2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Classification-Guided Segmentation Algorithm Based on Deep Learning for Epithelium Segmentation in Histopathological Images of Radicular Cysts

Lu Qiu, Meichang Huang, Xiaowei Xu, Wangyuan Zhao, Lu Zhao, Hai Zhong, Yaling Tang#, Jun Zhao#, *Member, IEEE*

***Abstract*— In histopathological analysis of radicular cysts (RCs), lesions in epithelium can provide pathologists with rich information on pathologic degree, which is helpful to determine the type of periapical lesions and make precise treatment planning. Automatic segmentation and localization of epithelium from whole slide images (WSIs) can assist pathologists to complete pathological diagnosis more quickly. However, the class imbalance problem caused by the small proportion of fragmented epithelium in RCs imposes challenge on the typical automatic one-stage segmentation method. In this paper, we proposed a classification-guided segmentation algorithm (CGSA) for accurate segmentation. Our method was a two-stage model, including a classification network for region of interest (ROI) location and a segmentation network guided by classification. The classification stage eliminated most irrelevant areas and alleviated the class imbalance problem faced by the segmentation model. The results of 5-fold cross validation demonstrated that CGSA outperformed the one-stage segmentation method which was lacking in prior epithelium localization information. The epithelium segmentation achieved an overall Dice's coefficient of 0.722, and intersection over union (IoU) of 0.593, which improved by 5.5% and 5.9% respectively compared with the one-stage segmentation method using UNet.**

***Clinical Relevance*— This work presents a framework for automatic epithelium segmentation in histopathological images of RCs. It can be applied to make up for the shortcomings of manual annotation which is labor-intensive, time-consuming and objective.**

I. Introduction

Radicular cysts (RCs) are the commonest inflammatory jaw cysts, representing 42% [1] of all apical periodontitis which are characterized by an inflammatory response and bone destruction in periapical tissues [2]. Although RCs are symptomless and progress slowly in the early stage, they are clinically significant because subsequent infection can cause swelling and pain. Moreover, RCs are often seen as the cause of swelling of the jaws and slowly enlarging swellings [3]. The histopathologic structure of RCs, especially the characteristic of the epithelium, is of particular importance. Lesions in epithelium can provide pathologists with rich information on pathologic degree and help the doctors to distinguish between RCs and other types of lesions (including granuloma and other types of odontogenic cysts), which is an essential prerequisite for making precise treatment planning and estimating the carcinogenic risk. At present, the segmentation of epithelium of RCs is mainly conducted manually by pathologists. However, the large size of whole slide images (WSIs) and the small proportion of fragmented epithelium make manual annotation labor-intensive, error-prone, and time-consuming.

With the advances of techniques in digitalized scanning and computational power, many studies have achieved excellent results in the epithelium segmentation in histopathological images. Quoc et al. [4] proposed a neural network using the decoder built upon a multi-path layout and dense shortcut connections between layers to segment the epithelium in the hematoxylin and eosin (H&E) stained breast tissues images. Wouter et al. [5] employed a deep learning method based on UNet to segment epithelial tissues in H&E stained prostatectomy slides using immunohistochemistry (IHC) as reference standard. Niu et al. [6] used a generalized densely connected encoder-decoder network to compensate for resolution loss induced by pooling layers and improved the performance of the network in segmenting epithelial and stromal regions. Sornapudi et al. [7] built an epithelium segmenting regression model named EpithNet to segment the epithelial regions of the cervical histology images. EpithNet fused the features from a set of sub-networks and output the probability that the pixel at the center of the given neighborhood belonged to the epithelium. Since the tumor epithelial regions are relatively large in size, these studies above apply the one-stage segmentation method without the localization of the region of interest (ROI). When faced with the small object segmentation task, such as epithelium segmentation in RCs, the performance of the one-stage model would be degraded due to the serious class imbalance problem.

In the detection and segmentation challenges of small objects, two-stage methods have the potential to achieve higher accuracy than one-stage methods because it can alleviate the foreground-background class imbalance problem [8]. For example, in the computer vision field of natural images, R-CNN-like detectors [9][10] are used to generate proposals in the proposal stage (e.g., Selective Search, Region

*This research is supported by National Key R&D Program of China (2016YFC0104604, 2016YFC0104608), Shanghai Jiao Tong University Medical Engineering Cross Research Funds (YG2017ZD10), the National Natural Science Foundation of China (No. 81371634), and Clinical Project of West China College of Stomatoloy, Sichuan University (LCYJ2019-8).

Lu Qiu, Xiaowei Xu, Wangyuan Zhao, Lu Zhao, Hai Zhong and Jun Zhao are with School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.

Meichang Huang and Yaling Tang are with State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Oral Pathology, West China Hospital of Stomatology (Sichuan University), Sichuan, 610064, China.

#Corresponding to: Jun Zhao (junzhao@sjtu.edu.cn), Yaling Tang (tangyaling@scu.edu.cn).

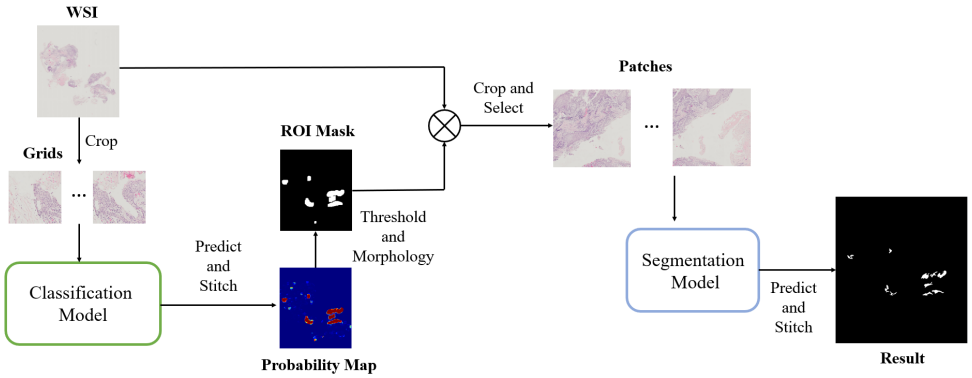


Figure 1. Schematic diagram of CGSA.

Proposal Network) before fine classification and segmentation. However, these methods of extracting proposals of instance are not suitable for histopathological analysis due to the fusion and entanglement of various tissues. Moreover, the large size of WSI makes it impossible to obtain the features of the entire image and train a similar end-to-end network.

In the histopathological analysis of RCs, epithelium ruptures and appears as fragmented tissue. The segmentation network exists serious class imbalance problem due to the small proportion of fragmented epithelium. The overwhelming majority of easily classified negative samples are unable to provide useful information. The loss calculated from easily classified negative samples dominates the final training loss, causing the decline of network performance. In this paper, we presented a two-stage classification-guided segmentation algorithm (CGSA) which was suitable for small object segmentation in histopathological images. CGSA included a classification stage for ROI location and a segmentation stage guided by classification. The classification stage aimed to locate the epithelium and obtain the ROI. The areas outside the ROI which would produce a large number of easily classified negative samples were filtered out. Compared with the one-stage segmentation method, CGSA eliminated most irrelevant areas through the classification network, alleviating the class imbalance problem faced by the segmentation network.

II. Methods

To alleviate the class imbalance problem in small object segmentation, we presented a two-stage CGSA method based on the characteristics of histopathological images. The framework of CGSA was shown in Fig.1. CGSA consisted of two parts, the classification model and the segmentation model. The classification network aimed to obtain the probability map and locate the epithelium. Then we got the ROI of epithelial regions, from where we extracted training patches for the segmentation network. Due to the large size of the histopathological images, both two parts were implemented based on the sliding window method. In the testing stage, we used the segmentation network to predict patches within ROI and stitched the segmentation results of patches to obtain the segmentation results of WSIs.

A. Classification Model

In classification stage, we used neural conditional random field (NCRF) model [11]. The framework has two major components: ResNet34 and fully connected conditional random field (CRF) (Fig.2). Firstly, a sliding window approach was used to extract small grids from WSIs. Each grid can be divided evenly into several patches without overlapping. ResNet34 was employed as the encoder to embed the patches into fixed-length vector representations, which were defined as embeddings. Secondly, CRF took the grid of embeddings as input and modeled the spatial correlation between adjacent patches. Finally, we obtained the probability maps from the prediction results of patches. Compared with the general classification model without CRF, NCRF reduced outliers and inconsistent predictions in the probability maps by considering the spatial correlations between neighboring patches through CRF.

The embeddings obtained from ResNet34 were defined as $x = \{x_i\}_{i=1}^{N}$, and $y = \{y_i\}_{i=1}^{N}$ represented the random variables for each patch i, where N was the number of nodes in CRF. The value of y_i ranged from 0 to 1. The conditional distribution $P(y|x)$ could be modeled as:

$$P(y|x) = \frac{1}{Z(x)}\exp\left(-E(y|x)\right), \tag{1}$$

$$E(y|x) = \sum_i \psi_u\left(y_i|x\right) + \sum_{i<j} \psi_p\left(y_i, y_j|x\right), \tag{2}$$

where $E(y|x)$ represented the energy function, and $Z(x)$ was the partition function. i, j ranged from 1 to N. $\psi_u(y_i|x)$ was the unary potential, and $\psi_p(y_i, y_j|x)$ was the pairwise potential. They measured the cost of single patch and pairs of patches respectively. The unary potential $\psi_u(y_i|x)$ was implemented as the negative logit for y_i before the softmax layer of the ResNet. The spatial correlations were modeled by the pairwise potential models which were implemented based on the cosine similarity between x_i, x_j:

$$\psi_p(y_i, y_j|x) = w_{i,j}\left(1 - \frac{x_i \cdot x_j}{\|x_i\|\|x_j\|}\right). \tag{3}$$

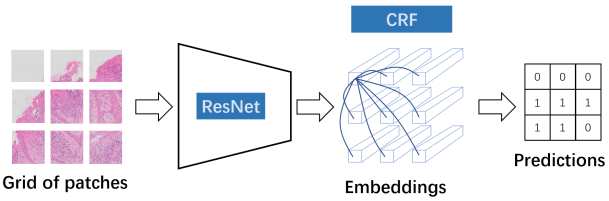
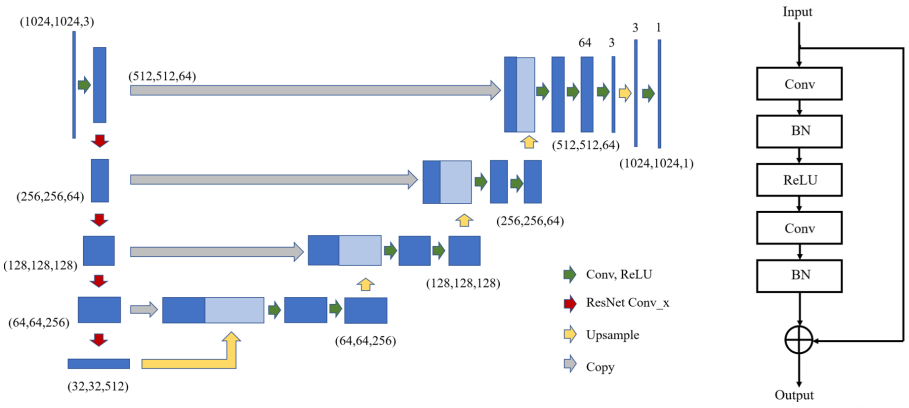


Figure 2. The architecture of classification model.



(a) (b)

Figure 3. The architecture of the segmentation model. (a) Res-UNet, (b) residual block.

Furthermore, an indicator function should be applied to check the label compatibility between y_i and y_j. $w_{i,j}$ was a trainable parameter that controlled the correlation strength between two patches. The prediction y_i was estimated by mean-field approximate inference.

The binary cross entropy loss L_c was utilized to optimize the classification model. y_{true_i} was the label of patch i. The L_c can be formulated as:

$$L_c = \frac{1}{N}\sum_i \left[y_{true_i} \log y_i + \left(1 - y_{true_i}\right) \log(1 - y_i)\right]. \quad (4)$$

After obtaining the probability maps of WSIs from the prediction results of patches, we performed post-processing on the probability maps to obtain the masks of ROI, including Gaussian smoothing with kernel 5×5, threshold processing, open and close operation with kernel 3×3.

B. Segmentation Model

We modified the typical UNet model by changing the encoding part to ResNet blocks, named Res-UNet. The architecture of the segmentation model was shown in Fig.3. The stride of the first convolution layer was 2. Conv_x layer was made up of the residual block. Conv_x represented the typical ResNet convolution modules. Resnet34 was selected as our model, and the four Conv_x layers contained 3,4,6, and 3 residual blocks respectively. The residual block was able to address the degradation problem caused by the increasing depth of network. Moreover, the ResNet blocks could be easily initialized by ImageNet pre-trained weights to improve the network performance. In practice, we found that the mixed loss L_{mix} which consisted of the pointwise cross entropy loss L_B and the soft dice loss L_D achieved a better result than single soft dice loss. The formula of L_{mix} was defined as (5), where W was a constant used to control the proportion of two loss functions.

$$L_{mix} = W \times L_D + (1 - W) \times L_B \quad (5)$$

III. Experiments & Results

A. Data Preparation

We conducted all the experiments based on histopathological images of RCs obtained from West China College of Stomatology, Sichuan University. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board. The dataset included 115 WSIs. Epithelium tissues were annotated by the pathologist. In each experiment of 5-fold cross validation, the datasets were divided into training, validation, testing following a ratio as 3:1:1. The Otsu algorithm [12] was used to extract the masks of WSIs and eliminate redundant areas.

B. Implementation Details

For the classification model, we set patch size as 256×256, $N = 3 \times 3 = 9$. The grid size was set as 768×768. For each patch, we only used the central 128×128 area to determine whether the patch was positive or negative. We used Adam [13] optimizer with weight decay=0.0001. Initial learning rate was set as 0.0001, and for every 100 epochs, learning rate dropped by half. 90 grids were randomly selected in each WSI for each epoch. Since the data set was imbalanced, grids containing only irrelevant tissues and grids containing epithelium were equally selected. The mean-field inference algorithm was performed 10 iterations. In the testing stage, we used the sliding window method with a step of 32. For each 768×768 grid, we only retained the prediction result of the 256×256 patch in center.

For the segmentation model, we extracted patches with 1024×1024 size from ROIs obtained in the classification stage. We used Adam optimizer with weight decay=0.0001. We used warm up strategy to adjust the learning rate, the learning rate was initially set as 0.0000001. The learning rate gradually increased up to 0.0001 after 20 epochs, and then decayed exponentially. W was set as 0.5. In the testing stage, the segmentation network was only used to predict patches within ROI, and results of patches outside ROI were set directly to 0.

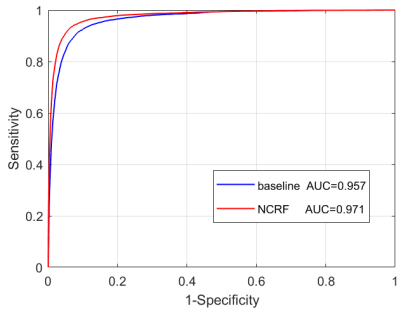


Figure 4. ROC curves of baseline method and NCRF.

During the training of both classification and segmentation, color jitter was added by using torch-vision transforms. We also used random flip and rotation for data augmentation. Our model was implemented with PyTorch-1.7.0 and torchvision-0.8.1. It was trained with a single NVIDIA GeForce GTX 1080 Ti GPU.

C. *Classification Results*

We used ResNet34 model as baseline, and compared the performance of the baseline method with NCRF. Both networks were initialized by ImageNet pre-trained weights. The performance of the NCRF model was significantly improved compared with the baseline method. Evaluations based on probability maps at the WSIs level were shown in Fig.4. For the NCRF method with transfer learning, the accuracy was 0.959, and the AUC was 0.971. The probability maps obtained from the NCRF method tended to eliminate isolated outliers and had smaller probability in non-epithelial tissue areas.

D. *Segmentation Results*

We used UNet and Res-UNet which was proposed in Methods Section for one-stage segmentation as baseline methods. Patches used in baseline methods were obtained by cropping the foreground areas of WSIs through the sliding window method. The average ratio of patches containing only irrelevant tissues to patches containing epithelium was 14.2:1, and in the most extreme case, the ratio was 83.7:1. Although we selected these two types of patches equally during training, there was a serious sample imbalance problem at pixel level compared with the segmentation stage of CGSA. For a definite segmentation network, the proposed framework CGSA outperformed the one-stage segmentation method. The epithelium segmentation task employed two metrics for evaluation: traditional Dice's coefficient and intersection over union (IoU). Dice's coefficient metric was applied to measure the amount of overlap between the segmentation results and the golden standard. The IoU metric was the ratio of the intersection to the union of two sets. We tested our framework through 5-fold cross validation, and the final results showed that our methods increased Dice's coefficient by 4.0% and IoU by 5.9% compared with the one-stage segmentation when using Res-UNet as the segmentation network. Fig.5 showed the segmentation results. Compared with the one-stage segmentation method, results obtained by CGSA had more accurate boundary and tended to eliminate irrelevant tissues.

TABLE 1. AVERAGE SEGMENTATION RESULTS OF WSIS LEVELES EXPRESSED AS MEAN±STD.

Method	Segmentation model	Dice's coefficient	IoU
One-stage	UNet	0.667±0.026	0.534±0.026
One-stage	Res-UNet	0.682±0.037	0.534±0.012
CGSA	UNet	0.709±0.027	0.573±0.028
CGSA	Res-UNet	**0.722±0.039**	**0.593±0.037**

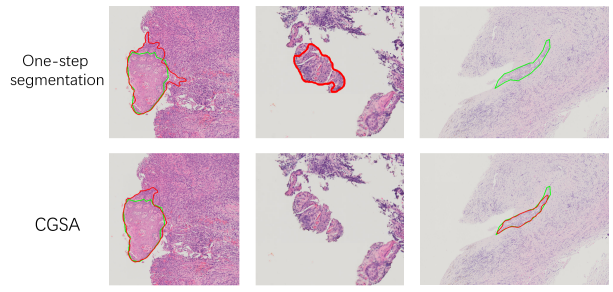


Figure 5. Segmentation results of one-stage segmentation method and CGSA. The areas surrounded by the green line are epithelium tissues annotated by the pathologist. The red line shows the segmentation results of the networks.

IV. CONCLUSION

In this paper, we proposed a two-stage segmentation method: classification-guided segmentation algorithm based on the characteristics of epithelium in RCs. We used the classification network based on ResNet34 and CRF to predict probability map and locate epithelium. The CRF model excluded most of other similar tissues by the spatial information of the surrounding patches. Then the segmentation network was used to fine segment patches extracted from ROI. Our method can eliminate the majority of irrelevant areas through the classification network and alleviate the class imbalance problem faced by segmentation network. Therefore, the segmentation network can focus on finer segmentation and achieve better performance. Quantitative experiments showed that our proposed method achieved better performance than one-stage segmentation method.

REFERENCES

[1] Shteyer, A., and E. Rozovsky. "Periapical lesions--types, incidence and clinical features." Refu'at ha-peh veha-shinayim (Tel Aviv, Israel: 1969) 21 (1972): 100-103.

[2] Braz-Silva, Paulo Henrique, et al. "Inflammatory profile of chronic apical periodontitis: a literature review." Acta Odontologica Scandinavica 77.3 (2019): 173-180.

[3] Shear, Mervyn, and Paul Speight. Cysts of the oral and maxillofacial regions. John Wiley & Sons, 2008.

[4] Vu, Quoc Dang, and Jin Tae Kwak. "A dense multi-path decoder for tissue segmentation in histopathology images." Computer methods and programs in biomedicine 173 (2019): 119-129.

[5] Bulten, Wouter, et al. "Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard." Scientific reports 9.1 (2019): 1-10.

[6] Niu, Chunyang, et al. "A Generalized Densely Connected Encoder-Decoder Network for epithelial and stromal regions segmentation in histopathological images." 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019.

[7] Sornapudi, Sudhir, et al. "EpithNet: Deep regression for epithelium segmentation in cervical histology images." Journal of Pathology Informatics 11 (2020).

[8] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[9] Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.

[10] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[11] Li, Y. , and W. Ping . "Cancer Metastasis Detection With Neural Conditional Random Field." (2018).

[12] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." IEEE transactions on systems, man, and cybernetics 9.1 (1979): 62-66.

[13] Kingma, D. , and J. Ba . "Adam: A Method for Stochastic Optimization." Computer Science (2014).