JOINT SEGMENTATION AND FINE-GRAINED CLASSIFICATION OF NUCLEI IN HISTOPATHOLOGY IMAGES

Hui Qu[†], Gregory Riedlinger[⋆], Pengxiang Wu[†], Qiaoying Huang[†], Jingru Yi[†]
Subhajyoti De[⋆], Dimitris Metaxas[†]

[†]Department of Computer Science, Rutgers University
[⋆] Cancer Institute of New Jersey

ABSTRACT

Nuclei segmentation and classification are two important tasks in the histopathology image analysis, because the morphological features of nuclei and spatial distributions of different types of nuclei are highly related to cancer diagnosis and prognosis. Existing methods handle the two problems independently, which are not able to obtain the features and spatial heterogeneity of different types of nuclei at the same time. In this paper, we propose a novel deep learning based method which solves both tasks in a unified framework. It can segment individual nuclei and classify them into tumor, lymphocyte and stroma nuclei. Perceptual loss is utilized to enhance the segmentation of details. We also take advantages of transfer learning to promote the training of deep neural networks on a relatively small lung cancer dataset. Experimental results prove the effectiveness of the proposed method. The code is publicly available[1].

Index Terms— Nuclei segmentation, nuclei classification, deep learning, histopathology image analysis

1. INTRODUCTION

Histopathological assessment remains a cornerstone of clinical diagnosis and classification of cancer. The underlying tissue architectures in histopathological images provide wealth information about the nature of disease, cytogenetic abnormalities, and characteristics of the microenvironment [1]. For example, malignant tumor cells can be distinguished from benign cells by the features of their nuclei [1]. Besides, phenotypic variations among tumor cells, which are indicative of intra-tumor heterogeneity have consequences for treatment strategies for cancer patients [2]. Therefore, development of software for refined segmentation and classification of histopathological structures such as lymphocytes and cancer nuclei can help improve the clinical management of cancer.

Nuclei segmentation and classification are both challenging tasks. The nuclear size is much smaller compared to glands or organs, and nuclei are often close to each other,

making it hard to segment individual nuclei accurately. The fine-grained classification of nuclei is also difficult due to the large inter-class and intra-class variances in nuclear shapes and textures. Traditional methods like thresholding, watershed, clustering, region growing [3] are not able to cope with these challenges well. Early learning-based methods [4, 5, 6] learn to segment or classify nuclei using low-level handcrafted features such as color, texture, gradients geometric features, which have limited representation capability.

Recently, deep convolutional neural networks (CNN) have achieved great success for image classification and segmentation [7, 8, 9, 10]. Many deep learning based methods have been proposed for nuclei segmentation or classification. Xing et al. [11] utilized CNN to obtain an initial shape of nuclei and then separate individual nuclei using a deformable model. In [12, 13], nuclei segmentation is performed by classifying each pixel into different classes using a patch around the pixel as the input to an image classification network. The computational cost is large because each patch only predicts one pixel. Fully convolutional neural networks (FCNs) [9, 10], which directly output the same size segmentation map for the input image, are more efficient and effective for image segmentation tasks, and have been used in nuclei segmentation [14]. Compared to nuclei segmentation, there are fewer works about nuclei fine-grained classification using deep learning. Sirinukunwattana et al. [15] built two CNNs to detect nuclei and then classify them into sub-categories. Zhou et al. [16] proposed a sibling CNN with objectiveness prior to detect and classify nuclei simultaneously.

Although the current methods have achieved good accuracy, they are not able to produce the pixel-wise masks of different types of nuclei at the same time, thus cannot generate both nucleus features and spatial distributions, which are essential for histopathology image analysis. Actually, the network structures for two tasks are similar, both need to extract feature representations from the input image. In this paper we propose a framework to solve the two tasks jointly. Different from previous methods, our model outputs the segmentation map for every type of nuclei and the background, which can segment individual nuclei, as well as classify them into tu-
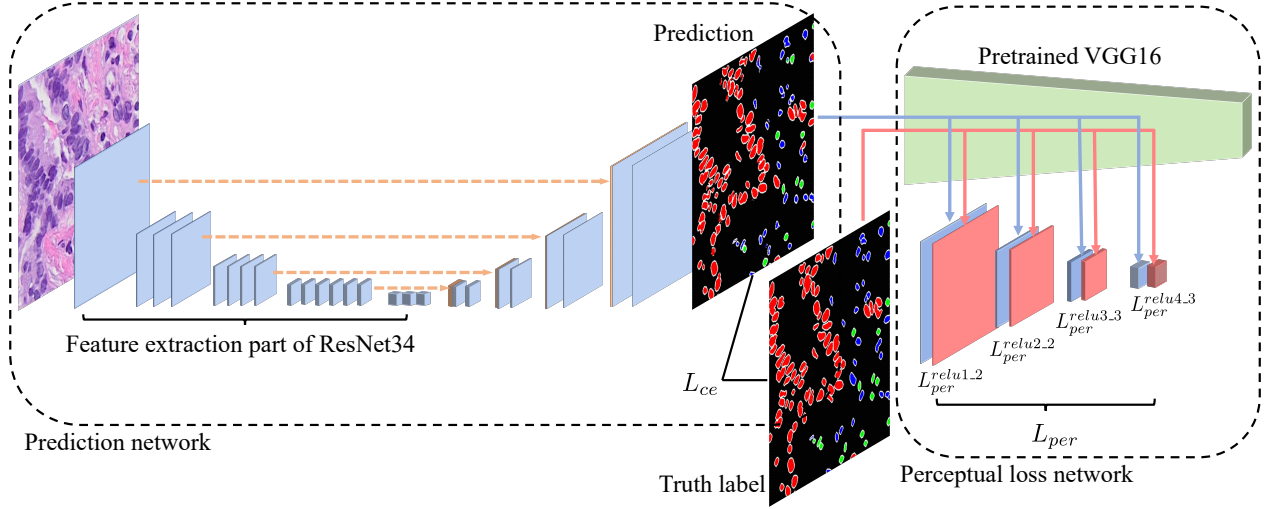
---

[1] https://github.com/huiqu18/NucleiSegClas

Fig. 1. System overview. Our framework consists of the prediction network and perceptual loss network. The prediction network takes the feature extraction part of ResNet34 as the encoder and outputs the segmentation map of different types of nuclei. The loss network uses the fixed pretrained VGG16 model as a feature extractor and computes the perceptual loss.

mor, lymphocytes and stroma sub-categories. Besides, one more channel is used to predict the contours of nuclei, aiming at separating touching nuclei. To further improve the segmentation accuracy, we take advantage of the perceptual loss [17] that can measure small differences in two images. In addition, transfer learning is utilized to promote the training due to the small size of the annotated dataset.

2. METHOD

2.1. Dataset and preprocessing

In order to evaluate the performance of joint nuclei segmentation and classification algorithms, we annotated a dataset that consists of 40 H&E stained tissue images from 8 different lung adenocarcinoma or lung squamous cell carcinoma cases, and each case has 5 images of size about 900×900. There are around 24000 annotated nuclei in the dataset and each nucleus is marked as one of the following three types: tumor nucleus, lymphocytes nucleus, stroma (fibroblasts, macrophages, neutrophils, endothelial cells, etc.) nucleus. For each image, we use one label image to encode the segmentation mask and classification class information of each nucleus. In a ground truth label, pixels of value 0 are background. Pixels that have a same positive integer belong to an individual nucleus. The integer value id also indicates the class of the nucleus: (1) tumor nucleus if $mod(id, 3) = 0$, (2) lymphocyte nucleus if $mod(id, 3) = 1$, (3) stroma nucleus if $mod(id, 3) = 2$, where mod is the modular operation. It is easy to extract both the class information and the segmentation mask of every nucleus from this encoding, even if the classes are imbalanced. Fig. 2 shows an example of original image and its labels.

The dataset is split into three parts: 24 images for training, 8 for validation and the remaining 8 for test. Each set contains at least one image of each case. Color normalization [18] is performed on all images to remove the color variations. Data augmentation is crucial for training deep neural networks when only a few training images are available, which is exactly the case in our task. We extract 16 image patches of size 250×250 uniformly from each training image with overlap, resulting in 384 small image patches. For each patch, a 224×224 image is randomly cropped as the network input. Other augmentations include random scale, random horizontal and vertical flip, random affine transformation, random elastic transformation, random rotation, normalization with mean and standard deviation by channel.

2.2. Network structure

The proposed framework is shown in Fig. 1. It consists of two parts: the prediction network that generates the segmentation mask of each type of nuclei, and the perceptual loss network that computes the perceptual loss between the predicted label and ground-truth label.

The prediction network is the routine encoder-decoder structure based on U-net [10]. We utilize the powerful representation ability of residual networks [8] to extract features. The encoder is from ResNet34 [8], without the average pooling and fully connected layers, and is initialized with the pretrained parameters from image classification tasks. There are skip connections between encoder and decoder, which helps to recover high resolution feature maps. The network outputs five probability maps: background, inner part of tumor nuclei, inner part of lymphocytes nuclei, inner part of stroma nuclei and contours of all nuclei. The contour map
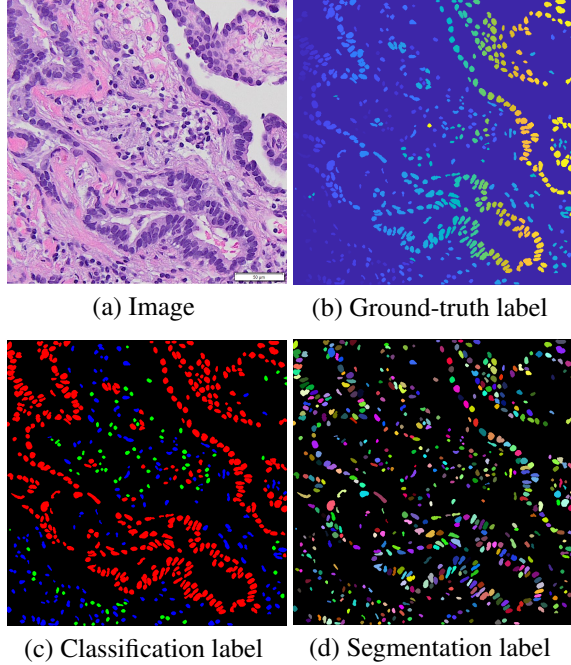
(a) Image   (b) Ground-truth label

(c) Classification label   (d) Segmentation label

Fig. 2. Example of an image and its labels. (a) Original image, (b) Ground-truth label, (c) Classification label, red, green and blue colors represent tumor, lymphocytes and stroma nuclei, respectively. (d) Segmentation label, distinct colors are different nuclei.

mainly aims to capture the contours of crowded and touching nuclei. As a result, the predicted inner parts of each nucleus are not connected. The final nuclei mask is generated by a simple morphological dilation operation. In this way, we can obtain each individual nucleus without much extra effort.

The perceptual loss network is utilized to improve the segmentation accuracy of details in the image. It originates from Johnson et al.'s work [17], in which the authors compute loss between high-level features of the transformed image and the original image. The pretrained VGG16 model is a feature extractor and is fixed during training and test. Four levels of features are extracted using this network for the output of the prediction network and the ground-truth label, i.e., feature maps after the last ReLU layer of the first, second, third and fourth blocks of VGG16 model, denoted as $relu1\_2, relu2\_2, relu3\_3, relu4\_3$. The mean square loss is then computed between the feature sets of two inputs.

2.3. Loss function

The loss function of the method consists of two parts. The first part is the cross entropy loss for five classes, i.e. background, inside tumor, inside lymphocyte, inside stroma and contour. It is defined as

$$L_{ce}(y, t, w) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{m=1}^{5} w_i t_i^{(m)} \log y_i^{(m)} \quad (1)$$

where N is the number of all pixels, $y_i^{(m)}$ is the probability of pixel i belonging to class m, $t_i^{(m)} \in \{0, 1\}$ is the corresponding ground-truth label of class m, w_i is the optional weight for pixel i and the default value is 1 for all pixels. In order to alleviate the problem of class imbalance, we assign larger weights for low frequent class pixels. The weight is calculated using a similar method in [10]:

$$w(x_i) = 1 + w_0 \cdot \exp\left(-\frac{(d_1(x_i) + d_2(x_i))^2}{2\sigma^2}\right) \quad (2)$$

where d_1, d_2 are the distances to the nearest and the second nearest nuclei. We set $\sigma = 5$ pixels and $w_0 = 10$.

The second part is the perceptual loss. Let's denote the trained VGG16 model as a function f. The features after the ReLU layer of the k-th block can be written as $f_k(x)$, where x is the input of VGG16. The size of k-th level features is denoted as $C_k \times H_k \times W_k$. The perceptual loss is

$$L_{per}(\hat{y}, t) = \frac{1}{4} \sum_{k=1}^{4} L_{per}^k(\hat{y}, t)$$
$$L_{per}^k(\hat{y}, t) = \frac{1}{C_k M_k N_k} \|f_k(\hat{y}) - f_k(t)\|_2^2 \quad (3)$$

where $\hat{y} = \arg\max y$ is the prediction map obtained from the output probability map y. The final loss function is

$$L = L_{ce} + \beta L_{per} \quad (4)$$

where β is a parameter that adjusts the weight of the perceptual loss, and is set to 0.1 in the experiments.

3. EXPERIMENTS

We test the proposed method on the lung cancer dataset mentioned in Section 2.1, and compare it to two popular approaches for segmentation. One is the fully convolutional network proposed by Long et al.[9], which is the first FCN used for segmentation task. The other is U-net [10], which has been widely used in medical image segmentation. Both networks output five probability maps as ours. For all models, we trained 300 epochs for convergence with Adam optimizer. The learning rate, batch size and weight decay are 0.0001, 8, and 0.0001, respectively.

3.1. Evaluation metrics

For nuclei segmentation, we use F1-score to measure the detection accuracy. A segmented nucleus is considered as true positive if it overlaps with a ground-truth nucleus more than 50% of the area of the true nucleus. Otherwise it is a false positive. All ground-truth nuclei that have no corresponding segmented nuclei are treated as false negative. Object-level dice coefficient and hausdorff distance are used to measure the segmentation accuracy. Dice coefficient measures how well the ground-truth and predicted nuclei overlap with each
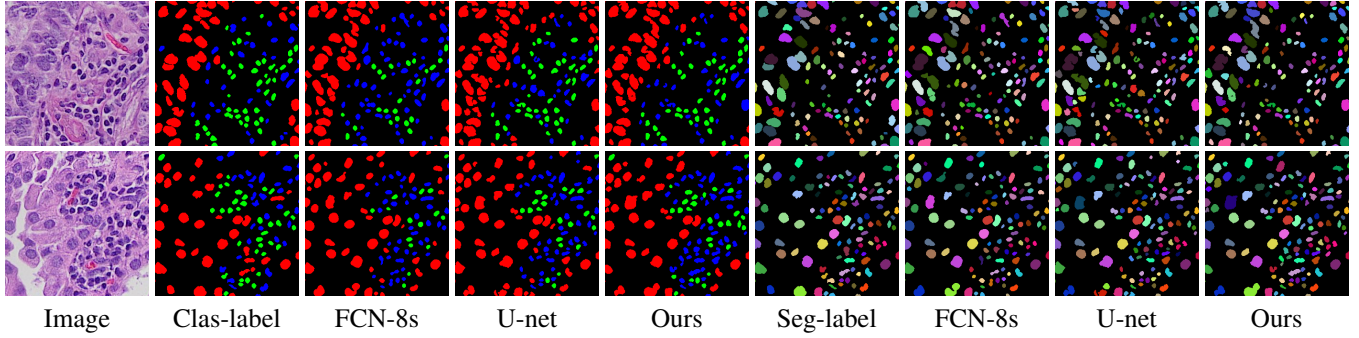
| Image | Clas-label | FCN-8s | U-net | Ours | Seg-label | FCN-8s | U-net | Ours |

Fig. 3. Some images results of ground-truth labels, FCN8s, U-net and the proposed method.

Table 1. Nuclei segmentation results on the test set.

Method	All			Tumor			Lymphocyte			Stroma		
	F1	Dice	Haus	F1	Dice	Haus	F1	Dice	Haus	F1	Dice	Haus
FCN-8s [9]	0.863	0.842	5.17	0.778	0.807	8.52	0.527	0.565	41.52	0.562	0.528	17.88
U-net [10]	0.874	0.865	4.68	0.802	0.831	7.58	0.620	0.622	36.99	0.593	0.566	15.93
Ours w.o. L_{per}	0.874	0.870	4.40	0.806	0.839	6.93	0.620	0.637	28.13	0.619	**0.599**	**14.36**
Ours no-pretrain	0.865	0.863	4.67	0.797	0.831	7.43	0.635	0.626	31.75	0.585	0.566	15.88
Ours	**0.886**	**0.876**	**4.14**	**0.826**	**0.846**	**6.66**	**0.671**	**0.668**	**27.26**	**0.622**	0.589	14.99

Table 2. Nuclei fine-grained classification accuracies (%) on the test set.

Method	All	Tumor	Lym	Stroma
FCN-8s [9]	80.96	85.14	**81.43**	72.64
U-net [10]	83.00	88.72	75.85	76.20
Ours w.o. L_{per}	83.86	89.56	80.00	75.90
Ours no-pretrain	83.19	**90.66**	72.73	76.32
Ours	**84.75**	90.29	75.44	**79.94**

other and hausdorff distance is for shape similarity. Formal definitions can be found in [19].

For fine-grained classification, we only consider the accuracy in true positives instead of all ground-truth nuclei, because not all nuclei have corresponding predicted ones.

3.2. Results and comparison

The quantitative results of different methods are shown in Table 1 and 2. It can be observed that all three models have achieved relatively good segmentation and fine-grained classification results, showing that our idea of combing the two tasks are feasible. Compared to FCN-8s and U-net, our method has improvements on the segmentation of all types of nuclei, especially on lymphocytes. The improvements on sub-category nuclei are larger than those on all nuclei because the classification results also affect the sub-class metrics, i.e wrongly classified nuclei have no corresponding ground-truth ones, thus reducing the F1 and dice scores and increasing hausdorff distance. Combined with the classification accu-

racies in Table 2, our method also outperforms FCN-8s and U-net on the fine-grained classification.

To illustrate the effects of transfer learning and perceptual loss, we also report the results for our model without perceptual loss or pretrained weights of the encoder part. It's evident that both techniques can promote the performance of segmentation and classification. The results without pretrained are worse than that without the perceptual loss, because the pretrained weights are more important than perceptual loss for this dataset due to its small size.

For the sub-category results of both tasks, the performance on tumor nuclei is the best because it is easier to be distinguished. The shape and size of some lymphocytes and stroma nuclei are very similar, resulting in relatively lower segmentation and classification accuracies. Some image results are shown in Fig. 3.

4. CONCLUSION

In this paper we proposed a framework that jointly segments and classifies different types of nuclei from histopathology images. We combine the cross entropy loss and perceptual loss to enhance the segmentation of details in the image. We also make use of transfer learning to better train the model on a small dataset. Experiments show that our method is able to achieve good segmentation and fine-grained classification results simultaneously. The segmentation maps of different types of nuclei generated by our method can be used to analyze the nuclear features and their spatial distributions, which is crucial for cancer diagnosis and prognosis.

5. REFERENCES

[1] Metin N Gurcan, Laura Boucheron, Ali Can, Anant Madabhushi, Nasir Rajpoot, and Bulent Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, pp. 147, 2009.

[2] Xiao-xiao Sun and Qiang Yu, "Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment," *Acta Pharmacologica Sinica*, vol. 36, no. 10, pp. 1219, 2015.

[3] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential," *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2014.

[4] Meng Wang, Xiaobo Zhou, Fuhai Li, Jeremy Huckins, Randall W King, and Stephen TC Wong, "Novel cell segmentation and online svm for cell cycle phase identification in automated microscopy," *Bioinformatics*, vol. 24, no. 1, pp. 94–101, 2007.

[5] Hui Kong, Metin Gurcan, and Kamel Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting," *IEEE transactions on medical imaging*, vol. 30, no. 9, pp. 1661–1677, 2011.

[6] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, 2010.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[11] Fuyong Xing, Yuanpu Xie, and Lin Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[12] Andrew Janowczyk and Anant Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, 2016.

[13] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[14] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 933–936.

[15] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[16] Yanning Zhou, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng, "Sfcn-opi: Detection and fine-grained classification of nuclei using sibling fcn with objectness prior interaction," *arXiv preprint arXiv:1712.08297*, 2017.

[17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[18] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

[19] Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al., "Gland segmentation in colon histology images: The glas challenge contest," *Medical image analysis*, vol. 35, pp. 489–502, 2017.