








Breast Cancer Histopathological Image Classification via Deep Active Learning and Confidence Boosting

Baolin Du^{1,2}, Qi Qi^{1,2}, Han Zheng^{1,2},
Yue Huang^{1,2}, and Xinghao Ding^{1,2}

¹ Fujian Key Laboratory of Sensing and Computing for Smart City,
Xiamen University, Xiamen 361005, Fujian, China
747127841@qq.com, 18150076754@163.com,
420100774@qq.com

² School of Information Science and Engineering, Xiamen University,
Xiamen 361005, Fujian, China
{yhuang2010, dxh}@xmu.edu.cn

Abstract. Classify image into benign and malignant is one of the basic image processing tools in digital pathology for breast cancer diagnosis. Deep learning methods have received more attention recently by training with large-scale labeled datas, but collecting and annotating clinical data is professional and time-consuming. The proposed work develops a deep active learning framework to reduce the annotation burden, where the method actively selects the valuable unlabeled samples to be annotated instead of random selecting. Besides, compared with standard query strategy in previous active learning methods, the proposed query strategy takes advantage of manual labeling and auto-labeling to emphasize the confidence boosting effect. We validate the proposed work on a public histopathological image dataset. The experimental results demonstrate that the proposed method is able to reduce up to 52% labeled data compared with random selection. It also outperforms deep active learning method with standard query strategy in the same tasks.

Keywords: Breast cancer · Histopathological image analysis
Deep active learning · Query strategy

1 Introduction

Breast cancer is ranked as the most common cancer in women worldwide, and it also featured with high morbidity and mortality among women worldwide [1]. The diagnosis by histopathological images under microscopy is one of the golden standards in clinical applications. With the development of imaging sensors, histopathological slides can be scanned and saved as digital images. As the digital image sizes increase dramatically with the magnification, it would be ideal to develop image processing and analysis tools, e.g. classification, in computer-aided diagnosis (CAD) for breast cancer.

Hand-crafted features, such as Scale invariant feature transform (SIFT), histogram of oriented gradient (HOG), gray-level co-occurrence matrix, kernel methods have been reported in the recognition or classification tasks in breast cancer histopathological image analysis. Some well-known classifiers, e.g. support vector machines (SVM), has been reported as well. Recently, deep learning methods, for example convolutional neural networks (CNN), has receive more attention and impressive performances in many tasks of histopathological image processing for breast cancer research, including recognition, classification and segmentation [2]. Chen et al. [3] detected cell mitosis in breast histology images using deep cascading CNN, which dramatically improves detection accuracy over other methods in 2014 ICPR MITOS-ATYPIA Challenge. Wang et al. [4] used CNN, which includes 27-layer breast cancer metastasis test and then won first place in Metastasis Detection Challenge of ISBI2016. Spanhol et al. [5] trained the classification of benign and malignant breast cancer pathological images by Alexnet [6], whose result is 6% higher than the traditional machine learning classification algorithm. Bayramoglu et al. [7] used deep learning to magnification independent breast pathology image classification and the recognition rate is 83%. Spanhol et al. [8] proposed an assessment of BC-recognition for caffeine-free features, increasing the accuracy to 89%. Wei et al. [9] proposed a novel breast cancer histopathological image classification method based on deep convolutional neural networks, named as BiCNN model, resulting to a higher classification accuracy (up to 97%).

The reported state-of-the-art methods strongly rely on the large-scale labeled data in training the network. However, in the view of real-world application, large-scale labeling in medical images are tedious and extremely expensive. Strong professional skills are usually required in the applications compared with annotating natural images. Very limited reports have been contributed to reduce the labeling burden in the proposed task. We proposed a deep domain adaptation method with PCANet and a domain alignment operation to reduce the labeling cost by transferring knowledge from the source dataset to the target one [10]. We also introduced self-taught learning to PCANet to reduce the burden of labeling [11]. However, labeled images in the training data are still randomly selected in the previous works.

In the proposed work, we want to improve the deep learning architecture for the classification task in breast cancer histopathological images by a deep active learning framework. Instead of random selection, active learning methods usually actively select samples with lowest confidence (highest entropy) as valuable samples, and they are added to query, and then the network can be fine-tuned incrementally [12]. In the proposed method, inspired by boosting, the query strategy is also improved, where samples with both high and low confidence are considered simultaneously to emphasize the confidence boosting. We consider that the network should be fine-tuned with additional supervision and its previous regularization simultaneously. The contributions of the proposed work can be summarized as: (1) the labeling cost can be reduced labeling effort with random selection; (2) The method outperforms standard active learning query strategy by the entropy boosting effect.

2 Proposed Method

As a topic in machine learning, active learning is to seek for the most informative samples in a large number of unlabeled dataset actively to annotation query, in order to reduce the labeling effort. We consider introducing the idea of active learning into our method to reduce the labeling cost required for deep learning methods in breast cancer pathological image classification. Firstly, the network is initiated with very limited random selected labeled data. Secondly, the key problem in active learning is how to define the criteria of ‘valuable’ samples. In the standard query strategy, ‘worthness’ is usually defined by the entropy calculated with deep architecture, as:

$$e_i = - \sum_{k=1}^Y \log(p_i^{j,k}) \cdot p_i^{j,k} \quad (1)$$

Where p_i is the confidence value of the network for a sample x_i , and Y represents the number of categories in the work. Entropy captures the uncertainty of classification system in each prediction. A larger entropy value denotes higher uncertainty of the system. In the standard query strategy, active learning methods select a certain number of high-entropy samples to the annotation query until the query size is full. Then the network is fine-tuned with the labeled samples incrementally.

In the proposed work, we believe that the evolution of the network should be fine-tuned incrementally by two factors, the additional supervision from manual labeling and the regulations from previous network. Thus in the proposed query strategy, inspired by the idea of boosting, samples with high entropy values and low entropy values are both considered for a boosting effect. It should be mentioned that the samples with high confidence or low entropy values are labeled by the previous network instead of manual annotation, so there is no additional cost of labeling with the standard active learning query strategy. The algorithm is detailed illustrated as follows.

Algorithm 1: The proposed query strategy.

Input : The training set for the specified dataset B_{train} ; Pre trained CNN M_0 ; Active learning times T_a ; Labelled queue size n ; Compare queue size m

Output : The final fine-tuned CNN model M_{T_a}

- 1 $UntagPool = B_{train}$
- 2 **for** $t = 1: T_a$ **do**
- 3 **for** $Samples \in UntagPool$ **do**
- 4 Entropy, EClass = ComputeEntropy ($M_t, Samples$);
- 5 **end**
- 6 Labelled queue, Compare queue, UntagPool = SelectSamplebyEntropy (Entropy, EClass, n, m);
- 7 Active batch = (Active batch, Labelled queue);
- 8 UntagPool = UpdateUntagPool (UntagPool, Active Batch);
- 9 Train batch = concatenate (Active batch, Compare queue);
- 10 $M_t = TrainNet(M_{t-1}, Train batch)$;
- 11 **end**

As shown in Algorithm 1, let B represents the whole dataset with n_B images, and it is divided into training set and test set B_{train} and B_{test} . The CNN model, denoted as M_0 , is initiated with n_i randomly selected samples in each category, n_i is set to be very small value, for example, two. For convenience, B_{train} is divided into labeled data B_l , and remaining unlabeled data B_u . The sizes of B_l and B_u are n_l and n_u respectively, where $n_l + n_u = n_B$. And n_l and n_u are changing during the incrementally network learning, since the main idea of active learning is to select most valuable samples from B_u to annotation queue A_t for manual annotation. In each query round, the network is fine-tuned with all the labeled samples in A_t , and then n_l turns to $n_l + n$, and B_u turns to $n_u - n$, where n is the size of A_t . A widely-used criteria is to select $n/2$ samples with highest entropy values in each category. The number of query is set to T_a , so the fine-tuned network after each query is denoted as $M_j, j = (0, \dots, T_a)$. In the proposed work, the network in each query is fine-tuned with samples with both high entropy and low entropy. Besides n manual annotated samples, A_t contains additional m samples with lowest entropy values in each category. It should be mentioned that the labels of these m samples are auto-labeled by the previous network.

3 Experiment

3.1 Dataset Description

The proposed framework is evaluated on a public dataset of breast cancer histopathological images, BreKHis [13]. The large-scale dataset contains 7909 images from 82 patients of breast cancer. The dataset is divided into benign and malignant tumors that are scanned with four magnification factors: 40X, 100X, 200X, and 400X. Pathological images are with size of 700×460 in RGB format. The details of the database are shown in Fig. 1.

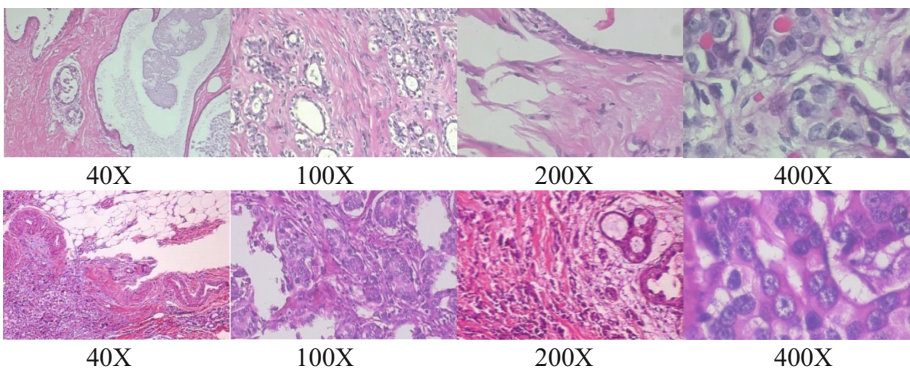


Fig. 1. Breast cancer histopathological image samples in the BreKHis. (Top: benign. Bottom: malignant.)

3.2 Implementation Details

In this section, the proposed algorithm is implemented with TensorFlow framework. The basic CNN architecture is AlexNet pre-trained at ImageNet [14]. The basic settings of the server is intel 2.2-GHz CPU and a NVIDIA GeForce GTX 1080Ti GPU. The dataset has also been divided into training data (70%) and testing data (30%) randomly with no overlapping. In both training and testing set, the size of each category is balanced to be the same. In our work, the proposed work is evaluated on the image-level binary classification, that is, each image is predicted with benign or malignant. Since two categories have been balanced, classification accuracy is used as the metric in the validation, as follow:

$$\text{Image level accuracy} = \frac{N_c}{N_{im}} \quad (2)$$

Where N_{im} the total number of images in the dataset, and the N_c represents the total number of images that are correctly classified.

The network is initiated with one benign sample and one malignant sample randomly selected from the training data. In each experiment, there are 5 query round, where query size for manual labeling in each round is N_m . It should be mentioned that the network is fine-tuned incrementally with 64 labeled images after each query, 48 of them are manual labeling, and the other 16 are auto-labeling.

3.3 Experiment Result

Experimental results on four magnification factors are demonstrated in Fig. 2 and Table 1. It can be observed and concluded from the figures that both standard deep active learning methods and proposed framework have consistent better performances compared with incremental learning with random selection in all the experiments. Deep active learning methods can save up to 52% of the labeling cost compared to random selection to achieve a similar accuracy. This demonstrated that in the view of real-world application, the proposed framework is a better option in recognition task with deep learning methods. It also can be concluded that our proposed method also outperforms deep active learning method with strategy of only high entropy.

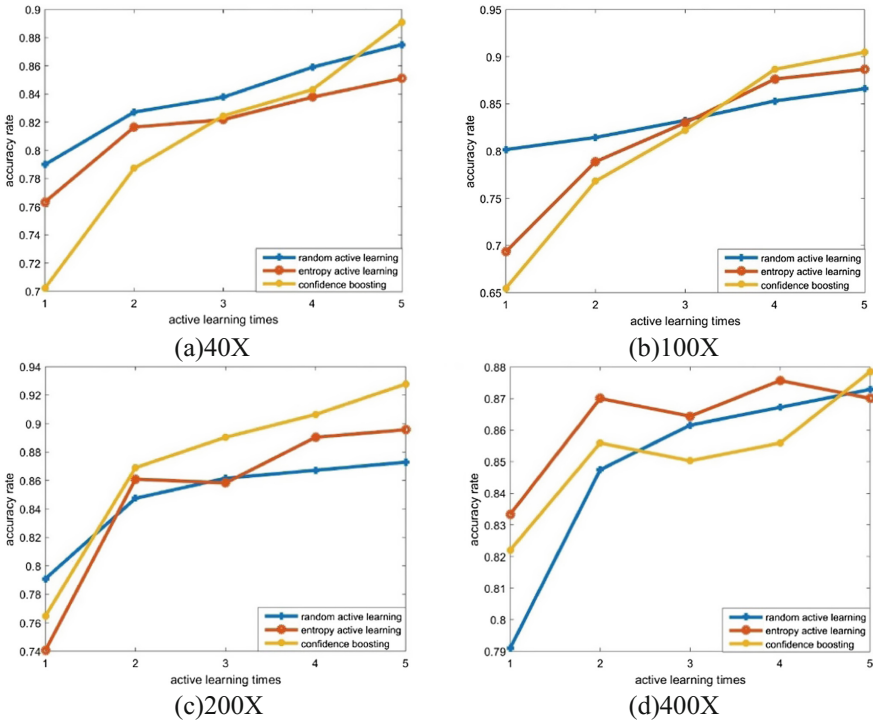


Fig. 2. Comparing the performance of entropy active learning, random active learning and our proposed method in 5 times active learning.

Table 1. Comparing to the annotation cost of our proposed method, random active learning and entropy active learning in similar accuracies. Thereinto, the cost refers to the number of labeled samples, which means the annotation cost.

Strategy	Magnification factors							
	40X		100X		200X		400X	
	Accuracy	cost	Accuracy	cost	Accuracy	cost	Accuracy	cost
Proposed	90.69%	288	90.46%	240	90.64%	192	90.96%	336
EntropyAL	90.96%	500	91.24%	400	91.98%	300	90.11%	450
RandomAL	90.96%	400	90.46%	400	90.37%	400	89.75%	400

4 Conclusions

We proposed a deep active learning framework in histopathological image analysis for breast cancer research. The main purpose of the work is to reduce the tedious labeling burden in the medical application if deep learning methods are used. Instead of randomly selecting samples for annotation as training samples, the framework actively seeking for the most valuable unlabeled data to be manual labeled, and then fine-tune the network incrementally. Besides, we also improve the query strategy with a confidence boosting operation, where both samples predicted with high confidence and low confidence are used in network training in each query round. The samples with high confidence are auto-labeled with the network, so there is no additional manual labeling cost compared with standard active learning methods. The experimental results validated on a large breast cancer histopathological images dataset have demonstrated that our proposed method significantly reduces the labeling cost compared with random selection. It also has better performances with higher accuracy when compared with standard query strategy.

References

1. Lakhani, S.R., Ellis. I.O., Schnitt, S.: WHO classification of tumours of the breast. In: International Agency for Research on Cancer, WHO Press, Lyon (2012)
2. Veta, M., Pluim, J.P.W., van Diest, P.J.: Breast cancer histopathology image analysis: a review. *IEEE Trans. Biomed. Eng.* **2**(5), 1400–1411 (2014)
3. Chen, H., Dou, Q., Wang, X.: Mitosis detection in breast cancer histology images via deep cascaded networks. In: Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona, pp. 1160–1166. AAAI Press (2016)
4. Wang, D., Khoslam, A., Gargeya, R.: Deep learning for identifying metastatic breast cancer. arXiv preprint [arXiv:1606.05718](https://arxiv.org/abs/1606.05718) (2016)
5. Spanhol, F.A., Oliveira, L.S.: Breast cancer histopathological image classification using convolutional neural networks. In: International Joint Conference on Neural Networks, Vancouver, BC, Canada, pp. 2561–2567. IEEE (2016)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, pp. 1097–1105. Curran Associates Inc. (2012)
7. Bayramoglu, N., Kannala, J., Heikkila, J.: Deep learning for magnification independent breast cancer histopathology image classification. In: International Conference on Pattern Recognition (ICPR), Cancun, Mexico, pp. 2440–2445. IEEE (2017)
8. Spanhol, F.A., Cavalin P.R., Oliveira, L.S.: Deep features for breast cancer histopathological image classification. In: IEEE International Conference on Systems, Los Angeles, CA, USA, pp. 1868–1873 (2017)
9. Weil, B., Han, Z., He, X.: Deep learning model based breast cancer histopathological image classification. In: 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, Chengdu, China, pp. 348–353. IEEE (2017)
10. Huang, Y., Zheng, H., Liu, C.: Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. *IEEE J. Biomed. Health Inform.* **21**(6), 1625–1632 (2017)

11. Yue Huang, Han Zheng, Chi Liu: Epithelium-stroma classification in histopathological images via convolutional neural networks and self-taught learning. In: IEEE International Conference on Acoustics, pp. 1073–1077. IEEE, New Orleans, LA, USA (2017)
12. Huang, S.-J., Jin, R., Zhou, Z.-H.: Active learning by querying informative and representative examples. In: International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, pp. 892–900. Curran Associates Inc. (2010)
13. Spanhol, F., Oliveira, L., Petitjean, C.: A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **61**(7), 1455–1462 (2016)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, pp. 1097–1105. Curran Associates Inc. (2012)