EPITHELIUM-STROMA CLASSIFICATION IN HISTOPATHOLOGICAL IMAGES VIA CONVOLUTIONAL NEURAL NETWORKS AND SELF-TAUGHT LEARNING

Yue Huang^{1,2}, Han Zheng¹, Chi Liu², Gustavo Rohde^{3,4}, Delu Zeng¹, Jiaqi Wang⁵, Xinghao Ding^{1*}

1. School of Information Science and Engineering, Xiamen University, Xiamen, 361005, China

2. Dept. of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, 15213, U.S.A.

3. Dept. of Biomedical Engineering, University of Virginia, Virginia, 22908, U.S.A.

4. Dept. of Electrical and Computer Engineering, University of Virginia, Virginia, 22908, U.S.A.

5. Beijing University of Post and Telecommunications, Beijing, 100876, China

*dxh@xmu.edu.cn

ABSTRACT

Epithelium-stroma classification is always considered as an important preprocessing step for morphological quantitative analysis in image-based histological researches of oncologic diseases. However, large-scale accurate ground-truth labeling is expensive in histopathological image analysis, thus the classification performances will still be limited with the insufficient labeled training samples. Considering that acquisition of public unlabeled histopathological images is much cheaper, an epithelium-stroma classification framework is developed, based on the deep convolutional neural network framework and the strategies of self-taught learning. The method has the ability of taking advantage of large-scale unlabeled public histopathological data as auxiliary data, and then transferring the knowledge to enhance the performances in epithelium-stroma classification with limited labeled training data. The experiments demonstrate that the proposed method outperforms traditional CNNs when the labeled training data size is decreasing dramatically.

Index Terms— epithelium-stroma classification, convolutional neural networks, self-taught learning, histopathological image analysis, transfer learning

1. INTRODUCTION

Epithelium and stroma are two basic types of animal tissues, as shown in Fig.1. Epithelium-stroma segmentation is recommended as one of the most widely-used preprocessing operations in histopathological image analysis. For example, epithelium-stroma ratio, is always recognized as an independent prognostic indicator in many in oncology researches [1-2]. It also has been reported in recent studies that morphological patterns in stroma are strongly associated with the prognostic information in breast cancer [3]. According to the recent literature, most of the contributions treat epitheliumstroma segmentation as a pattern recognition problem. That is, histopathological images are partitioned into many image patches at first, and then these image patches are binarily classified into epithelium or stroma with some supervised learning methods [4-5]. In the work described here, we only focus on the step of image patches classification, since higher classification accuracy denotes better pixel-wise epitheliumstroma segmentation performances. Feature extraction is a crucial step in supervised recognition frameworks, and it can be briefly classified into two categories for epithelium-stroma recognition: hand-crafted features and data-driven features. Texture, color, and their combinations, are most widely-used hand-crafted features in epithelium-stroma classification [6]. Deep learning strategies have received tons of attentions in computer vision recently, especially in pattern classification problems [7]. These end-to-end data-driven approaches in deep neural networks attempt to extract hierarchical features with strong representive power from large amount of training data automatically via the multi-layer architecture [8]. Epithelium-stroma classification framework with convolutional neural networks (CNNs) have been reported, and the results demonstrated that this CNNs based method outperform hand-crafted features, e.g. local binary pattern [5, 9]. Large-scale labeled training data is prerequisite in the CNNs based classification. However, in the real-world application, it will be huge cost for the pathologists to acquire large annotated labeled data in each specified dataset, since the large-scale ground-truth labeling is tedious and time-consuming.

The work described here are motivated based on the two observations. Firstly, more and more histopathological images datasets are public online as challenges or for algorithm validations, and tons of glass slides are generated and scanned into digital images in the hospitals every day in the real-world

This work was supported in part by the National Natural Science Foundation of China under Grants 61571382, 81671766, 81301278, 81671766, 61172179 and 61103121, in part by the Guangdong Natural Science Foundation under Grant 2015A030313007, in part by the Fundamental Research Funds for the Central Universities under Grants 20720160075, 20720150169 and 20720150093, in part of CCF-Tencent Open Fund.

application [10-12]. Large-scale histopathological image data is easy to acquire, however, most of them are unlabeled, since accurate ground-truth from pathologists is very expensive. Secondly, although the public histopathological images have significantly different appearances from various datasets due to different generation procedures, they still share limited number of basic structural patterns, e.g. nuclear, cytoplasm, gland, and mitosis.

Self-taught learning, a category in transfer learning, is proposed to predict the labels of testing samples in the target domain with the limited training samples in the target domain and the large-scale unlabeled auxiliary data in the source domain simultaneously [13-15]. The method assumes that the large-scale unlabeled data in source domain can be utilized to learn the basic structures that have the ability of representing data in the target domain. Considered that unlabeled histopathological images are much cheaper to acquire than labeled data, and the structure patterns are somehow similar across different histopathological images, the proposed work aims to exploit and take advantage of the information shared across different histopathological datasets without labels information. Thus the epithelium-stroma classification performances with limited labeled training data in the target domain will be enhanced, and then the burdens of training data labeling will be reduced. It should be mentioned that self-taught learning is different with semi-supervised learning, as the data in source and target domains may belong to different classes [15], for example, in the proposed work, target domain contains epithelium and stroma data, but source domain contains various kinds of histopathological image data, which are not limited to epithelium or stroma.

Therefore, an epithelium-stroma classification method based on CNNs and self-taught learning is proposed. The contributions of the proposed works can be summarized that the method introduces the strategies of self-taught learning to CNNs, in order to take advantages of large-scale public unlabeled histopathological data as the auxiliary data. Therefore, the epithelium-stroma classification performances can be enhanced when the labeled training samples are limited, and then the burden of ground-true labeling is reduced in the real-world applications. In addition, there is no requirement of fine-tune in CNNs anymore.

2. METHODS

In the section, the basic idea of self-taught learning will be reviewed briefly at first, and then proposed method will be described in details. As defined in transfer learning, source domain is denoted as S, and target domain is denoted as T[13]. In self-taught learning, data in source domain is denoted as $\{\mathbf{x}_i^S\}_{i=1}^m$, m is the sample size. Data in target domain is denoted as $\{\mathbf{x}_i^T\}_{i=1}^n$, where n is the sample size in target data [14]. The data in target domain are divided into labeled training samples and unlabeled testing samples. The method



Fig. 1. Examples of image patches from epithelium (top row) and stroma (bottom row) from histopathology images.

is designed to predict the labels of testing sample, with the limited labeled training samples in target domain and the unlabeled auxiliary data from source domain. Self-taught learning consists of two steps of sparse coding [15].

Step 1 A dictionary $\mathbf{D} = [\mathbf{d}_1, ... \mathbf{d}_r]$ is learned from the auxiliary unlabeled data in source data, with the object function with minimizing reconstruction error:

$$J_{s}(\mathbf{D}, \alpha_{i}^{S}) = \sum_{i=1}^{m} (||\mathbf{x}_{i}^{S} - \mathbf{D}\alpha_{i}^{S}||_{2}^{2} + \lambda ||\alpha_{i}^{S}||_{1})$$

s.t. $||\mathbf{d}_{j}||_{2} \le 1, \forall 1 \le j \le r$ (1)

where r is the atom number of the dictionary \mathbf{D} , $\lambda > 0$ is the penalty parameter, α_i^S is the representation coefficient.

Step 2 The data in target domain are sparse represented by the dictionary \mathbf{D} learnt in step 1, and then the representation coefficients are assigned to the data in target domain as the extracted features [15].

$$J_t(\alpha_i^T) = ||\mathbf{x}_i^T - \mathbf{D}\alpha_i^T||_2^2 + \lambda ||\alpha_i^T||_1, \forall 1 \le i \le n$$
 (2)

where α_i^T is the representation of \mathbf{x}_i^T with respect to \mathbf{D} . Therefore, the labels of the testing samples in target domain can be predicted by the extracted features α_i^T and any supervised classifier, e.g. logistic regression. It can be observed that the feature extraction procedure in step 2 can be treated as a procedure of single-layer representation. Thus, the major contribution of self-taught learning can be summarized that the basic structure patterns can be learnt from the large-scale unlabeled auxiliary data in source dataset at step 1, and then they are considered as the knowledge to be transferred to the target domain to enhance the reconstruction-based classification performances in step 2.

As mentioned in the introduction, tons of literatures have demonstrated that CNNs has the ability of extracting highlevel representative features due to its multi-layer architecture [16-17]. In CNNs, convolutional kernels are the key parameters, which are randomly initialized at first, and then are estimated from the large-scale labeled data via back-propagation. The knowledge to be transferred to target domains are actually some patterns which are assumed to be shared with the target domain and the source domain. Therefore, we proposed a CNNs framework with the strategies of self-taught learning, where the atoms in dictionary from the representation procedure in eq.(1) are assigned as the convolutional kernels in CNNs to replace random initialization. The framework is designed to take advantage of the knowledges learnt from the unlabeled data, which is much cheaper to acquire than the labeled training data in target domain.

3. EXPERIMENTS

3.1. Datasets descriptions

There are two kinds of datasets utilized in the proposed work, breast cancer datasets (D1), and a dataset of combination of public H&E stained histopathological images (D2). Two breast cancer datasets in D1, which are acquired from two separate and independent cohorts, Netherland Cancer Institute (NKI, 248 patients, 778 images) and Vancouver General Hospital (VGH, 328 patients, 664 images) [3, 10]. Both datasets consist hematoxylin H&E eosin(H&E) stained histological images from breast cancer tissue microarrays (TMAs). Images from both datasets are scanned at a 20X optical magnification, and each of them is with size of 1128×720 . The epithelium and stroma areas are manually labeled by pathologists as the ground-truth in all the images of both NKI and VGH. Image patches with the size of 50×50 in epithelium and stroma areas are randomly selected, and then training and testing data can be generated, as shown in Table 1.

Second dataset D2 consist of a combination of public H&E histopathological images from different resources. 45 images are downloaded from the public datasets [11-12], and 30 images are coming from the searching results from google images with the retrieval words of 'H&E histopathological images'. Since H&E is the most widely-used staining method in histology research, it is very easy to find large-scale H&E stained histopathological images in the internet. Therefore, the H&E histopathological images in the dataset are generated from different cancers, with different magnifications, and are acquired with different scanners. All the images in the size of 50×50 in unlabeled D2 are randomly selected as the auxiliary data.

3.2. Implementation details

To emphasize the contributions of introducing self-taught learning strategy, in the proposed work, the architecture of CNNs without auxiliary data is implemented with LeNet, which includes two convolutional layers and two maximum

Table 1. The number of training and testing image patches in

 D1 for algorithm validation.

e			
Data	Tissue	Training data size	Testing data size
NKI	Epithelium	5,500	20,000
	Stroma	5,500	20,000
VGH	Epithelium	3,000	25,000
	Stroma	3,000	25,000



Fig. 2. Example images in dataset D2, *left*: an image from public dataset [12]; *right*: an image from google image

pooling layers [16]. The results will be treated as the baseline for following comparisons. The size of convolutional kernels is set to be 5×5 , and the number of kernels each convolutional layer is eight. In the implementation of the proposed method, since the dictionary learning in the first step in self-taught learning is based on image reconstruction, we use principle component analysis (PCA) as a simplified version. The eigenvectors corresponding to the first eight components are chosen as the dictionary, which are then utilized as the convolutional kernels in the following CNNs. That is, instead of random initialization and fine-tune with back-propagation in traditional CNNs, the convolutional kernels are estimated with eq.(1) directly from D2 without any fine-tune. The implementation of CNNs(LeNet) with pre-defined kernels are the same as in [18]. Similar as in the baseline, there are two convolutional layers and two maximum pooling layers in the architecture of proposed method, and the features of training and testing samples can be extracted after the fully connection layer followed by the last pooling layer. Finally, support vector machine (SVM) is applied as the classifier in the binary recognition. The size of convolutional kernel and the number of feature maps in each layer of proposed method are set to the same as in CNNs, for a fair comparison.

Algorithms validation on epithelium-stroma classification is implemented in the two datasets in D1 individually. As shown in Table 2, six experiments belonging to three categories are designed. The first category (first and second rows) is called baseline, where the classification is trained and tested with the decreasing labeled training data in the target domain without auxiliary data. Classifications in the second category employ the alternative dataset in D1 as the auxiliary data, as the third and the fourth rows in Table 2. It should be mentioned that the labels in the source domain are not used,

Source Domain		
(auxiliary data)	Size of auxiliary data	Target Domain
		NKI
		VGH
VGH	12,000	NKI
NKI	12,000	VGH
D2	12,000	NKI
D2	12,000	VGH

Table 2. Experiments for algorithm validation on epitheliumstroma classification with different settings of source data and target data.

although the ground-truth is available in D1. In addition, image patches in NKI or VGH as auxiliary data are randomly selected from the images with size of 50×50 , so they may be different from the training and testing data in Table 1. In the third category, unlabeled dataset D2 is applied as the auxiliary data. In all the experiments, the labeled training sample size is decreasing from 20% to 4% of the original labeled training data size in Table 1. The testing sample size is the same as in Table 1 for a fair comparison.



Fig. 3. Average classification accuracy comparisons of epithelium-stroma classification in NKI via traditional CNNs and the proposed method. The results of proposed method with the auxiliary data from both VGH and D2 are provided.

3.3. Experiments results

The results of experiments from three categories are provided and plotted in Fig.3 and Fig.4. It can be observed that proposed method outperforms traditional CNNs with the auxiliary data either from D1 or D2 in both VGH and NKI, when the size of labeled training data is decreasing dramatically. It also can be observed the comparisons between two kinds of auxiliary data with the proposed method. The classifica-



Fig. 4. Average classification accuracy comparisons of epithelium-stroma classification in VGH via traditional CNNs and the proposed method. The results of proposed method with the auxiliary data from both NKI and D2 are provided.

tion performances is better when the images from D1 are employed as the auxiliary data, and this can be explained that the two datasets in D1 are coming from the same cancer, and are scanned at the same scanning magnification, thus they share more similar basic structure patterns with each other.

4. CONCLUSIONS

Epithelium-stroma classification is one of the most common preprocessing steps in image analysis for quantitative histology research. Labeled data in histopathological images are always insufficient to estimate the large amount of parameters in CNNs, but it is much cheaper to retrieval large number of unlabeled histopathological images from the internet. Therefore, the paper proposes a epithelium-stroma classification framework with CNNs and the strategies of self-taught learning. Basic structure patterns are learnt as knowledge from large scale unlabeled auxiliary data during self-taught learning, and then are transferred to estimate the convolutional kernels in CNNs. Based on the experimental results, the transfer knowledges have the ability of enhancing the performances of proposed image patch based epithelium-stroma recognition, if the labeled training samples are strongly insufficient. The proposed method can be also considered as a tool for other pattern recognition problems in histopathological image analysis, since the high-quality ground-truth acquisition from pathologist is always very expensive in real-world application.

5. REFERENCES

[1] J. Liu, J. Liu, J. Li, et al., "Tumor-stroma ratio is an independent predictor for survival in early cervical carcinoma," Gynecologic Oncology, vol. 132, pp. 81-86, 2014.

- [2] k. Wang, W. Ma, J. Wang, et al., "Tumor-Stroma Ratio Is an Independent Predictor for Survival in Esophageal Squamous Cell Carcinoma," *Journal of Thoracic Oncol*ogy, vol. 7, pp. 1457-1461, 2012.
- [3] A. H. Beck, A. R. Sangoi, S. Leung, et al., "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci Transl Med*, vol. 3, p. 108ra113, Nov 9 2011.
- [4] F. Bunyak, A. Hafiane, Z. Al-Milaji, et al., "A segmentation-based multi-scale framework for the classification of epithelial and stromal tissues in H&E images," *in Bioinformatics and Biomedicine (BIBM)*, 2015 IEEE International Conference on, 2015, pp. 450-453.
- [5] A. Sethi, L. Sha, A. R. Vahadane, et al., "Empirical comparison of color normalization methods for epithelialstromal classification in H and E images," *J Pathol Inform*, vol. 7, p. 17, 2016.
- [6] F. Bianconi, A. Ivarez-Larrn, and A. Fernndez, "Discrimination between tumour epithelium and stroma via perception-based features," *Neurocomputing*, vol. 154, pp. 119-126, 2015.
- [7] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput*, vol. 18, pp. 1527-54, Jul 2006.
- [8] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," *in Pattern Recognition (ICPR)*, 2012 21st International Conference on, 2012, pp. 3288-3291.
- [9] J. Xu, X. Luo, G. Wang, et al., "A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214-223, 2016.
- [10] https://tma.im/tma-portal/C-Path/index.html.
- [11] http://path.upmc.edu/cases.html
- [12] http://bioimage.ucsb.edu/research/bio-segmentation.
- [13] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345-1359, 2010.
- [14] H. Wang, F. Nie, and H. Huang, "Robust and Discriminative Self-taught learning," *Proceesings of the 30th International Conference on Machine Learning*, pp. 298-306, 2013.

- [15] R. Raina, A. Battle, H. Lee, et al., "Self-taught learning: transfer learning from unlabeled data," *Proceedings* of the 24th international conference on Machine learning, 2007: pp. 759-766.
- [16] Y. LeCun, L. Bottou, Y. Bengio, et al., "Gradient-based learning applied to document recognition," *Proceedings* of the IEEE, vol. 86, pp. 2278-2324, 1998.
- [17] C. Farabet, C. Couprie, L. Najman, et al., "Learning hierarchical features for scene labeling," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, pp. 1915-29, Aug 2013.
- [18] Y. Huang, R. Wu, Y. Sun, et al., "Vehicle Logo Recognition System Based on Convolutional Neural Networks With a Pretraining Strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 1951-1960, 2015.