)	Resnet50-GPP	Resnet50-GICN-GPP (*k* = 7) (proposed method)
Score	46.55	37.4	50.82	59.18	61.33	63.70
Params	134 M	7 M	31 M	832 K	2619 K	3176 K
Test time	210 ms	73 ms	143 ms	69 ms	77 ms	102 ms
Stride	32	16	32	8	16	8

Score means segmentation accuracy (%). Params is the number of network model parameters. Test time means how long a 512 × 512 size image takes in the test. Stride means the maximum multiple of subsampling

4 Experiments and results analysis

In order to verify the effectiveness of the ResNet50 -GICN-GPP digital pathological image segmentation algorithm proposed in this paper, we will train the model on the Camelyon16 and Gastric WSIs Data to test the segmentation accuracy of the model, and compare it with other semantic segmentation algorithms. Experiments use the Tensorflow deep learning framework as a development and training tool. Inter (R) Xeon (R) CPU E5-2683 v3 @ 2.00GHZ dual-core processor machine is equipped with Ubuntu 14.04 operating system as a hardware experimental environment. Its memory is 256GB, and GPU processor is the NVIDIA M40.

4.1 Camelyon16

The Camelyon16 Challenge is the classification and location of pathological sections of breast cancer transference in the lymph nodes. The competition provided 110 pieces of tumor tissue and 130 pieces of normal tissue and marked the cancer area. This article focuses only on tumor data to achieve the segmentation of cancerous regions.

The maximum resolution × 40 image matrix of Camelyon16 is about 300,000 × 150,000, while a single sample requires 5G of storage space, beyond the PC's computing power. It is not practical to use CNN directly. Before training the model, it is necessary to segment the ROI region from 80% of the white background by using the technique in Section 3.1.1. After data preprocessing, over 700,000 sample data of 512 × 512 size and the corresponding label data were obtained.

Due to limited hardware environment, we finally randomly select 10 tumor data as a training set, two tumor data as a test set. The training set contains a total of 85,261 images, and 1000 k training iterations, with the initial learning rate of 5e-4, 10 k times per iteration and the learning rate reduce to 0.98 before.

Table 1 shows the segmentation accuracy of various segmentation methods. The Dice coefficient of ResNet50-GICN-GPP proposed in this paper reaches 63.7%, which is 10~20% higher than other methods, and is better than ever in model parameters and single test time.

Table 2 Scores of different *k*-sized GICNs on the test set

k	1	3	5	7	9	11	13
Score	58.3	60.0	62.6	63.5	64.2	64.0	64.4

4.2 Gastric WSIs Data

Gastric WSIs Data is a digital pathological sample of gastric cancer provided by the "Key Laboratory of Biomedical Imaging and Imaging Big Data of Shanxi Province". It is for routine HE staining at a magnification of × 20 and with a picture size of 2048 × 2048 pixels. The competition data is used for partial area of whole section. All data will be marked by pathology experts in the form of "double-blind assessment + validation". The data will be marked with or without cancer and the tumor area profile was drawn with lines (double-blind assessment + validation).

Gastric WSIs Data contains 100 patient cases with a total of 1000 pathological images in the ratio of training set to test set by 7: 3. To study the effect of different convolution kernel sizes on the accuracy of pathological image segmentation, the ResNet50-GICN model was trained on a data set. The network input is 512 × 512, with 58,000 training samples of the original data. The initial learning rate was set at 1e-4 for a total of 500 K iterations. The 380 K–460 K learning rate is reduced to 1e-5 and 460 k is set to 1e-6 later. The stochastic gradient descent method was used to train the network. After training the model, the average accuracy in the test set is as shown in Table 2.

The results in Table 2 show that as kernel size increases; score is also growing. However, the control group of the first experiment does not directly explain the performance improvement brought by GICN, so we designed the second experiment in this chapter. By replacing the CNN with kernel size *k***k* with the GICN structure, the model

Table 3 GICN and CNN experimental comparison, params means model training parameters

k	3	5	7	9	11
Score (GICN)	59.0	62.6	63.5	64.2	64.0
Score (CNN)	56.4	57.1	57.3	56.8	56.2
Params (GICN)	358 K	595 K	832 K	1069 K	1306 K
Params (CNN)	614 K	1302 K	1990 K	2678 K	3366 K

Table 4 Segmentation accuracy of model to multi-scale target

	FCN-VGG16	U-Net	SegNet	ResNet50-GICN ($k = 7$)	ResNet50-GPP	ResNet50-GICN-GPP ($k = 7$) (proposed method)
XS score	34.80	45.36	48.52	45.20	59.29	63.11
XL score	48.14	22.67	52.43	62.48	57.17	64.03

was trained on the experimental data I. After comparison, the experimental results are shown in Table 3.

Table 3 demonstrates that GICN works better than using a large kernel directly and requires fewer training parameters. Another purpose of this paper is to improve the segmentation effect of the model on cancerous tissues with different scales. Experiment III is designed, and the comparison results are shown in Table 4. This experiment set the connectivity area pixels and cancerous tissue less than 50,000 as small targets (XS representation), while others as large size targets (XL).

In Table 4, the higher the score is, the better the segmentation effect is. It can be seen that the ResNet50-GICN does not work well for small-size target segmentation, even

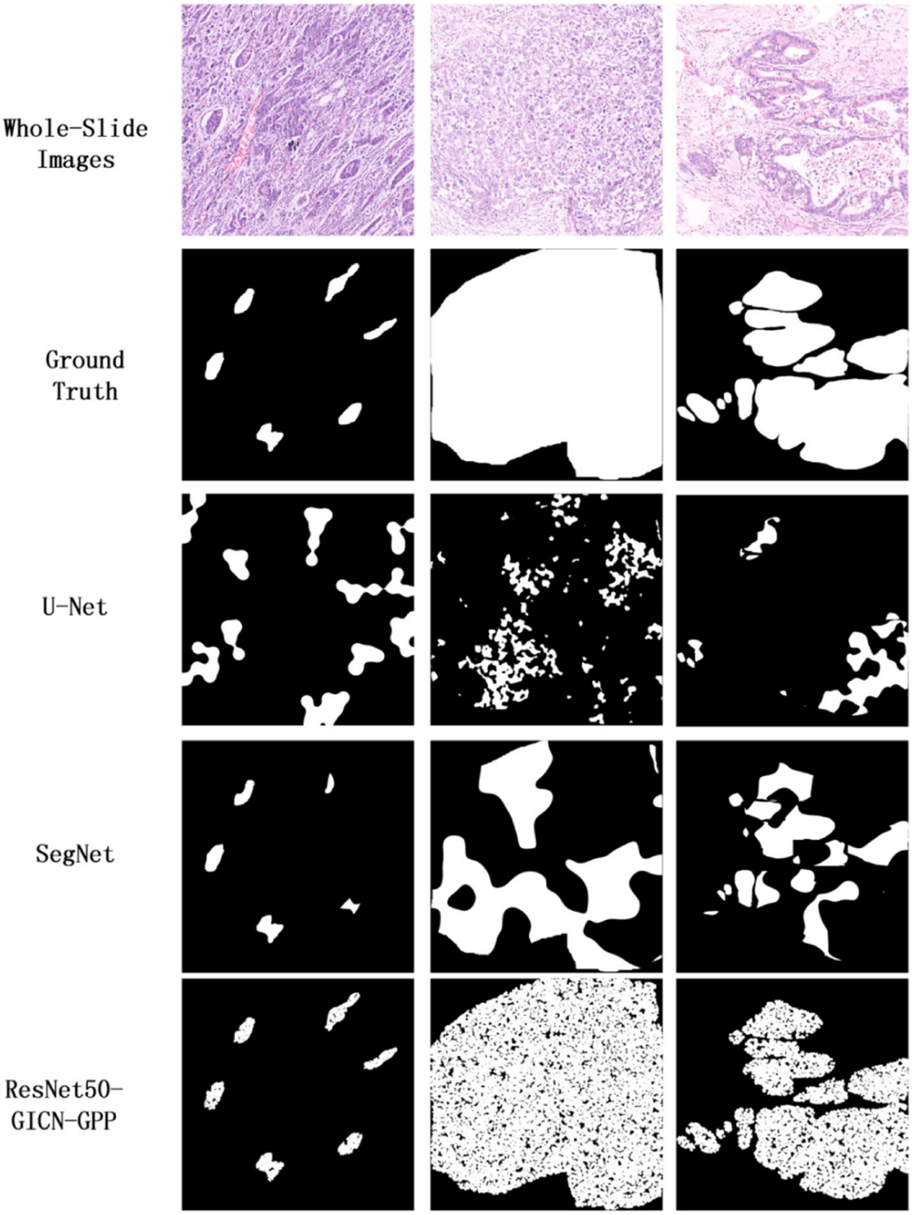


Fig. 3 Comparison of U-Net, SegNet and ResNet50-GICN-GPP segmentation results

lower than U-Net and SegNet. But it greatly improves on large-size target segmentation over FCN, U-Net and SegNet. ResNet50-GPP performs well on all scales of target segmentation. ResNet50-GICN-GPP connects the GICN and GPP features quickly, and the segmentation accuracy of the small object is significantly higher than that of the ResNet50-GICN. And the accuracy of the large object is also high, which shows the feasibility of the method in this paper.

Finally, in order to demonstrate the segmentation effect of different network architectures more intuitively, U-Net, SegNet and ResNet50-GICN-GPP models are trained on the experimental data I. The test results are shown in Fig. 3. Due to the shallowness of the U-Net network, the learning of the image space information is insufficient, and only a small part of the cancerous tissue can be segmented. When the image content becomes complicated, the segmentation accuracy is greatly descending. SegNet uses VGG-16-like network structure in the coding layer. Through the learning of the characteristic boundary information, SegNet is greatly effective in multiscale tasks of cancerous tissue segmentation. The last line in Fig. 3 is the segmentation result of ResNet50-GICN-GPP. Compared with the first two, it can segment more cancerous tissues and has better segmentation result and higher segmentation accuracy.

First row shows the medical image that we need to segment; the 2nd row shows the ground truth of the image segmentation result. From 3rd row to 5th row, they show the segmentation results of different methods, where ResNet50-GICN is the proposed method in this this paper.

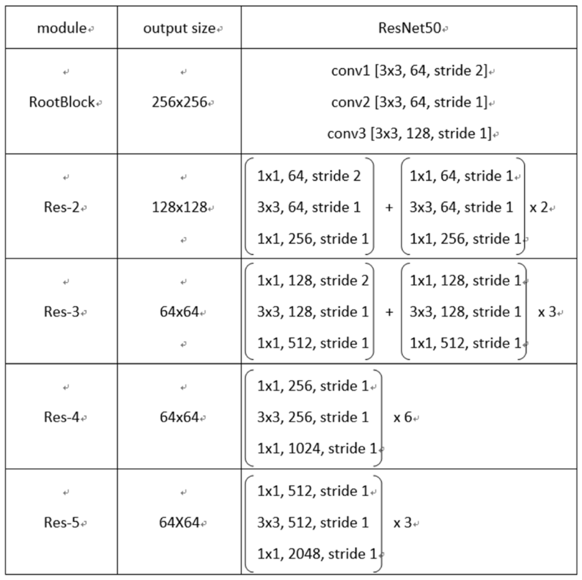
5 Results and discussion

Histopathology image analysis is a gold standard for cancer recognition and diagnosis. But typical problems with histopathology images that hamper automatic analysis include complex clinical features, insufficient training data, and large size of a single image (always up to gigapixels). In this paper, an image semantic segmentation algorithm based on feature Pyramid (ResNet50-GICN-GPP) is proposed. Firstly, the patch sampling method is adopted to resample the image, reduce the size of a single sample, and expand the number of training samples. Secondly, design the whole convolution network based on ResNet50 learning feature location information, then use GICN structure and deconvolution network to integrate multi-level features. Finally, in order to solve the problem that the GICN structure may lose the small object, the GPP structure should be joined to explore the multi-scale semantic information. The proposed method achieves 63% of the average segmentation accuracy (Dice coefficient) on Camelyon16 and Gastric WSIs Data, compared with U-Net, FCN and SegNet which has 10~20% improvement, and fully demonstrates the effectiveness of this method in different types of cancer. By experimentally comparing the segmentation accuracy of various scales of cancerous tissues, the performance of ResNet50-GICN-GPP is balanced and the multi-scale information localization is more accurate.

6 Conclusions

In this paper, we propose a digital pathology image segmentation algorithm based on feature pyramid, which integrates high-level semantic feature information and high-resolution low level location information, increasing the accuracy of classification and positioning, and making full use of the multi-level features of Pyramid. It also addresses the loss of semantic information for small objects in multi-level features; multi-scale feature pyramids are designed to extract global context information from high-level features. Through experimental comparison, the two feature pyramid models all help to improve the segmentation accuracy, and at the same time, the effect is better. The proposed method is tested on the Camelyon16 and Gastric WSIs Data datasets, and the average accuracy is higher than other methods, which fully demonstrates the effectiveness of the proposed method. The segmentation results in Section 4.1 are rough, and the next step is to introduce the CRF into the network and to train a new end-to-end network to further improve the accuracy.

1 Appendix



module	output size	ResNet50
RootBlock	256x256	conv1 [3x3, 64, stride 2] conv2 [3x3, 64, stride 1] conv3 [3x3, 128, stride 1]
Res-2	128x128	(1x1, 64, stride 2 3x3, 64, stride 1 1x1, 256, stride 1) + (1x1, 64, stride 1 3x3, 64, stride 1 1x1, 256, stride 1) x 2
Res-3	64x64	(1x1, 128, stride 2 3x3, 128, stride 1 1x1, 512, stride 1) + (1x1, 128, stride 1 3x3, 128, stride 1 1x1, 512, stride 1) x 3
Res-4	64x64	(1x1, 256, stride 1 3x3, 256, stride 1 1x1, 1024, stride 1) x 6
Res-5	64X64	(1x1, 512, stride 1 3x3, 512, stride 1 1x1, 2048, stride 1) x 3

Fig. 4 The structure of ResNet50 network with the shape of "1x1, 64, stride 2" indicates that the size of the convolution kernel is 1 × 1, the number of the output feature maps is 64, and stride 2 indicates the convolution step

Abbreviations

CAD: Computer-aided diagnosis; CNNs: Convolutional neural networks; FCN: Full convolutional neural network; GICN: Global Inception Convolution Networks; GPP: Global pyramid pooling; IGS: Image-guided surgery; ORT: Oncology radiation therapy; WSIs: Whole-Slide Images

Acknowledgements

Thank for the editor and reviewers.

About the authors

Pinle Qin received the PhD degree in computer application technology from Dalian University of Technology (DLUT), Dalian, Liaoning, P.R. China, in 2008. He is currently an associate professor with the School of Data Science and Technology, North University of China (NUC). His current research interests include computer vision, medical image processing and deep learning.
Jun Chen received his undergraduate degree in computer application technology from North University of China(NUC), Taiyuan, Shanxi, in 2014. Currently, he is pursing Master degree from NUC, and his areas of interest are digital image processing, medical image processing and computer vision.
Jianchao Zeng graduated from Xi'an Jiaotong University in 1985 and taught there, Ph.D., professor and doctoral supervisor, is a special allowance specialist of the State Council and vice president of North University of China(NUC). He is mainly engaged in the research of complex systems and community intelligence, intelligent computing, and health management of complex systems.
Rui Chai received the PhD degree in computer application technology from Beijing University of Technology (BUT), Beijing, P.R. China, in 2016. Currently, his research interests include image processing, medical image processing, and deep learning.
LiFang Wang received her undergraduate degree from Shanxi Normal University, Taiyuan, Shanxi, in 2000. Currently, she is pursing Master and Ph.D. degree from NUC, and her areas of interest are digital image processing and medical image processing.

Funding

The paper is subsidized by the Natural Science Key Project of Shanxi Province, China. No. 2015011045.

Availability of data and materials

Data will not be shared; reason for not sharing the data and materials is that the work submitted for review is not completed. The research is still ongoing, and those data and materials are still required by the author and co-authors for further investigations.

Authors' contributions

PQ and JZ designed the research. JC, RC, and LW analyzed the data. PQ wrote and edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

We approved.

Consent for publication

We agreed.

Competing interests

There are no potential competing interests in my paper. And authors have seen the manuscript and approved to submit to your journal. We confirmed that the content of the manuscript has not been published or submitted for publication elsewhere.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 17 May 2018 Accepted: 15 August 2018
Published online: 23 August 2018

References

1. J. Guipin, Q. Wenjian, Z. Shoujun, et al., State-of-the-art in medical image segmentation[J]. Chinese Journal of Computers **38**(6), 1222–1242 (2015)
2. R.S. Weinstein, A.R. Graham, L.C. Richter, et al., Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future[J]. Hum. Pathol. **40**(8), 1057–1069 (2009)
3. D.E. Ilea, P.F. Whelan, O. Ghita, *Unsupervised image segmentation based on the multi-resolution integration of adaptive local texture descriptors*.[J], vol 226 (2010), pp. 134–141
4. A. Tashk, M.S. Helfroush, H. Danyali, et al., A novel CAD system for mitosis detection using histopathology slide images[J]. J Med Signals Sens **4**(2), 139–149 (2014)
5. A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, et al., Breast histopathology image segmentation using spatio-colour-texture based graph partition method[J]. J. Microsc. **262**(3), 260 (2016)
6. J. Xu, L. Xiang, G. Wang, et al., Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis[J]. Comput Med Imaging Graph: the Official Journal of the Computerized Medical Imaging Society **46 Pt 1**, 20 (2015)
7. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks[C]// international conference on neural information processing systems. Curran Associates Inc., 1097–1105 (2012)
8. O. Russakovsky, J. Deng, H. Su, et al., ImageNet large scale visual recognition challenge[J]. Int. J. Comput. Vis. **115**(3), 211–252 (2014)
9. J. Deng, W. Dong, R. Socher, et al., ImageNet: a large-scale hierarchical image database[C]// computer vision and pattern recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 248–255 (2009)
10. Y. Xu, Z. Jia, L.B. Wang, et al., Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features[J]. Bmc Bioinformatics **18**(1), 281 (2017)
11. Z. Jia, X. Huang, E.I. Chang, et al., Constrained deep weak supervision for histopathology image segmentation[J]. IEEE Trans. Med. Imaging **PP(99)**, 1 (2017)
12. O. Frigo, N. Sabater, J. Delon, et al., Split and match: example-based adaptive patch sampling for unsupervised style transfer[C]// computer vision and pattern recognition. IEEE, 553–561 (2016)
13. Peng C, Zhang X, Yu G, et al. Large Kernel Matters -- Improve Semantic Segmentation by Global Convolutional Network[J]. 2017
14. Lin M, Chen Q, Yan S. Network in network[J]. Computer Science, 2013
15. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation[C]// IEEE conference on computer vision and pattern recognition. IEEE computer Society, 3431–3440 (2015)
16. H. Noh, S. Hong, B. Han, Learning Deconvolution network for semantic segmentation[C]// IEEE international conference on computer vision. IEEE Computer Society, 1520–1528 (2015)
17. O. Ronneberger, P. Fischer, T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation[M]// Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015* (Springer International Publishing, 2015), pp. 234–241
18. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for scene segmentation.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence **PP(99)**, 1 (2017)
19. G. Ghiasi, C.C. Fowlkes, *Laplacian Pyramid Reconstruction and Refinement for Semantic Segmentation[C]// European Conference on Computer Vision* (Springer, Cham, 2016), pp. 519–534
20. G. Lin, A. Milan, C. Shen, and I. Reid. Refinenet: Multipath refinement networks with identity mappings for high resolution semantic segmentation. arXiv:1611.06612, 2016
21. F. Yu, V. Koltun, in *ICLR*. Multi-Scale Context Aggregation by Dilated Convolutions (2016)
22. Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[J]. arXiv:1612.01105, 2016
23. L.C. Chen, G. Papandreou, I. Kokkinos, et al., DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs[J]. IEEE Trans Pattern Anal Mach Intell **PP(99)**, 1 (2016)
24. S. Zheng, S. Jayasumana, B. Romera-Paredes, et al., *Conditional Random Fields as Recurrent Neural Networks[J]* (2015), pp. 1529–1537