DEEP ACTIVE LEARNING FOR NUCLEUS CLASSIFICATION IN PATHOLOGY IMAGES

Wei Shao[1], *Liang Sun*[1], *Daoqiang Zhang*[1]

[1]Dept. of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China

ABSTRACT

The systematic study of nuclei patterns in pathology images is very important for fully charactering the grade of cancerous tissues. Nowadays, with the great advance of deep neural networks (*i.e.,* DNN), intense interest in adopting DNN to distinguish different types of pathology nuclei is widely spread. However, most of the existing methods need to annotate lots of nuclei images in the training stage, and this is not always an option for the labelling cost are high. To address this problem, we propose a novel approach called DAPC (*i.e.,* deep active learning with pairwise constraints) to actively select the most valuable nuclei for annotation. Specifically, we firstly design a novel pairwise-constraint regularized deep convolutional neural network (*i.e.,* CNN) that can simultaneously preserve the distribution of different subjects and optimize the objective criterion of conventional CNN. Then, through the properly designed CNN, we query the most informative nuclei in the unlabelled dataset for human annotation, and the parameters of the designed CNN is subsequently updated by incorporating the newly annotated samples to enhance the CNNs performance incrementally. We evaluate our method on a public available pathology colon dataset, the experimental results show that the proposed method could achieves to a weighted F1-score of 79.2% by only annotating 60% nuclei in the training set, which is better than the comparing methods.

Index Terms— Deep Learning, Active Learning, Pairwise Constraints, Cell Nucleus Classification.

1. INTRODUCTION

It is widely recognized that the characteristics (*i.e.,* size ,shape and texture) of nuclei are important factors in diagnosing the grade of cancerous tissues [1-2], and thus one important task in the research of pathology is to distinguish different types of nuclei at cellular level [3]. Recently, with the breakthrough of microscopic imaging technology, scientists are collecting large volumes of hematoxylin-eosin staining images to analysis their covered nucleus patterns[4][5]. Hence, finding an automatic computational way to complete the nuclei classification task has been becoming a new focus in pathology.

From the perspective of machine learning, the task of distinguishing different types of nuclei can be treated as a classification problem, and most of the existing methods have solved it by adopting a two-step framework, where they firstly endeavor to figure out a proper feature representation way for encoding the image data, which then will be fed into an appropriate classifier for category decision. For instance, Cosatto *et al* [4] have used the shape and texture feature for nuclear image description and a AdaBoost classifier for nuclei pleomorphism grading, Yuan *et al* [6] have adopted SVM classifier to classify nucleus into cancer, lymphocyte or stromal based on the morphological features

Nowadays, with the great advance of DNN [7], deep learning approaches have been shown to produce much more satisfactory results than the traditional methods on the task of nuclei classification. In [5], Malon *et al* have trained a deep CNN classifier to classify mitotic and non-mitotic cells. Zhang *et al* [8] naturally integrate deep learning and transfer learning into a single framework for classifying cervical nucleus. In [9], Xu *et al* have proposed a novel stacked sparse autoencoder framework consisting of two sparse autoencoder layers followed by a softmax classifier to distinguish between nuclear and non-nuclear patches. Other efforts include [10] has proposed a novel neighboring ensemble predictor coupled with CNN to more accurately predict the class label of the nucleus.

Although state-of-the-art nucleus classification results can be obtained by applying deep learning method, it requires a large amount of precisely annotated nuclei images to train an effective model. Such training examples are manually annotated by the experts, leading to high cost of both time and money. Active learning methods [11] interactively query labels from the experts, and expect to train the model with less data by querying annotations only for the most valuable data. In this paper, we propose a novel approach called DAPC (*i.e.,* deep active learning with pairwise constraints) to actively select the most informative cell nuclei for annotation. Here, informative means that the selected nucleus can reduce the uncertainty of the learning model [11], and we can eliminate the redundancy of annotation by labelling the informative samples to train the deep learning model.

We show the flowchart of our method in Fig.1. Specifically, we firstly design a novel pairwise-constrained deep CNN
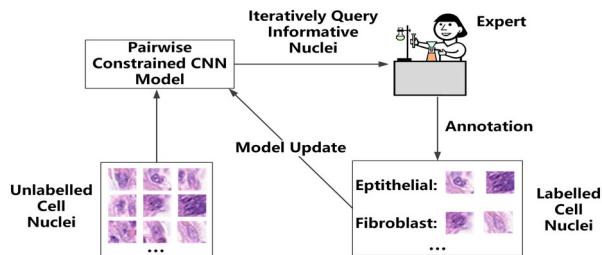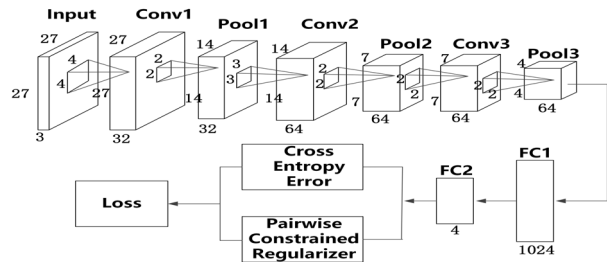
Fig. 1. The flowchart of our method.

Fig. 2. The architecture of the pairwise constrained deep CN-N model

(detailed shown in Fig.2)model, which aims at preserving the distribution information of intra-class subjects and inter-class subjects in the learned deep representation space. Then, through the properly designed deep CNN model, we query the most informative unlabelled nuclei that can reduce the uncertainty of the learning model for human annotation. Next, we update the parameter of the deep CNN model by incorporating the newly annotated samples to enhance the CNN's performance incrementally. The query-update process will repeat several times until satisfactory nuclei classification performance has been achieved. The experimental results show that our method can save 40% annotation efforts for human experts while still achieve better classification performance than the comparing methods.

2. MATERIALS AND METHODS

2.1. Dataset and Preprocessing

In this study, we use the recently publicly released histopathology colon dataset [10] as our benchmark dataset. It contains 22444 cell nuclei images, and the resolution of each image is 27×27. Each nuclear image belongs to one of the four nuclei types, namely, epithelial, inflammatory, fibroblast, and miscellaneous. Here, miscellaneous refers to the nuclei that do not belong to the first 3 categories. (*i.e.,* epithelial, inflammatory, and fibroblast). Within the dataset, there are in total 7722 epithelial, 5712 fibroblast, 6971 inflammatory, and

2039 miscellaneous nucleus patches.

In the pre-processing stage, we firstly augment the nucleus images in the training dataset by rotating them from 0^0 to 270^0 at the intervals of 90^0. Then, in order to avoid the variation of color distribution in histology images, we follow the method in [10], which arbitrary perturbed the color distribution of the training patches in HSV space.

2.2. Pairwise Constrained Deep CNN Model

Fig.2 shows the architecture of the pairwise constrained deep CNN model. Specifically, the first 3 layers (*i.e.,* Conv1, Conv2, and Conv3) are convolutional layers. Different convolutional layers are interleaved with pooling layers (*i.e.,* Pool1, Pool2 and Pool3), which are used to reduce the output from the convolutional layers. The last two layers (*i.e.,* FC1, FC2) are fully connected layers, by which we can obtain high-level representation of nucleus for the final prediction task. It is worth noting that, the main difference between our proposed deep CNN and the conventional deep CNN is that, besides the cross entropy error, we also add pairwise constrained regularization term to optimize the parameters of the deep network.

To be more specific, we firstly introduce some notations used in our method. Let x_i be the i-th patch in the training set $D = \{x_1, x_2, ..., x_n\}$, and $y_i \in \{1, 2, 3, 4\}$ be its corresponding label, denoting if the i-th nuclei patch belongs to Epithelial ($y_i = 1$), inflammatory ($y_i = 2$), fibroblast ($y_i = 3$) or miscellaneous ($y_i = 4$). The deep CNN extracts layer-wise representations of x_i, and we denote the output of its last fully connected layer (*i.e.,* FC2 in Fig.2) as a column vector, $h_i = [h_{i,1}, h_{i,2}, h_{i,3}, h_{i,4}]^T \in R^4$. Then, the deep representation of all the nuclei patches in the training set can be represented as $H = [h_1, h_2, ...h_n] \in R^{4 \times n}$, in which each column is the representation of one sample.

In order to preserve the distribution information of intra-class subjects (*i.e.,* subjects from the same class) and inter-class subjects (*i.e.,* subjects from different classes) in the learned deep representation space, we introduce pairwise constraint knowledge[12], which specify whether a pair of examples belongs to the same class (must-link constraints) or not (cannot-link constraints) to help train the deep CNN model. We define the objective function of the pairwise-constrained CNN, which need to be minimized, as follows:

$$J(H) = c\_entr(\boldsymbol{H}, \boldsymbol{Y}) + \alpha \times p\_cons(\boldsymbol{H}) \qquad (1)$$

Here, $c\_entr(\boldsymbol{H}, \boldsymbol{Y})$ is the cross entropy loss, the second term is the pairwise constrained regularization term, and we show it in Eq.(2).

$$p\_cons(H) = \frac{1}{n_M} \sum_{i,j \in \boldsymbol{M}} \|h_i - h_j\|_2 - \frac{\beta}{n_c} \\ \times \sum_{i,j \in \boldsymbol{C}} \|h_i - h_j\|_2 \qquad (2)$$

Where, n_C and n_M are the number of cannot-link (*i.e.,* C) and must-link (*i.e.,* M) constraints. β is a parameter, which is used to balance the contribution of must-link and cannot-link . The intuition behind Eq.2 is to let the distance between instances involved by the cannot-link constraints in C as large as possible, while the distances between instances involved by the must-link constraints in M as small as possible in the learned deep representation space. It can be expected that more discriminative high-level features can be extracted if we add the pairwise constrained regularizer to the objective function of the conventional deep CNN.

By introducing a coefficient matrix S defined as:

$$S_{i,j} = \begin{cases} 1/n_M & if(h_i, h_j) \in M \\ -\beta/n_C & if(h_i, h_j) \in C \end{cases} \quad (3)$$

Eq.(2) can be formulated as :

$$p\_cons(H) = tr(H(D-S)H^T) = tr(HLH^T) \quad (4)$$

where, D is the diagonal matrix, whose diagonal element is defined as $D_{i,i} = \sum_{j=1}^{n} s_{i,j}$, and *tr(.)* is the trace operator.

2.3. Query Informative Cell Nuclei for Annotation

After the pre-processing step, we have got lots of unlabelled nuclei patches, which are required for annotation. Intuitively, labelling all the nuclei data in the training set would be labor-intensive and time-consuming, and thus our goal aims at achieving high accuracy by annotating as few cell nuclei as possible,thereby reducing the workload for experts.

To be more specific, Suppose, given n nuclei images $D = \{x_1, x_2, \cdots, x_n\}$. Initially, we have l nuclei images annotated by the experts. Without loss of generality, we denote them as $L = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$, where y_i indicates the label of x_i. The remaining $n - l$ nuclei patches form the unlabelled set $U = \{x_{1+1}, x_{1+2}, \cdots, x_n\}$, which are used for active learning. In our proposed active learning algorithm, we iteratively query the best nuclei subset $Q \in U$ with b nuclei patches for annotation.

In this study, we query the most informative nuclei patches for annotation. Here, informative query strategy means that the selected nuclei patches should have the lowest confidence in the existing prediction model [11], and we implement it by picking samples with the largest class entropy. Specifically, we firstly use the labeled nuclei patches L to train the proposed pairwise constrained deep CNN model, and then the deep representation of the unlabelled set U, $H_U = [h_{l+1}, h_{l+2}, ..., h_n]^T$, can be obtained from the last fully connected layer (*i.e.,* FC2) of the trained model. Next, for the deep representation of each unlabelled nucleus, $h_k \in R^4$ $(k = l+1, l+2, \cdots, n)$, we use the softmax function to calculate its class probabilities, $p_{k,j}(j = 1, 2, 3, 4)$, and further adopt these label probability information to compute entropy value of each sample *i.e.,* $E_k = -\sum_{j=1}^{4} p_{k,j} \log p_{k,j}$.

Finally, since the informative samples usually come with larger entropy values, we select b unlabelled nuclei patches with the largest entropy values for annotation and the parameters of the model is then updated by incorporating the newly annotated samples to enhance the CNN's performance incrementally. This query-update cycle repeat several times until acceptable nuclei classification performance has been achieved.

3. EXPERIMENTAL RESULTS

3.1. Experimental Settings

In the procedure of active learning, we first randomly pick 10% nuclei patches from the training set to train the deep C-NN model. Then, at each iteration of active learning, 10% informative nuclei patches are labelled to help update the parameters of the deep CNN. For the deep model settings, we fix the parameter β in Eq.(2) as 10^{-2},while the regularization term α in Eq.(1) is tuned from $\{10^{-2}, 10^{-1}, \cdots, 10^3\}$. We train our models using stochastic gradient descent with a batch size of 80 examples, from which we also randomly select 2000 must and cannot links to train the deep CNN model. We evaluate the performance of our method by weighted F1 score, which can be calculated by weighting the F1 score of each class according to the number of nucleus it contains.

3.2. Comparisons with Random Query Strategy

In order to evaluate the effects of querying the most informative nucleus for annotation, we compare it with the RANDOM query strategy, by which we randomly select nuclei for annotation. Fig.3 shows the weighted-F1 score for different active learning approaches. Here, for the sake of fairness, we use pairwise constrained deep CNN to classify cell nuclei for both random and informative query strategies.

From the results in Fig.3, we can observe that our method performs consistently better than the RANDOM method during the whole active learning process, which shows the advantage of selecting the most informative nucleus for annotation. On the other hand, if all the nuclei patches in the training dataset are chosen to train the pairwise constrained deep CN-N, the weighted F1 score can achieve to 0.794. Rather than querying all the nucleus in the training set, our method can achieve to similar weighted-F1 score, 0.792, by only labelling 60% nucleus in the training dataset, which demonstrate that our method can reduce the high labelling cost for experts.

3.3. Comparisons with the Existing Nuclei Classification Methods.

We also compare our method with DCNN, NEP [10], SSAE [9], and CRImage [6] methods by calculating the F1 score for each class of nuclei. Here, DCNN refers to the conventional deep CNN method without pairwise constraint. For the sake
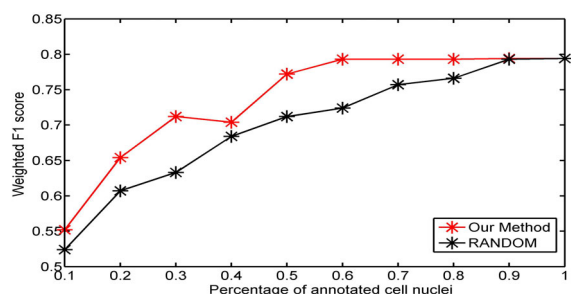
Fig. 3. Comparisons of the weighted-F1 score for different active learning approaches

of fairness, we compare the classification results of DCCN and our method with the same network structure (*i.e.,* shown in Fig.2). It is worth noting that, for our proposed DAPC algorithm, we only annotate 60% nucleus patches in the training set. The results are shown in Fig.4.
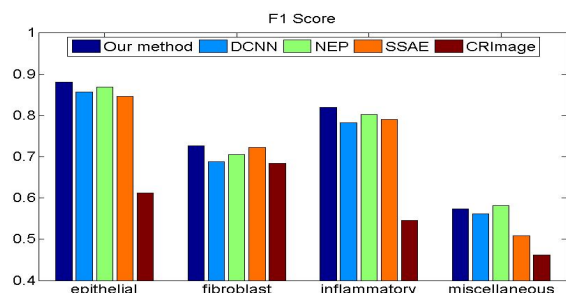

Fig. 4. Comparisons of the F1 score for each nuclei class. For our method we only annotate 60% patches in the training set

As can be seen from Fig. 4, on one hand, our proposed DAPC model is superior to the conventional CNN method, which demonstrates the advantage of incorporating pairwise constrained information to train the deep model. On the other hand, our method only annotates 60% nucleus, while still can achieve better classification results than the comparing methods in most nuclei classes. These results again validate that our method can alleviate the burden of the annotators.

4. CONCLUSION

In this paper, we propose an interactive cell nuclei classification method (*i.e.,* DAPC). The best merit of our approach is its capability of reducing the workload of expert for annotation. The experimental results on a public available colon dataset demonstrate that our method can save experts 40% time for annotation while still can achieve better classification performance when comparing with several existing methods.

5. REFERENCES

[1] D. Lawson,et al, "Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells", *Nature*, vol. 526, no. 7571, pp. 131-135, 2015.

[2] Y. Alkofahi, W. Lassoued, et al,"Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841-852, 2010.

[3] H. Su, F. Xing, et al,"Robust Cell Detection of Histopathological Brain Tumor Images Using Sparse Reconstruction and Adaptive Dictionary Selection", *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1575-1586, 2016.

[4] E. Cosatto, M. Miller,et al, "Grading nuclear pleomorphism on histological micrographs", *In International Conference on Pattern Recognition*, pp. 11-14, 2008.

[5] C. Malon, and E. Cosatto,"Classification of mitotic figures with convolutional neural networks and seeded blob features", *Journal of Pathology Informatics*, vol. 4, no. 1, pp. 9-13, 2013.

[6] Y. Yuan, H. Failmezger, et al,"Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling", *Science Translational Medicine*, vol. 4, no. 157, pp. 143-157, 2012.

[7] Y. LeCun, Y. Bengio, et al, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[8] L. Zhang,et al, "DeepPap: Deep Convolutional Networks for Cervical Cell Classification", *IEEE Journal of Biomedical and Health Informatics*, in press, 2017.

[9] J. Xu, L. Xiang, et al, "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images", *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119-129, 2016.

[10] E. Shan, et al, "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.

[11] B. Settles, "Active learning literature survey", *University of Wisconsin, Madison*, vol. 39, no. 2, pp. 127-131, 2010.

[12] S. Ding , Lin. L, et al, *Deep feature learning with relative distance comparison for person re-identification*, Pattern Recognition, vol.48, no. 10, pp. 2993-3003, 2015.