



# Semi-automatic Segmentation of Tissue Regions in Digital Histopathological Image

Xin He<sup>1</sup> , Kairun Chen<sup>2</sup> , and Mengning Yang<sup>1</sup>  

<sup>1</sup> School of Big Data and Software Engineering, Chongqing University, Chongqing, China  
mnyang@cqu.edu.cn

<sup>2</sup> Chongqing Meehoo Technology Co., Ltd., Chongqing, China

**Abstract.** Segmentation of tissue regions in the digital histopathological images refers to the identification and segmentation of tissues such as epithelium, glandular cavity, fibers, etc. Precise segmentation of tissues is key to pre-determining the regions with the greatest diagnostic value, which can support clinical diagnosis, particularly with regard to etiology and severity. In view of the uneven quality of histopathological images and the difficulty of manual segmentation. In this paper, an approach based on weakly supervised learning and deep learning has been proposed to build a semi-automatic segmentation model of tissue regions. The model uses superpixel classification to pre-segment the tissues, the tissue region boundary is preserved, and the automatic segmentation of the tissues is finally achieved based on the deep convolutional neural network. The effectiveness of the model is evaluated on 600 cervical histopathology images provided by the hospital. The results show that the proposed method achieves 82.52% mean IoU of epithelial segmentation and 81.67% mean IoU of glandular lumen segmentation in cervical histopathological images. The model is superior to traditional manual feature representation methods and classical deep convolution neural network methods in segmentation accuracy and efficiency.

**Keywords:** Digital histopathological image · Tissues segmentation · Superpixels · Deep convolutional neural network

## 1 Introduction

Digital histopathological images are obtained by making stained sections of patients' suspected lesion tissues and then photographing them in high resolution using whole slide imaging (WSI) technology. They are analyzed and judged by the pathologists, and this histopathological examination process has become the

“gold standard” of cancer detection and diagnosis. Histopathological images can be used by doctors to know about the status of tumor cells, which is of great significance for the diagnosis, classification, and prognosis of tumors. These images include many tissues such as stroma, epithelium, and glandular lumen, etc. Different tissues are closely related to the type and severity of the disease. For example, the diagnosis of cervical squamous cell carcinoma is often based on the proliferation and arrangement of cells in cervical epithelial tissue [1]. Therefore, it is important for pathologists to focus on the tissue region with the greatest diagnostic value in histopathological images before judging the disease. However, the traditional manual segmentation method is limited by time-consuming, unstable, strong subjectivity, and high error rate due to visual fatigue. Developing an intelligent automated algorithm to efficiently segment different tissues is crucial for the development of pathology-assisted diagnosis [2,3].

The field of histopathological image segmentation faces the following challenges: (1) Compared with natural images, histopathological images have unclear semantic regions, unobvious boundaries and high similarity among tissues, and the analysis of histopathological images is more challenging. (2) Histopathological images are usually super-resolution images, which have greater computational complexity. (3) Lack of large-scale annotated datasets [4], the lesion dataset changes greatly and needs to be labeled by clinical experts. Therefore, classical segmentation algorithms for natural images, such as threshold segmentation, region growth and edge detection, are not fully applicable to histopathological images. In order to segment histopathological image tissue regions, this paper starts with the accurate dividing of the tissue boundaries, superpixels [5] have been shown to be able to efficiently and completely segment local regions. Convolutional neural networks with excellent performance in image processing are used to classify superpixel in an inexact supervised learning way. To simplify operations and further improve segmentation accuracy, the obtained tissue images are used to make deep learning datasets, an automatic segmentation model is trained end-to-end based on a deep convolutional neural network, so as to achieve faster and more accurate segmentation of tissue regions. The innovation of this article lies in:

1. Improve the accuracy and interpretability of the model through the key human-computer interaction.
2. Superpixel-level labeling replaces the pixel-level labeling, which improves the labeling efficiency and preserves tissue boundaries.
3. The semi-automatic segmentation model has strong universality and can be applied to segmentation tasks of various tumor histopathological images.

The organizational structure of the paper is as follows: The second section reviews the related work of tissue region segmentation in histopathological images. The third section introduces preliminaries, including image preprocessing and semi-automatic segmentation methods. In the fourth section, the experimental results are discussed and analyzed, and the effectiveness of the method is verified. The last section gives conclusion and future research contents.

## 2 Related Work

In the past few decades, with the mature development of image scanning technology and the improvement of computing power, as well as the emergence of automatic analysis algorithms, significant progress has been made in histopathology image segmentation. To deeply explore the tissue segmentation methods of histopathological images, this section sorts out and analyzes the effective techniques which are used to segment different interest objects in histopathological images. Research on these techniques can be broadly divided into two main categories: methods based on hand-crafted features and methods based on deep learning.

### 2.1 Methods Based on Hand-Crafted Features

Support vector machine (SVM) model based on local binary patterns (LBP) was used to automatically distinguish epithelial and stromal in digitized tumor tissue microarrays (TMAs) of colorectal cancer [6], the image is narrowed and divided into square blocks, and the blocks are then independently classified using an SVM model. Also a bayesian model was used to automatically segment the stromal tissue in the immunohistochemical (IHC) image based on color and texture features [7]. The patch was classified by the deep learning method [8], and patch level statistics and morphological characteristics were input into the random forest (RF) regression model to classify the whole slide image. All of the above approaches are based on hand-crafted feature representation, the segmentation precision is unsatisfactory due to high computation and limited feature extraction. In recent years, artificial intelligence technologies such as deep learning have made breakthroughs in various fields. More and more researchers have turned their attention to the application of deep learning in histopathological images and achieved outstanding results.

### 2.2 Methods Based on Deep Learning

Convolutional Neural Networks (CNNs) perform well in histopathological image processing [9]. The supervised classification of a CNN was combined with unsupervised image segmentation to distinguish the epithelial and stromal tissues of H&E images [10], the combination of deep learning and boundary localization improved boundary segmentation accuracy, but the performance of this method was limited for images with fuzzy boundaries. De [11] proposed a method for segmenting renal tissue using CNN, experiments were conducted using three different network architectures, with about 90% accuracy. Nirschl [12] provided a deep learning framework for the segmentation of muscle cells and stroma in H&E stained heart biopsy samples. The framework uses AlexNet architecture to train pixel-level classifiers for segmentation. Compared with a random forest classifier with 333 intensity and texture features, the framework is superior in AUC and F-score.

Fully Convolutional Networks (FCN) has made significant progress in image semantic segmentation with the advantages of unlimited input size and end-to-end training of the model [13]. Chen [14] segmented the colon glands based on the Deep Contour-Aware Networks (DCAN). DCAN adopts an auxiliary supervision mechanism to overcome the problem of gradient disappearance when training. This method ranked first in the 2015 MICCAI glandular segmentation challenge and 2015 MICCAI nuclear segmentation challenge. Lahiani [15] used an end-to-end color deconvolution deep learning method to segment tissues in multi-staining immunohistochemical images, digital histopathological images with multiple staining effects can be automatically segmented into the tumor, healthy tissue, necrotic region and background based on a FCN, however, the scheme is difficult to judge when the image source is unknown or the imaging quality is poor.

U-Net as the baseline of medical image processing methods [16], many scholars have processed and analyzed medical images based on U-Net or its improved version. A group of sparsely annotated histopathological images was used to train U-Net and FCNs of different depths [17], and the pixel-based AUC score was 0.97. However, this method has a large network scale, many parameters and a long training time. Based on the U-Net architecture, a positive predictive value of  $0.89 \pm 0.16$  and a sensitivity of  $0.92 \pm 0.1$  were obtained in the epidermal or non-epidermal pixel classification task [18], but the entire epidermis region cannot be divided.

Other methods based on deep learning are also used in tissue segmentation of histopathological images. In [19], HistoSegNet is proposed for the semantic segmentation of tissues in the histopathological image. It is superior to the more complex weakly supervised semantic segmentation method and can be extended to other datasets without retraining. A general neural network method is designed to segment disease-related regions in medical images [20], it requires only two types of tags at the sample level. Using the labeled samples to train a meta-network, which deduces a segmented neural network to segment the disease-related regions in the image, and identify tumor regions or tumor-free regions reliably. However, the diversity of training data needs to be improved.

It can be known from literature research that dataset, segmentation accuracy, and model portability are still the difficulties in tissue segmentation. Therefore, we constructed a semi-automatic segmentation model suitable for various tumor image tissues based on pathological knowledge and deep learning. Semi-automatic refers to obtaining high-quality segmentation samples through the interactive pre-segmentation model, which can be used to train the tissue automatic segmentation model end to end.

### 3 Preliminaries

This section introduces the image preprocessing, and describes in detail the semi-automatic segmentation methods, including pre-segmentation and automatic segmentation.

### 3.1 Methodology Overview

The semi-automatic segmentation model of tissues is divided into two stages: pre-segmentation based on superpixel classification and automatic segmentation based on deep CNNs. The pre-segmentation process yields a large number of high-quality tissue segmentation results, providing a dataset quickly for automatic segmentation model in the next stage. Automatic segmentation utilizes the deep learning method to learn the segmented image dataset end-to-end and ultimately achieves the automatic segmentation of tissues. The semi-automatic segmentation model greatly reduces the burden of manual segmentation and improves the accuracy of the automatic segmentation of tissue regions. The segmentation process is shown in Fig. 1.

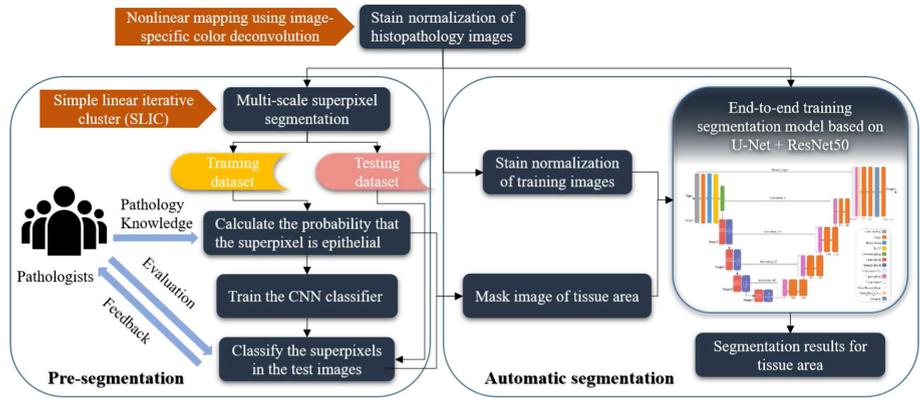


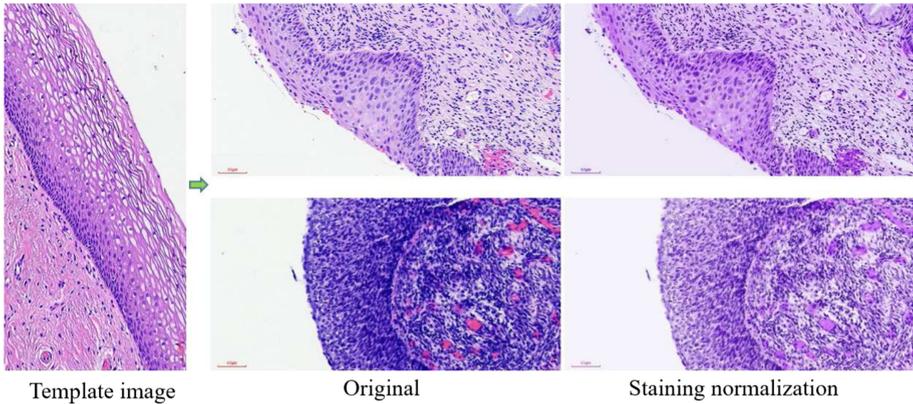
Fig. 1. Segmentation process of tissue region in histopathological image.

### 3.2 Histopathological Images Preprocessing: Staining Normalization

The imaging of histopathological images is influenced by factors such as staining degree and scanning equipment. Deep learning algorithms are extremely sensitive to the color structure of images, so it is necessary to normalize image to the color distribution of the template image to reduce the variance. Based on the representation derived from color deconvolution, the nonlinear mapping from the source image to the template image is found, and the staining normalization is realized [21]. It is a spectral normalization method that converts all images into a spectral distribution of the template image (see Fig. 2).

### 3.3 Pre-segmentation of Tissue Regions

**Multi-scale Superpixel Segmentation.** Superpixel [5], first proposed by Ren and Malik, is a clustering-based segmentation algorithm, which clusters a series



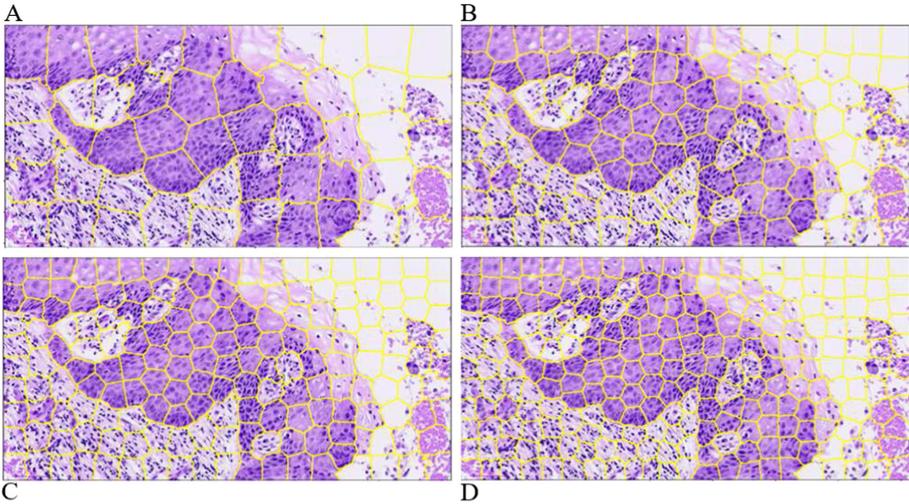
**Fig. 2.** Stain normalization results of histopathological images.

of pixels that are adjacent to each other and have similar color, brightness, texture into small regions. The simple linear iterative clustering (SLIC) method proposed by Achanta [22] is simple, fast in running, and capable of generating uniformly distributed and compact superpixels. SLIC method is used to segment superpixels with different scales in this paper, the multi-scale facilitates the model to reconcile the labeling cost and segmentation effect. The multi-scale superpixel segmentation is shown in Fig. 3.

**Rectangularization and Labeling of Superpixel.** Superpixels are rectangularized to be able to feed them into the CNN for classification. The external rectangle of the superpixel is obtained by topologically analyzing the superpixel mask. According to Table 1, the preset size of the superpixel blocks is cropped out in the external rectangle, where  $SP_{number}$  indicates the number of superpixels segmented in an image ( $1430 \times 712$ ),  $SP_{size}$  indicates the size of the superpixels after rectangularization.

**Table 1.** Predefined size for cropping superpixel rectangle.

$SP_{number}$	$SP_{size}$
250	64
180	86
130	100
50	128
40	156



**Fig. 3.** Superpixel segmentation. The images are divided into 50(A), 130(B), 180(C) and 250(D) superpixels.

Based on Eqs. 1 and 2, the coordinates  $(x_-, y_-)$  of the top-left point of the rectangularized superpixel block are calculated.  $x, y$  are the top-left point coordinates of the external rectangle, and  $W, H$  are the width and height of the external rectangle. Starting from the top-left point, pixels are taken to the right and down, and reverses when the boundary is encountered. The pseudocode is shown in Table 2. The results are shown in Fig. 4.

$$x_- = x - \lceil 1/2(SP_{size} - W) \rceil \tag{1}$$

$$y_- = y - \lceil 1/2(SP_{size} - H) \rceil \tag{2}$$

Superpixel labeling is done by people trained in basic pathology, the probability that each superpixel rectangular belongs to a specific tissue is determined, which is used for inexact supervised learning. As shown in Eq. 3, where  $K$  is the number of superpixels.

$$p_i \approx \text{area}(\text{target tissue}) / \text{area}(\text{superpixel}) \tag{3}$$

$$p_i \in [0, 1], i \in K$$

**Pre-segmentation Architecture.** CNN is constructed to train the superpixel classification model in an inexact supervised learning way, the network structure is shown in Fig. 5. The classification of superpixels is the initial segmentation of tissues. The process is shown from 4 to 12.

$$a^1 = \text{Relu}(X * W^1 + b^1) \quad n\_kernel = 20, size\_kernel = 5 * 5 \tag{4}$$

$$a^2 = \text{pool}(a^1) \quad size\_pool\_kernel = 2 * 2, stride = 2 * 2 \tag{5}$$

**Table 2.** Pseudocode for the method of superpixel rectangularization.

---

**Input:** A histopathological image  $img$ , the top-left point coordinates  $(x,y)$ , width  $W$  and height  $H$  of the superpixel external rectangle.

**Output:** Rectangularized superpixel  $roi$ .

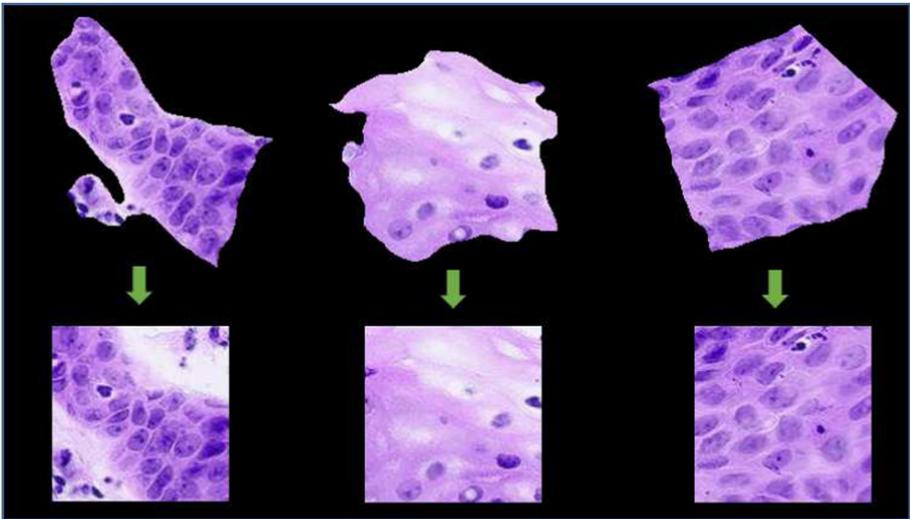
---

```

 $x_- \leftarrow x - \lceil (SP_{size} - W)/2 \rceil$ 
if  $x_- < 0$ :
     $x_- \leftarrow 0$ 
elif  $x_- + SP_{size} > img.shape[0]$ :
     $x_- \leftarrow img.shape[0] - SP_{size}$ 
 $y_- \leftarrow y - \lceil (SP_{size} - H)/2 \rceil$ 
if  $y_- < 0$ :
     $y_- \leftarrow 0$ 
elif  $y_- + SP_{size} > img.shape[1]$ :
     $y_- \leftarrow img.shape[1] - SP_{size}$ 
 $roi \leftarrow img[x_- : x_- + SP_{size}, y_- : y_- + SP_{size}]$ 

```

---

**Fig. 4.** The results of superpixel rectangularization.

$$a^3 = Relu(a^2 * W^3 + b^3) \quad n\_kernel = 40, size\_kernel = 4 * 4 \quad (6)$$

$$a^4 = pool(a^3) \quad size\_pool\_kernel = 2 * 2, stride = 2 * 2 \quad (7)$$

$$a^5 = \text{flatten}(a^4) \tag{8}$$

$$a^6 = \text{softmax}(W^6 * a^5 + b^6) \tag{9}$$

$$a^7 = \text{dropout}(a^6) \tag{10}$$

$$a^8 = W^8 * a^7 + b^7 \tag{11}$$

$$\hat{y} = \text{argmax}(a^8) \tag{12}$$

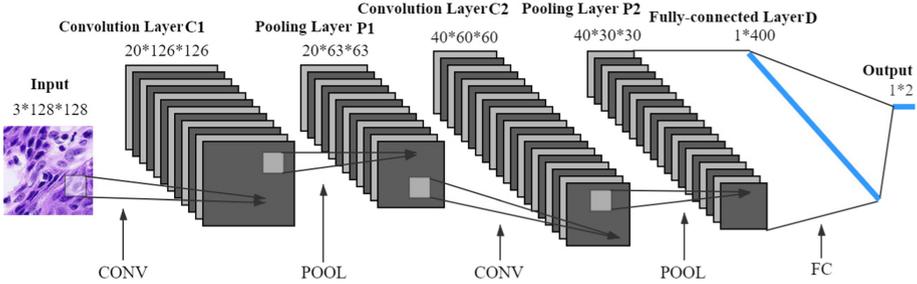


Fig. 5. Convolutional neural network structure.

where the input  $X$  is a three-dimensional superpixel image, and  $a^1, a^3$  are the outputs of the convolutional layers, respectively. *Relu* is the activation function of the convolutional layer,  $n\_kernel$  is the number of convolutional kernels, the size of convolutional kernel is  $size\_kernel$ . *flatten* function makes the  $a^4$  matrix into a one-dimensional vector  $a^5$  for input to the fully connected layer.  $a^2, a^4$  is the output of max-pooling layer,  $size\_pool\_kernel$  is the size of pooling kernel,  $stride$  is the step size.  $a^6$  is the fully connected layer output after activation, the activation function is the *softmax* function,  $a^7$  is the result of discard units from the network with a certain probability to prevent overfitting.  $a^8$  is the output of fully connected layer with no activation, finally, it returns the subscript  $\hat{y}$  of the maximum value in  $a^8$ , which is the output. The cross-entropy loss is minimized to train superpixel classification model, which is defined as Eq. 13.

$$L = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \tag{13}$$

where  $y^{(i)}$  is the real category of the superpixel,  $\hat{y}^{(i)}$  is the classification of superpixel by model,  $L$  represents the difference between the predicted output and the real category.

### 3.4 Automatic Segmentation of Tissue Regions

**Obtaining Ground Truth for Automatic Segmentation.** The mask obtained by binarization of the pre-segmented result is taken as ground truth, the tissue part of the RGB is 255. The input and label images are data enhanced. Each image is cut to half size of the original image. The staining normalized image is the input and the mask of the corresponding tissue is the label.

**Automatic Segmentation Architecture.** Both low-level and high-level features of images are important for tissue segmentation. Skip-connection and U-structure of the U-Net enable learning of both high-level and low-level features. An improved U-Net architecture for replacing the VGG Net [16] with the first four layers of the ResNet50 is used to avoid gradient disappearance for deep network training, which is called U-Net+ResNet50 in this paper, U-Net+ResNet50 structure is shown in Fig. 6.

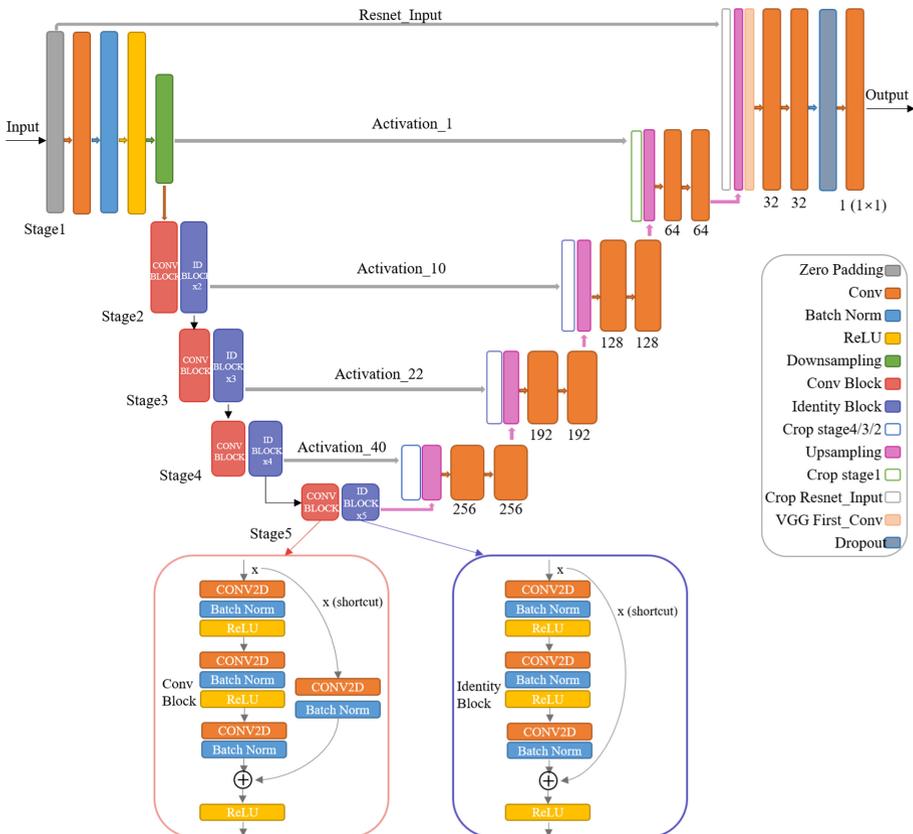


Fig. 6. U-Net + ResNet50 structure.

The ResNet50 downsampling process includes Conv Block and Identity Block, the role of Conv Block is to change the dimensionality of the feature vector. Identity Block with consistent input and output dimensions for identity mapping. The bottom left Conv Block is used to handle mismatches between input images and output dimensions, where the convolutional layers on the shortcut connection is used to adjust the dimensionality of the input, ensuring that inputs and outputs on the main path can be summed. The shortcut connection in the bottom right Identity Block spans three hidden layers. In addition, the U-Net+ResNet50 adopts the “same” mode for zero-padding filling at the edge of the image. Batch normalization is performed between each convolutional operation and the activation function to prevent overfitting to some extent. Max-pooling is used for the downsampling of pathological features and bilinear interpolation for upsampling.

Automatic segmentation model is trained based on the U-Net + ResNet50, the input is staining normalized images, the output is tissue masks. The loss function *Bce\_Dice\_Loss* for end-to-end training consists of two parts: *Dice loss* and *Binary crossentropy loss*. *Dice loss* describes the similarity of two contour regions, denoted by  $A$  and  $B$  as the set of pixels contained in two contour regions of category  $y, \hat{y}$ . The *Dice loss* is derived from  $DSC(A, B)$ , as shown in Eq. 14, where  $p$  and  $r$  are defined as shown in Eqs. 15 and 16.

$$\begin{aligned}
 \text{Dice loss} &= 1 - DSC(A, B) \\
 &= 1 - \frac{\sum_{n=1}^N p_n r_n + \varepsilon}{\sum_{n=1}^N p_n + r_n + \varepsilon} - \frac{\sum_{n=1}^N (1 - p_n)(1 - r_n) + \varepsilon}{\sum_{n=1}^N 2 - p_n - r_n + \varepsilon} \tag{14}
 \end{aligned}$$

$$p = TP / (TP + FP) \tag{15}$$

$$r = TP / (TP + FN) \tag{16}$$

where  $TP, FP, FN$  are the number of true positives, false positives, and false negatives, respectively,  $p_n$  is the accuracy rate,  $r_n$  is the recall rate, and  $\varepsilon$  is the smoothing parameter. The *Binary crossentropy loss* is defined in Eq. 17.

$$\text{Binary crossentropy loss} = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \tag{17}$$

The sum of the Dice loss and the *Binary crossentropy loss* is taken as loss function, *Bce\_Dice\_Loss* is shown in Eq. 18.

$$Bce\_Dice\_Loss = Binary\ crossentropy\ loss + Dice\ loss \tag{18}$$

## 4 Experiments and Results Analysis

### 4.1 Experimental Objective

1. Different hyperparameters are set to optimize pre-segmentation and automatic segmentation models respectively.

- Four widely used methods are compared with our proposed method to measure the segmentation performance of our method.

## 4.2 Dataset

**Pre-segmentation Dataset.** Cervical histopathological image dataset provided and authorized by the hospital. All personal information is withheld to protect patient privacy. Pre-process 600 cropped images, they are divided into superpixels. Data enhancement of labeled superpixels by rotating, flipping, adding noise. Finally, 22,032 superpixels are obtained as a superpixel classification dataset (SCD).

**Automatic Segmentation Dataset.** The automatic segmentation dataset (ASD) is constructed based on the pre-segmentation results. The ASD consists of 1662 staining normalized images and corresponding tissue masks (ground truth).

## 4.3 Experimental Setup

The experimental settings of pre-segmentation and automatic segmentation are the same, and the positive and negative samples of the dataset are balanced, of which 90% is taken as a training set, the remaining 10% is taken as a validation set, real-time samples from the hospital as a test set. Each evaluation metric is cross-validated by a 5-fold cross-validation and the final results averaged. All experiments were performed on an Amax NVIDIA Titan V server with a 12G GPU.

The methods widely used in image segmentation are selected for comparison with our method, SVM-RBF [23] and Random Forest (RF) [24], as the most commonly used and better performing segmentation methods based on manual feature, are used to verify the limitations of classical traditional methods in pathological image processing. FCN [13] and U-Net [16] are widely used deep learning segmentation methods, and U-Net is also the baseline of the proposed method.

## 4.4 Experimental Results and Analysis

**Metrics of Pre-segmentation Model.** Superpixel classification is evaluated by Mean Cross-Entropy (MCE) Loss and Accuracy. *MCE Loss* and *Accuracy* are defined as shown in Eqs. 19 and 20. *MCE loss* characterizes the difference between the predicted output and the true label. *TP, TN, FP, FN* in Eq. 20 are elements of the confusion matrix. *Accuracy* can partly indicate whether the classifier is effective.

$$MCE\ Loss = -(1/N) \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \quad (19)$$

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (20)$$

Pathology-trained personnel performs a rapid evaluation of the superpixel classification results, and makes continuous improvements in the input superpixel size and network hyperparameters until an optimal classification model is obtained. Finally, the superpixel classification results are reorganized into tissue segmentation results according to the division rules and categories.

Pre-segmentation includes superpixel segmentation and rectangularization. In addition, a histopathological image is divided into superpixels for independent classification, ignoring the correlation between the superpixels in close positions. In view of this, the contents of the image are directly learned based on deep learning to avoid superpixel segmentation. It simplifies the production of the dataset and improves the accuracy of tissue segmentation by learning the overall features.

**Pre-segmentation Results of Tissue Region.** The classification of the superpixels based on the CNN is shown in Table 3, where *Epochs* is the number of training rounds and Batch Size (BS) is the batch size of each input data during training. It is clear that at *Epochs* of 24 and *Batch Size* of 20, *Mean Loss* and *Accuracy* are iterated to optimal, achieving 85% classification accuracy.

**Table 3.** Classification results of superpixels with size 128 \* 128.

	Epochs = 30	Epochs = 24	Epochs = 23		Epochs = 19
	BS = 23	BS = 20	BS = 10	BS = 25	BS = 30
Mean loss	0.2404	<b>0.2350</b>	0.3200	0.2620	0.2683
Accuracy	0.8472	<b>0.8489</b>	0.8358	0.8466	0.8471

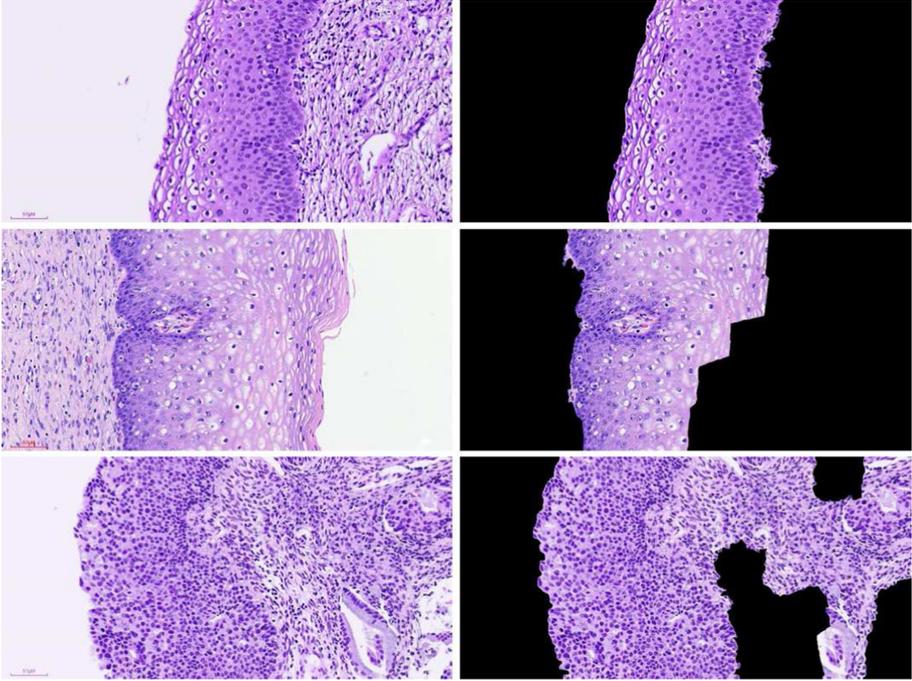
Experiments also verify the effect of different sizes of superpixels. If a superpixel is too small, and contains too little contextual information, which will lead to poor classification accuracy. While it is too large to properly segment tissue boundaries. As shown in Table 4, the classification model achieved more accurate results when the superpixel size is 128 × 128, that is, each pathological image is divided into 50 superpixels.

**Table 4.** Classification results of superpixels with different sizes.

	Size = 64 × 64	Size = 86 × 86	Size = 100 × 100	Size = 128 × 128	Size = 156 × 156
Mean loss	0.2855	0.2820	0.2649	<b>0.2350</b>	0.2417
Accuracy	0.8213	0.8275	0.8311	<b>0.8489</b>	0.8401

The superpixels are recombined base on categories to obtain the segmentation result of the tissue, which is visually shown in Fig. 7. Pre-segmentation provides

a better segmentation of images with clear borders. However, for images with more complex staining distribution and more disordered cell arrangement, the pre-segmentation model has a greater error.



**Fig. 7.** Pre-segmentation results of tissue regions in histopathological images.

**Metrics of Automatic Segmentation Model.** *Accuracy* and the *Mean\_IoU* as evaluation metrics of models. They are defined as shown in Eqs. 21 and 22, respectively. The Intersection-over-Union (IoU) refers to the ratio of intersection and union between the target region generated by the model and the originally marked region. In Eq. 22,  $P$  and  $G$  represent the predicted and ground truth, and  $N$  is the number of samples.

$$Accuracy = (TP + TN) / ((TP + FP) + (TN + FN)) \quad (21)$$

$$\begin{aligned} Mean\_IoU &= (1/N) * (area(P) \cap area(G)) / ((area(P) \cup area(G))) \\ &= (1/N) * TP / (FN + TP + FP) \end{aligned} \quad (22)$$

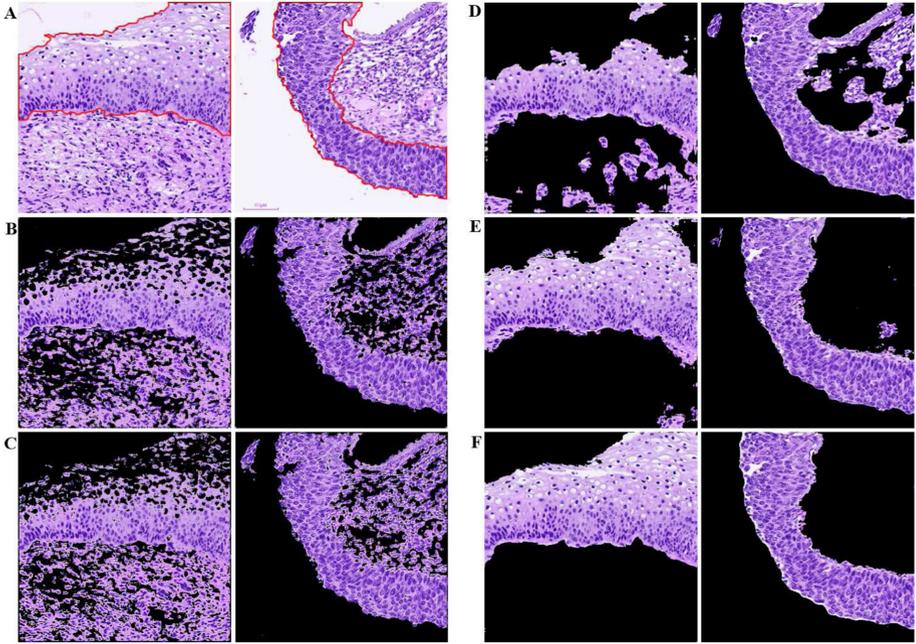
**Automatic Segmentation Results of Tissue Region.** Table 5 shows the *Accuracy* and *Mean\_IoU* of the five models for segmenting the tissues. For such massive and dense data as histopathological images, SVM-RBF and RF are not comparable in terms of time and accuracy with the deep CNN model due to limitations in non-linear mapping capabilities and parameter estimation, achieving only about 86% but taking up to 15 h or more. FCN is able to utilize information from multiple layers simultaneously, but it is not sensitive to details and lacks spatial consistency, achieving only 89% accuracy and the Mean\_IoU of 0.6341. The baseline network U-Net achieves 94% segmentation accuracy and 0.7866 Mean\_IoU, which is more accurate than the above methods. However, U-Net usually needs random initialization and has many parameters. The improved U-net +ResNet50 in this paper achieves 95% accuracy and 0.8252 Mean\_IoU, compared to other deep learning models, it achieved better segmentation results, and each image can be segmented in less than one second, which satisfies the need for fast segmentation. The intuitive segmentation of each model is shown in Fig. 8.

**Table 5.** Results of automatic segmentation of tissue regions in histopathological images.

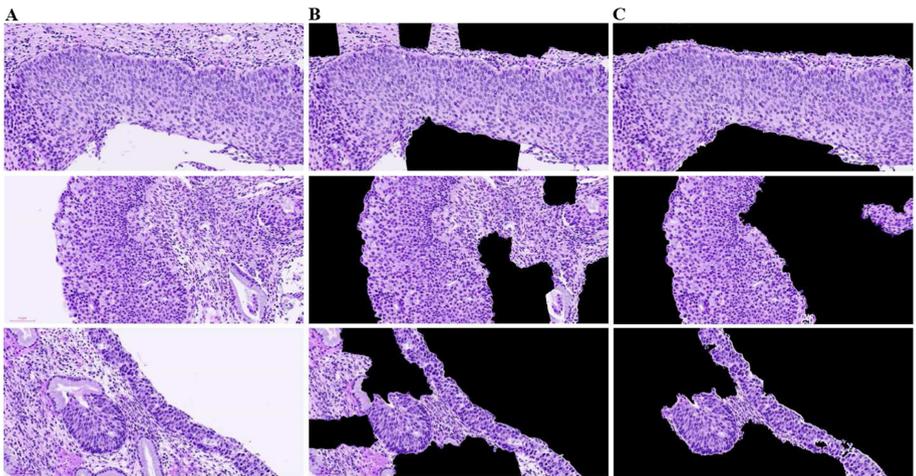
Methods	Accuracy	Mean_IoU	Training time (min)
SVM-RBF	0.8603	0.5269	839
RF	0.8578	0.5537	924
FCN	0.8944	0.6341	56
U-Net	0.9416	0.7866	<b>53</b>
U-Net+ResNet50	<b>0.9542</b>	<b>0.8252</b>	70

The results of pre-segmentation and automatic segmentation of epithelial tissue are visually compared, as shown in Fig. 9. The automatic segmentation model based on deep learning learns the unique cell arrangement patterns and specific pathological features of the epithelial tissue and thus performs well even on cervical pathology images with disordered cell alignment and uneven staining.

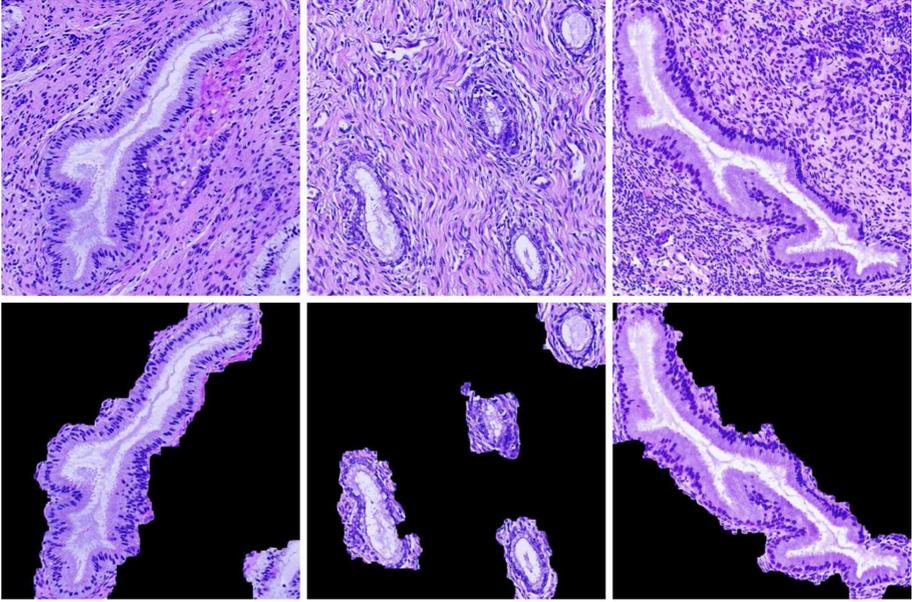
The semi-automatic segmentation model consists of a pre-segmentation model and an automatic segmentation model. To further validate the generality of the model, it is also used to segment glandular cavity tissues in cervical pathology images. 620 histopathological images are taken as datasets, each of which contains one or more glandular cavity tissues. The visual display of the segmentation results is shown in Fig. 10, which achieves medically acceptable results. This further demonstrates that the semi-automatic segmentation model has a strong generalization ability and provides a versatile solution to segment lesion regions in various tumor images.



**Fig. 8.** Automatic segmentation results of tissue regions. Original (A), SVM-RBF (B), RF (C), FCN (D), U-Net (E), U-Net+ResNet50 (F). In the original image, the red outline is the epithelial tissue, the white region is the background, and the rest is fibrous tissue. (Color figure online)



**Fig. 9.** Comparison of pre-segmentation results and automatic segmentation results. Original (A) and pre-segmentation results (B), automatic segmentation results (C).



**Fig. 10.** Segmentation results of glandular cavity.

## 5 Conclusion and Future Work

In this study, a semi-automatic segmentation model for tissue regions is constructed to accurately segment tissues in small datasets. The model starts with the superpixel classification of histopathological images, traditional pixel-level labeling is replaced by superpixel-level labeling. Superpixels preserve the boundary of tissues, and can be labeled quickly. This inexact supervised learning approach greatly reduces physician burden and improves labeling efficiency. The guidance of pathological knowledge makes the results of tissue segmentation more accurate and improves the interpretability of the model. A high-quality training dataset for the deep learning model is constructed based on tissue pre-segmentation results, and it is learned end-to-end to segment tissues more quickly and accurately. The model provides a versatile solution for rapid and accurate segmentation of various tissue regions, and the techniques for constructing deep model datasets greatly reduce the reliance of medical research on public datasets. Individual physicians or small teams can also follow the method to annotate the data in their field, so that they can start research and contribute to the development of the medical field.

Accurate segmentation of tissues like epithelium in histopathological images can provide more precise regions of interest for intelligent diagnosis, thus improving the accuracy of diagnosis. Another member of our team learns pathological features in the segmented epithelial tissue, such as color, texture and cell density characteristics, and then uses deep learning techniques for lesion grading,

demonstrating that focusing on the tissue regions yields more accurate results than diagnosing on the whole image.

Moreover, there are still some issues to be resolved in this paper. First, the rewards and penalties of reinforcement learning can be used for superpixel classification in order to obtain an optimal model. Second, immunohistochemical images such as ki67 and p16 can be combined to provide richer pathological features, thus improving the accuracy of tissue segmentation and disease diagnosis. Finally, our research cannot be limited to the segmentation of tissues, microscopic cell segmentation and morphological analysis, and macroscopic studies of tissue structure and location can further aid computerized diagnosis.

**Funding Information.** This study is supported by Fundamental Research Funds for the Central Universities under Grant 2020CDCGRJ013 and the Science and Technology innovation ability enhancement project of Third Military Medical University under Grant 2019XQY14.

## References

1. Doorbar, J., Griffin, H.: Refining our understanding of cervical neoplasia and its cellular origins. *Papillomavirus Res.* **7**, 176–179 (2019)
2. Zhang, X., Liu, W., Dundar, M., et al.: Towards large-scale histopathological image analysis: hashing-based image retrieval. *IEEE Trans. Med. Imag.* **34**, 496–506 (2015)
3. Xu, J., Luo, X., Wang, G., et al.: A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **191**, 214–223 (2016)
4. Ren, X., Malik, J.: Learning a classification model for segmentation. In: Tomaszewski, J.E., Gurcan, M.N. (eds.) *ICCV 2003, LNCS*, vol. 1, pp. 10–17. IEEE (2003). <https://doi.org/10.1109/ICCV.2003.1238308>
5. Lin, B., Deng, S., et al.: FocAnnot: patch-wise active learning for intensive cell image segmentation. In: Gao, H., Wang, X., Iqbal, M., Yin, Y., Yin, J., Gu, N. (eds.) *Collaborative computing: networking, applications and worksharing. CollaborateCom 2020, LNCS, Social Informatics and Telecommunications Engineering*, vol 350. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-67540-0\\_21](https://doi.org/10.1007/978-3-030-67540-0_21)
6. Linder, N., Konsti, J., Turkki, R., et al.: Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagn. Pathol.* **7**(1), 22 (2012)
7. Hiary, H., Alomari, R.S., Saadah, M., Chaudhary, V.: Automated segmentation of stromal tissue in histology images using a voting Bayesian model. *Signal, Image Video Process.* **7**(6), 1229–1237 (2012). <https://doi.org/10.1007/s11760-012-0393-2>
8. Vu, Q.D., Graham, S., Kurc, T., et al.: Methods for segmentation and classification of digital microscopy tissue images. *Front. Bioeng. Biotechnol.* **7**, 53 (2019)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017)
10. Al-Milaji, Z., Ersoy, I., Hafiane, A., et al.: Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images. *Patt. Recogn. Lett.* **119**(MAR.), 214–221 (2017)

11. De Bel, T., Hermsen, M., Smeets, B., et al.: Automatic segmentation of histopathological slides of renal tissue using deep learning. *Medical Imaging 2018*. In: Proceedings of the SPIE, LNCS, vol. 10581, pp. 1058112 (2018). <https://doi.org/10.1117/12.2293717>
12. Nirschl, J.J., Janowczyk, A., Peyster, E.G., et al.: Deep learning tissue segmentation in cardiac histopathology images. In: *Deep Learning for Medical Image Analysis*, pp. 179–195 (2017)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.* **39**(4), 640–651 (2015)
14. Chen, H., Qi, X., Yu, L., et al.: DCAN: deep contour-aware networks for object instance segmentation from histology images. *Med. Image Anal.* **36**, 135–146 (2017)
15. Lahiani, A., Gildenblat, J., Klamann, I., et al.: Generalising multistain immunohistochemistry tissue segmentation using end-to-end colour deconvolution deep neural networks. *IET Image Process.* **13**(7), 1066–1073 (2019)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
17. Bulten, W., Hulsbergen-van, d.K.C., van d.L.J., et al.: Automated segmentation of epithelial tissue in prostatectomy slides using deep learning. *Medical Imaging 2018*. In: Proceedings of the SPIE, LNCS, vol. 10581 (2018). <https://doi.org/10.1117/12.2292872>
18. Oskal, K.R.J., Risdal, M., Janssen, E.A.M., Undersrud, E.S., Gulrud, T.O.: A U-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN Appl. Sci.* **1**(7), 1–12 (2019). <https://doi.org/10.1007/s42452-019-0694-y>
19. Chan, L., Hosseini, M.S., Rowsell, C., et al.: Histosegnet: semantic segmentation of histological tissue type in whole slide images. In: *Proceedings of the IEEE International Conference on Computer Vision 2019*, LNCS, pp. 10661–10670. IEEE (2019). <https://doi.org/10.1109/ICCV.2019.01076>
20. Schuhmacher, D., Gerwert, K., Mosig, A.: A generic neural network approach to infer segmenting classifiers for disease-associated regions in medical image data. *medRxiv* (2020). <https://doi.org/10.1101/2020.02.27.20028845>
21. Khan, A.M., Rajpoot, N., Treanor, D., et al.: A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans. Biomed. Eng.* **61**(6), 1729–1738 (2014)
22. Achanta, R., Shaji, A., Smith, K., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Patt. Anal. Mach. Intell.* **34**(11), 2274–2282 (2012)
23. Cruz-Roa, A., Díaz, G., Romero, E., et al.: Automatic annotation of histopathological images using a latent topic model based on non-negative matrix factorization. *J. Pathol. Inform.* **2**(2), S4 (2011)
24. Urbán, S., Tanács, A.: Atlas-based global and local RF segmentation of head and neck organs on multimodal MRI images. In: *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis 2017*, pp. 99–103. IEEE. <https://doi.org/10.1109/ISPA.2017.8073577>