ADAPTING FISHER VECTORS FOR HISTOPATHOLOGY IMAGE CLASSIFICATION

Yang Song[1], *Ju Jia Zou*[2], *Hang Chang*[3], *Weidong Cai*[1]

[1]School of Information Technologies, University of Sydney, Australia
[2]School of Computing, Engineering and Mathematics, Western Sydney University, Australia
[3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, USA

ABSTRACT

Histopathology image classification can provide automated support towards cancer diagnosis. In this paper, we present a transfer learning-based approach for histopathology image classification. We first represent the image feature by Fisher Vector (FV) encoding of local features that are extracted using the Convolutional Neural Network (CNN) model pretrained on ImageNet. Next, to better transfer the pretrained model to the histopathology image dataset, we design a new adaptation layer to further transform the FV descriptors for higher discriminative power and classification accuracy. We used the publicly available BreaKHis image dataset for classifying between benign and malignant breast tumors, and obtained improved performance over the state-of-the-art.

Index Terms— Convolutional Neural Network, Fisher Vector, transfer learning, image classification

1. INTRODUCTION

Visual analysis of histopathology images is regularly performed in the clinical routine of cancer management. For example, the differentiation of benign and malignant tumors typically relies on the diagnosis from histopathology images. The manual process is however time-consuming, and computerized approaches that can provide automated classification of histopathology images have been proposed to reduce the workload on pathologists. While some approaches have reported good classification performance with standard or customized features that are handcrafted [1, 2], the recent trend is to learn features automatically especially with the unsupervised methods (e.g. autoencoder) [3, 4, 5].

Supervised models, in particular CNN, have also been applied to histopathology image classification. For example, a CNN model is designed to perform patch-level classification of breast cancer images [6]. The patch-level processing helps to increase the amount of training data, which is essential to achieve high classification performance with CNN. This property renders CNN particularly suitable for problems with naturally large amounts of training data, such as cell detection and segmentation [7, 8]. On the other hand, when the training data is limited, one way to leverage knowledge from more image data is to use CNN models pretrained on ImageNet [9]. The vast amount of images in ImageNet helps to represent a large variety of image patterns, and the resultant CNN models can be transferred to other image datasets effectively. Also, to further adapt the ImageNet models to the specific problem domain, fine-tuning can be performed by training the CNN model up to a certain convolutional or fully connected layer. Such a transfer learning approach has been adopted in various biomedical imaging studies, demonstrating that pretrained models are useful in biomedical applications despite the significant differences in visual characteristics from the general images [10, 11].

In this study, we design a transfer learning approach for histopathology image classification. To represent the images, we first use the CNN model pretrained on ImageNet to extract a dense set of local features. FV encoding [12] is then used to aggregate these local features into a high-dimensional image-level descriptor. Next, we propose to better adapt the FV descriptors to the histopathology images with a new adaptation layer, which is formulated as a locally connected layer in the neural network structure. The transformed FV descriptors are finally classified using Support Vector Machine (SVM) to obtain the image label. Note that in our method, the CNN model is applied at image-level rather than patch-level, and fine-tuning is conducted via the additional adaptation layer rather than the existing layers in the CNN. For evaluation, we used the publicly available BreaKHis database [13], and achieved good performance improvement over existing studies based on handcrafted features [13] and fusion of domain-specific CNN models [6].

2. METHODS

2.1. CNN-based Fisher Vector

FV descriptors represent an image by aggregating dense local features based on Gaussian Mixture Model (GMM). To gen-

This work was supported in part by ARC grants.

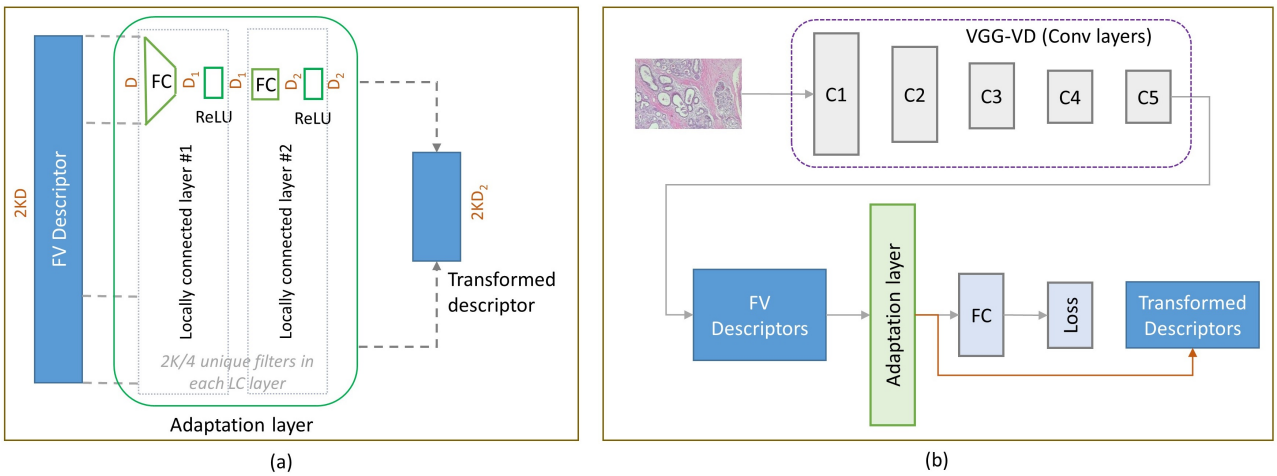

Fig. 1. Illustration of our proposed method. (a) The design of the adaptation layer. (b) The overall flow for training the parameters and computing the feature descriptors.

erate the FV descriptor, local or patch-level features are first computed from the training images. Based on these local features, a GMM with K components is then generated. Given a test image I with N local features, each local feature f_n is soft-assigned to each of the Gaussian components. Then for each Gaussian component, its first and second order differences from each local feature are computed and accumulated based on the soft assignments. Subsequently, these difference vectors for all K Gaussian components are concatenated to produce the FV descriptor h of image I. Note that assuming the local feature f_n has a dimension of D, the dimension of FV descriptor h is $2KD$.

The local features can be computed in many ways. For example, the original FV descriptor uses the Dense Scale-Invariant Feature Transform (DSIFT) features. FV encoding of CNN local features has also shown excellent performance of texture classification in general imaging [14]. In this work, we use the VGG-VD model [15] pretrained on ImageNet to obtain the local features. VGG-VD is a deep CNN model with 19 convolutional and fully connected layers. The last convolutional layer produces a number of local features of $D = 512$ dimensions each, which are used as the dense local features to derive the FV descriptor h.

2.2. Adaptation Layer

Recall that the CNN-based local features are extracted using the VGG-VD model pretrained on ImageNet. It is intuitive to consider fine-tuning the VGG-VD model using the problem dataset before generating the FV descriptors for higher classification performance. The advantage of such fine-tuning has been reported on general imaging [16]. However, we found no performance improvement with this approach on our histopathology image dataset. Instead, we introduce a new type of neural network layer, called the "adaptation layer", to be applied after the FV descriptor is generated. With this adaptation layer, the FV descriptor is further transformed and better adapted to the problem dataset, and the descriptor obtained from the adaptation layer is used to perform image classification using a linear-kernel SVM.

We design the adaptation layer with a locally connected structure. As illustrated in Fig. 1a, given an FV descriptor h of $2KD$ dimensions, we divide the descriptor into $2K$ sections of $D = 512$ dimensions each. Each section is passed through a filter, which is modeled as a fully connected layer of D_1 neurons (we set $D_1 = 64$). The filter weights are locally shared among every four sections, hence there are a total of $2K/4$ unique filters. Each filter is also followed by ReLU activation [9]. The collection of these filters (with ReLU) can be considered as a locally connected layer. Then, another locally connected layer is added with filter size D_2 (we set $D_2 = D_1$) and ReLU, to further transform the descriptors. These two locally connected layers then constitute the adaptation layer, and the output of the adaptation layer has a total dimension of $2KD_2$. The transformed descriptor is then classified using linear-kernel SVM to obtain the image category.

To learn the filters, FV descriptors of training images are generated as inputs to the adaptation layer. A fully connected layer of L neurons (L being the number of classes in the dataset) and a softmax loss function are added to the output of the adaptation layer, as shown in Fig. 1b. During training, to initialize the filter weights, we first train the filters individually by appending a fully connected layer and loss layer to the D_2-dimensional output of each section. These section-wise filter weights are then used to initialize the overall network. Our empirical results show that such initialization provides better classification results than random weight initialization. In addition, when considering the overall network (Fig. 1b),

Table 1. The classification accuracies (%), comparing our method (CNN-based FV descriptor with adaptation layer) with existing results on the BreaKHis dataset, including the benchmark approach [13] (only patient-level results available) and the state-of-the-art [6]. We also include the results using CNN-based FV descriptors only without the adaptation layer. In addition, the results obtained using features derived from the last fully connected (FC) layer of VGG-VD are shown as well.

	Method	Magnification factors			
		40×	100×	200×	400×
Patient level	Handcrafted features with SVM [13]	81.6±3.0	79.9±5.4	85.1±3.1	82.3±3.8
	CNN with random patches [6]	88.6±5.6	84.5±2.4	83.3±3.4	81.7±4.9
	Max pooling of four CNN models [6]	90.0±6.7	88.4±4.8	84.6±4.2	86.1±6.2
	Our method	**90.0±3.2**	**88.9±5.0**	**86.9±5.2**	**86.3±7.0**
	FV without adaptation	90.0±5.8	88.5±6.1	85.4±5.0	86.0±8.0
	FC feature VGG-VD	86.9±5.2	85.4±5.7	85.2±4.4	85.7±8.8
Image level	CNN with random patches [6]	**89.6±6.5**	85.0±4.8	82.8±2.1	80.2±3.4
	Max pooling of four CNN models [6]	85.6±4.8	83.5±3.9	82.7±1.7	80.7±2.9
	Our method	87.0±2.6	**86.2±3.7**	**85.2±2.1**	**82.9±3.7**
	FV without adaptation	86.8±2.5	85.6±3.8	83.8±2.5	81.6±4.4
	FC feature VGG-VD	80.9±1.6	81.1±3.0	82.2±1.9	80.2±3.8

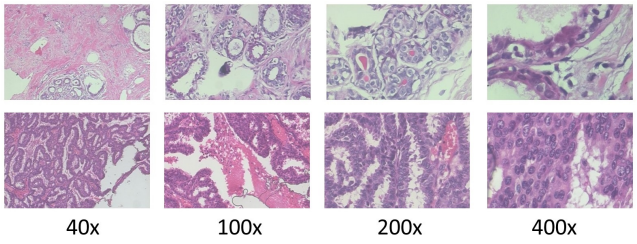

Fig. 2. Sample images of different magnification factors from the BreaKHis database. The top row shows benign cases and the bottom row shows malignant cases.

this learning process is equivalent to backpropagation only to the adaptation layer without affecting the earlier convolutional layers and FV encoding.

2.3. Dataset and Implementation

Experimental evaluation was performed on the BreaKHis database. This dataset contains 7909 hematoxylin and eosin (H&E) stained microscopic biopsy images of benign and malignant breast tumors. There are 2480 benign and 5429 malignant samples, which are collected from 82 patients with varying magnification factors, i.e. 40×, 100×, 200×, and 400×. Each image has 700×460 pixels in RGB format, and example images are shown in Fig. 2. The goal was to classify the images into benign or malignant classes.

To compute the FV descriptors, the images were scaled with multiple factors $(2^s, s = -3, -2.5, \ldots, 1.5)$ so that the CNN-based local features were extracted at multiple scales. For each image, all of its multi-scale local features were aggregated using a single GMM to generate the FV descriptor. With $K = 64$ Gaussian components and $D = 512$ dimensional CNN-based local features, the FV descriptor h was thus $2 \times 64 \times 512 = 65536$ dimensional. For the adaptation layer, we set $D_1 = D_2 = 64$ and the transformed descriptor at the output of the adaptation layer was $2 \times 64 \times 64 = 8192$ dimensional. The learning rate was set to 0.05.

For training and testing, we performed cross validation, using the same folds released with the dataset. A total of five splits were tested, with each split containing 70% of images as training data and 30% as testing data. Images of the same patient were grouped into either the training or testing set only. Similar to the existing study [6], results were measured by classification accuracies at both image and patient levels.

3. RESULTS

The patient- and image-level classification results are listed in Table 1. To compute the patient-level results, following the existing approaches [13, 6], a patient case was classified as benign or malignant based on the majority voting from the classification outputs of the images from the same patient. The results show that our method achieved overall the best performance at different magnification factors for both image- and patient-level classification.

The current state-of-the-art [6] presented a domain-specific CNN model, and four different techniques to generate image patches for training the CNN model from scratch. The techniques included random sampling or sliding window patch selection, with different patch sizes. Among the four techniques, it was shown that random sampling of 1000 patches of 64×64 pixels from each image provided the best result overall. The classification results were further improved by integrating the classification outputs from the four CNN models using max pooling. These approaches have shown large improvement over the initial approach [13], which was based on the Parameter-Free Threshold Adjacency Statistics (PFTAS) feature. This indicates the benefit of using automated feature learning for the histopathology images.

One of the differences between our method and [6] is that we did not train a new CNN model using the image patches from the histopathology images. Instead, we used the pretrained VGG-VD model to obtain the local features, and we used the entire image as the input without subdividing the image into patches. FV encoding then aggregated the local information to represent the image-level characteristics. It can be seen from Table 1 that even without the adaptation layer, these FV descriptors could provide better classification results than the state-of-the-art. However, if image-level features were obtained from the last fully connected layer of VGG-VD, as a more standard way to use a pretrained CNN model, the classification results were lower than the customized CNN models in [6]. These results demonstrate that FV encoding of CNN-based local features from the last convolutional layer was an effective method for transferring the pretrained VGG-VD model to the histopathology images, even though these histopathology images appear different from the general images contained in ImageNet. Then, by transforming the FV descriptors using the adaptation layer, our final results show further improvement in classification accuracy, especially for the seemingly more difficult $200\times$ and $400\times$ cases. Also, we obtained overall higher improvement in image-level classification compared to patient-level classification. This was because with the majority voting, some changes in image label would not affect the patient label.

4. CONCLUSIONS

In this study, we proposed a method for automated classification of histopathology images. Our method comprises two main components. First, image features are computed by FV encoding of CNN-based local features based on the VGG-VD model that is pretrained on ImageNet. Second, the FV descriptors are further adapted to the problem domain with a neural network-based adaptation layer. We applied our method to classify benign and malignant breast cancer images using the publicly available BreaKHis dataset, and achieved improved performance over the state-of-the-art for both image- and patient-level classification.

5. REFERENCES

[1] Y. Xu, J. Zhu, E. I. Chang, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Med. Image Anal.*, vol. 18, no. 3, pp. 591–604, 2014.

[2] M. Kandemir, C. Zhang, and F. A. Hamprecht, "Empowering multiple instance histopathology cancer diagnosis by cell graphs," *MICCAI*, pp. 228–235, 2014.

[3] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin, "Classification of histology sections via multispectral convolutional sparse coding," *CVPR*, pp. 3081–3088, 2014.

[4] A. BenTaieb, H. Li-Chang, D. Huntsman, and G. Hamarneh, "Automatic diagnosis of ovarian carcinomas via sparse multiresolution tissue representation," *MICCAI*, pp. 629–636, 2015.

[5] Y. Song, Q. Li, H. Huang, D. Feng, M. Chen, and W. Cai, "Histopathology image categorization with discriminative dimension reduction of fisher vectors," *ECCV Workshops*, pp. 306–317, 2016.

[6] F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," *IJCNN*, pp. 1–8, 2016.

[7] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "Neutrophils identification by deep learning and voronoi diagram of clusters," *MICCAI*, pp. 226–233, 2015.

[8] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, pp. 1–9, 2012.

[10] G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," *MICCAI*, pp. 652–660, 2015.

[11] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architecture, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, 2016.

[12] F. Perronnin, J. Sanchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," *ECCV*, pp. 143–156, 2010.

[13] F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng*, vol. 63, no. 7, pp. 1455–1462, 2016.

[14] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," *CVPR*, pp. 3828–3836, 2015.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR, arXiv:1409.1556*, 2015.

[16] T. Lin and S. Maji, "Visualizing and understanding deep texture representations," *CVPR*, pp. 2791–2799, 2016.