



Multi-task deep learning for fine-grained classification and grading in breast cancer histopathological images

Lingqiao Li^{1,2} · Xipeng Pan¹  · Huihua Yang^{1,2} · Zhenbing Liu² · Yubei He³ · Zhongming Li¹ · Yongxian Fan² · Zhiwei Cao¹ · Longhao Zhang¹

Received: 23 August 2018 / Revised: 20 November 2018 / Accepted: 27 November 2018 /

Published online: 7 December 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Fine-grained classification and grading of breast cancer (BC) histopathological images are of great value in clinical application. However, automatic classification and grading of BC histopathological images are complicated by (1) small inter-class variance and large intra-class variance exist in BC histopathological images, and (2) features extracted from similar histopathological images with different magnification are quite different. To address these issues, an improved deep convolution neural network model is proposed and the procedure can be divided into three main stages. Firstly, in the representation learning process, multi-class recognition task and verification task of image pair are combined. Secondly, in the feature extraction process, a prior knowledge is built, which is “the variances in feature outputs between different subclasses is relatively large while the variance between the same subclass is small.” Additionally, the prior information that histopathological images with different magnification belong to the same subclass are embedded in the feature extraction process, which contributes to less sensitive with image magnification. The experimental results based on three different histopathological image datasets show that the performance of the proposed method is better than state of the art, with better robustness and generalization ability.

Keywords Multi-task deep learning · Histopathological image classification · Fine-grained · Convolutional neural network · Breast cancer

1 Introduction

Cancer is one of the top deadliest diseases in modern society. Breast cancer (BC), with different subtypes and risk stratification, is the most common cancer in women. According to the International Cancer Research Center (IARC) data released in 2014, BC is the

Lingqiao Li and Xipeng Pan are contributed equally

✉ Huihua Yang
yhh@bupt.edu.cn

Extended author information available on the last page of the article.

second leading cause of cancer death, and its incidence is increasing year by year at a younger age [35]. Detection and diagnosis of BC can be achieved by imaging procedures such as mammography, magnetic resonance imaging (MRI) and ultrasound. However, the analysis of digital pathological images is an important standard for the final diagnosis of BC, and the accurate classification of pathological images is an important basis for treatment plan making by doctors. Therefore, there is important clinical significance for automatic classification of BC pathological images. Histopathological image analysis is a time-consuming and laborious task, and the diagnosis results are easily influenced by many subjective factors. With the help of Computer-Aided Diagnosis (CAD) system, the automatic classification of pathological image can not only improve the efficiency of diagnosis, but also provide more accurate and objective diagnostic results for doctors [44, 46]. Noticeably, fine-grained classification and grading of pathological images [8, 15, 16] are much more significant than binary one [12, 13, 32–34]. They will help patients get accurate diagnosis, guide doctors to develop more scientific and reasonable treatment plan, reduce treatment insufficiency or overtreatment. Last but not least, early detection and intervention can improve the prognosis.

The automatic classification of BC pathological images is a challenging task. The fundamental reasons consist of three aspects: 1) tissue preparation, fixation and other steps are conducted by different personnel with different technology and proficiency, while different dyeing procedures lead to the diversity of appearance changes in pathological slices [28, 29]. 2) Breast pathological images are always characterized by small inter-class variance and large intra-class variance, which bring many particular difficulties for fine-grained classification. 3) Features extracted from similar pathological images with different magnification are quite different, which makes the design of classifier more difficult. Therefore, the classifier should have multi-scale attributes and be independent of magnification. Some details of BC pathological images are shown in Figs. 1 and 2. In Fig. 1, samples (a) - (e) are

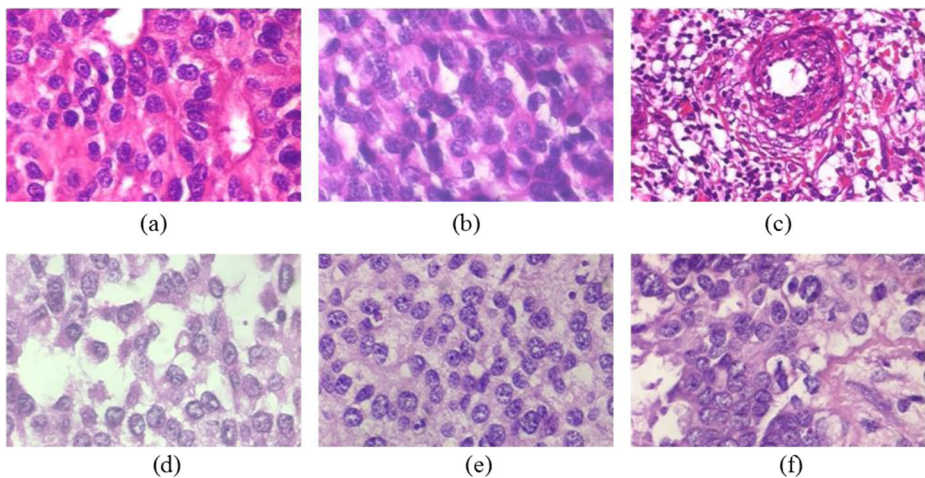


Fig. 1 Breast cancer histopathology images, samples **a-e** are ductal carcinoma(DC), sample **f** is phyllodes tumor. The images are from BreakHis database and the magnification factor of the them is 400x

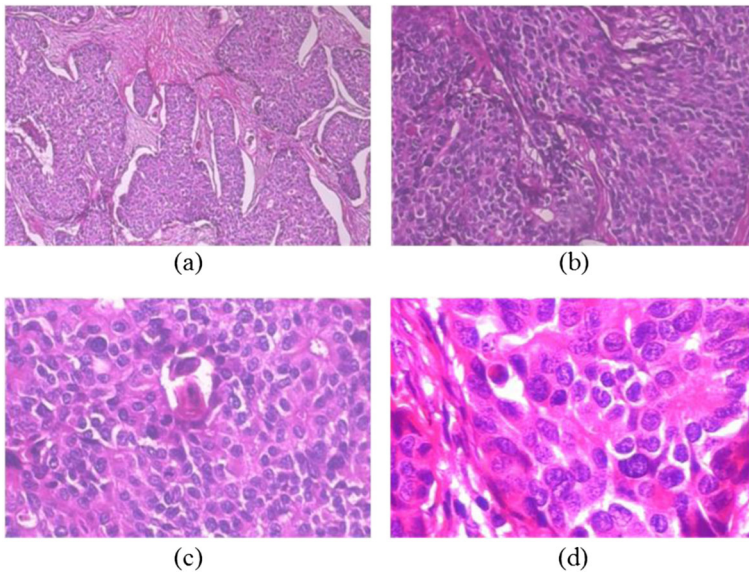


Fig. 2 Slides of breast ductal carcinoma in different magnification factors from the same patient: **a** 40x, **b** 100x, **c** 200x, and **d** 400x. The images are from BreakHis database

ductal carcinoma (DC). Sample (f) is phyllodes tumor. Although samples (a) - (e) are all belong to DC, there are many differences in shades and shapes of cells. At the same time, there is a great similarity in color and cell morphology between samples (e) and (f), but they belong to different classes. The pathological images with different magnification factors are shown in Fig. 2. Although they are all DC and from the same patient, we can see that the differences of visual characteristics among different magnification images are huge.

Spanhol et al. [33] introduced a BC pathological images dataset named BreakHis. Based on the dataset, six kinds of feature descriptors, including local binary patterns, gray level co-occurrence matrix, and four kinds of classifiers, such as support vector machine (SVM) and random forest, are applied to classify the data. The accuracy rate of binary classification is about 80~85%, showing room for improvement is left. At the same time, the complementarity of the magnification factors could be fruitfully investigated in the future. Bayramoglu et al. [2] proposed two different architectures (single task CNN and multi-task CNN) to detect the image magnification level and classify benign and malignant tumors simultaneously. However, fine-grained classification is often very important for accurate diagnosis and personalized treatment. Janowczyk et al. [16] proposed a framework on deep learning for digital pathological image analysis. The framework obtained the promising performance on seven different tasks, but the computing cost was heavy because of the patch-based style and the performance obtained by means of a five steps pipeline.

One limitation of the above methods is that they only employ class labels to drive fine-grained classification of images, while it would be better if class similarity constraint is embedded. Siamese network [6] defines dissimilar and similar image pairs, and specifies

that the distance between dissimilar image pair should be greater than a certain threshold, while the one between similar image pair is smaller. This similarity constraint can effectively obtain features in the process of feature representation learning for many kinds of tasks [3, 36, 38, 43]. An intuitive improvement is to combine the classification and the similarity constraints together for better performance. Therefore, other than using classification constraint alone (e.g., softmax), contrastive constraint is embedded to the procedure of feature representation learning for pathological images. It improved traditional CNN because contrastive constraint will add some prior knowledge for training the model.

In this work, we propose a novel fine-grained classification and grading approach for large-scale complex pathological images. The major contributions lie in three-fold. First, we propose an improved deep convolution neural network model to achieve accurate and precise classification or grading of breast cancer pathological images. Meanwhile, online data augmentation and transfer learning strategies are employed to avoid model overfitting effectively. Second, multi-class recognition task and verification task of image pairs are combined in the representation learning process; in addition to this, a prior knowledge is build, which is “the variances in feature outputs between different subclasses are relatively large while the variance between the same subclass is small”, which will effectively overcome the intractable problem (small inter-class variance and large intra-class variance in pathological image). At the same time, the verification task only related to the category of the pathological image, and is independent of magnification, In other words, the priori information (pathological images with different magnification belong to the same subclass) is embedded in the feature extraction process. It contributes to less sensitive with image magnification. Finally, the experimental results based on three pathological image datasets show that the performance of our method is better than that of state of the arts, with good robustness and generalization ability.

The rest of the work is organized as follow. Recent methods or algorithms for the classification of BC pathological images will be reviewed in Section 2. Section 3 elaborates our methodological contributions in detail, while the experimental results and comparisons with the recently published methods are described in Section 4. Section 5 draws a conclusion to this paper.

2 Related work

A great deal of research progress has been acquired about the automatic classification of BC pathological images. It mainly contains two categories: 1) classification algorithms based on human feature engineering and classical machine learning [21, 24, 25, 39], 2) the recent booming methods based on deep learning.

2.1 Classical classification algorithms

Gupta et al. [11] proposed a framework over multiple magnifications for BC histopathological image classification. The authors employed joint color-texture features and classifiers to demonstrate that some of these features and classifiers were indeed effective. Kowal et al. [17] suggested four different nuclei segmentation methods and deployed them in a medical decision system for BC diagnosis. The classification accuracy was 96~100% for the 500 medical images from 50 patients. In [12, 13], extensive experiments showed that there is no need for stain normalization and the classification can be made magnification invariant when given effective features and ensemble classifiers. Zhang et al. [45]

proposed one-class KPCA model ensemble for medical images classification and the averaged classification accuracy for 361 BC pathological image is about 92%. Wang et al. [40] proposed a framework for cell nuclei segmentation and classification of BC pathological images. For the classification step, 4 shape-based features and 138 textural features based on color spaces are extracted. Optimal feature set is obtained by SVM with chain-like agent genetic algorithm. The proposed method obtained a promising performance on 68 breast cell histopathology images. Dimitropoulos et al. [8] published a dataset with 300 annotated breast carcinoma images of grades 1, 2 and 3, and presented a manifold learning model for grading of invasive breast carcinoma.

It is worth noting that the above classification methods lack a unified standard of comparison, and there is no comparability between the accuracy metric. More importantly, these algorithms are based on manual feature extraction method, which not only needs domain knowledge, but also consumes a lot of time and energy to complete, in addition, the key issue is that extraction of high-quality discriminative features is still challenging [18–20].

2.2 Deep learning methods

Deep learning can automatically learn features from the data, which will avoid the complexity and some limitations of the traditional algorithms. Convolutional neural network (CNN) is a member of deep learning family and has been widely used in the field of machine translation, object detection, visual tracking and image classification [5, 22, 23, 26, 27]. These successful cases provide some references for CNN in the classification of breast pathological images [4, 41, 47]. Spanhol et al. [32, 34] employed AlexNet to extract the deep feature and combined different feature fusion strategies for BC recognition. The performance of the proposed model is much better than the traditional ones. Wei et al. [42] proposed a novel method based on deep CNN (named as BiCNN) to address the two-class BC pathological image classification. This model considered class and sub-class labels of BC as prior knowledge, which could restrain the distance of features of different BC pathological images. Garud et al. [10] presented a GoogLeNet architecture based classification model for the diagnosis of the cell samples using their microscopic high-magnification multi-views. Han et al. [15] employed GoogLeNet as the basic network and proposed a BC multi-classification method. The structured model had achieved remarkable performance on a large-scale dataset, which was a potential tool for BC multi-classification in clinical settings. Inspired by inception modules [37], Akbar et al. [1] proposed a regularization technique named the transition module, which was beneficial to the model generalization ability and the gradual decrease in network size. Zhi et al. [48] investigated using transfer learning on convolutional neural networks (VGGNet and the custom model) to diagnose BC from histopathological images. Song et al. [30, 31] combined convnet with fisher vector (FV) and designed a new adaptation layer to further boost the discriminative power and classification accuracy for histopathology image classification. Most of the existing classification methods of breast pathological images are based on binary classification. However, fine-grained classification of pathological images is of more important significance.

3 Proposed approach

In this paper, we propose an improved fine-grained pathological image classification model based on Xception network [5]. Xception is an improved deep learning model devised by Google which presented excellent classification performance in the large scale image

datasets (ImageNet, JFT). The depthwise separable convolution is used to replace the original convolution operation of Inception V3. However, the number of medical images (pathological images) is often much smaller than that of natural images. The small data number fails to support large capacity model, which tends to be overfitting and fails to achieve good classification performance. To resolve this issue, we adopt two schemes to

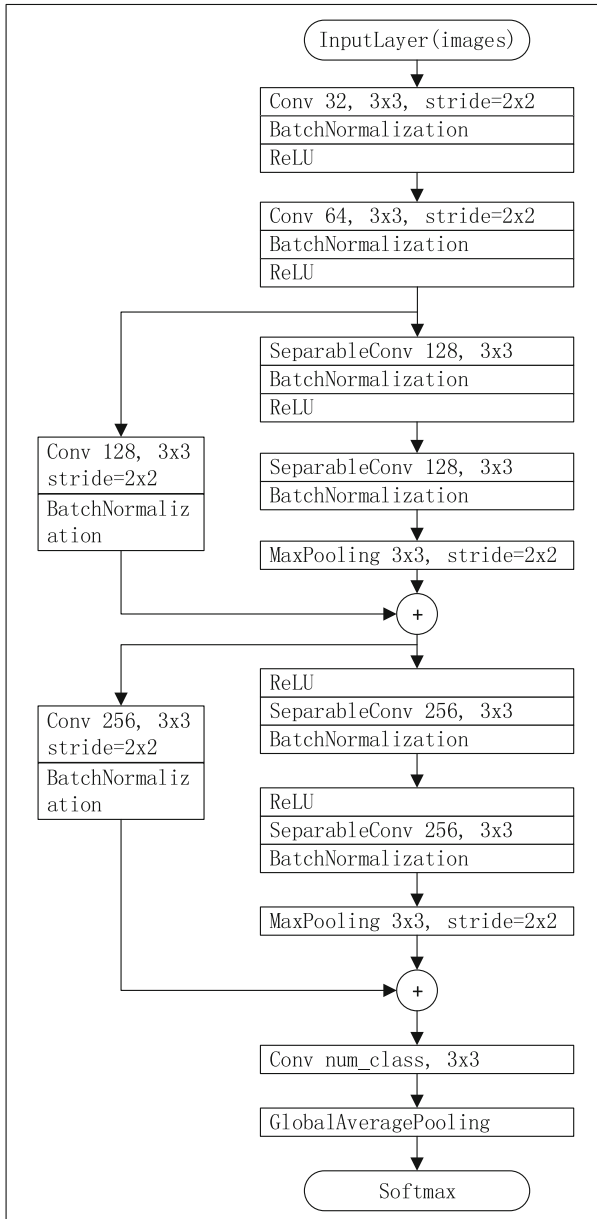


Fig. 3 Xception based network architecture

improve the classification performance of pathological images. First, parts of network layers are extracting from the Xception network to form a new model, which is used to extract the pathological image features. Two prior knowledge that features extracted from different subclasses have relative large distance while features from the same subclass have smaller distance; and pathological images with different magnification belong to the same subclass are embedded in the feature extraction process, which are beneficial to obtain a discriminative model for fine-grained classification of pathological image. Second, transfer learning is employed to fine tune the models trained on ImageNet dataset. More details are explained in Section 3.2.

3.1 Deep convolutional neural networks

(a) Architecture design

In order to prevent overfitting and improve the training speed, we choose parts of network layers in the Xception network to extract the feature of pathological image. The new model architecture mainly composed of input layer, convolution layer, depthwise separable convolution, batchnormalization layer, maxpooling layer and activation function layer (ReLU), which are shown in Fig. 3.

Embedding the prior information in the feature extraction process is beneficial to train a fine-grained classification model for pathological images with strong discriminative ability. The multi-task network model is designed, as shown in Fig. 4. Specifically, the input consists of the paired pathological images, their corresponding labels, and the attribute value that whether the images belong to the same class. The Xception based network structure with the top layer removed and connects with the output of the proposed model, which formulates the training model. The outputs of the proposed model are composed of the softmax probability distribution of the image pair and the distance between the features of the image pair extracted by the network structure. Cross entropy loss is obtained by cross entropy function with softmax probability distribution and one-hot form labels as inputs. Contrastive

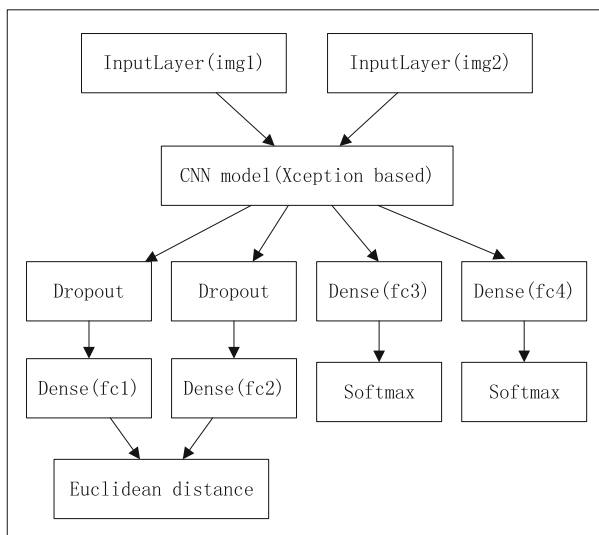


Fig. 4 A network model for multi-task fine-grained classification of pathological images

loss is obtained by contrast loss function with the distance of image pairs as input. The two losses are combined through the weights as the final loss, which is used for the training of the proposed model, as shown in Section 3.3.

(b) Implementation detail

The implementation is based on keras (<https://github.com/fchollet/keras>), and the back-end is tensorflow (<https://github.com/tensorflow/tensorflow>). The data set is randomly divided into three parts: 20% validation set, 20% test set and the other is the training set, none of which overlap with others. The training set is employed for model training and parameter learning; validation set is used to optimize the model and test the model in training process, automatically adjust learning rate and decide whether early stop according to test performances of given training steps; and test set is used for validating the recognition and generalization ability of the proposed model. The experimental results are the mean and variance of the 5 random data set experiments. In order to verify the effectiveness of transfer learning, two training strategies are adopted: random initialization training and transfer learning. Furthermore, the result of only softmax with loss is comparison with the combination with softmax with loss as well as contrastive loss to verify the effectiveness of multitask learning.

3.2 Data augmentation and transfer learning

Lack of large-scale training data is one of the main challenges to apply deep convolutional neural network (DCNN) to medical image classification. However, obtaining large-scale medical images is difficult, and expensive, especially with the professional labels from the pathologists. In order to alleviate the above difficulties, the following two solutions are employed in the classification and grading tasks of three pathological image datasets.

Data augmentation: The training data set is augmented by affine transformation and some data augmentation techniques (e.g., small rotations, zoom, mirror operation, horizontal flipping and vertical flipping) are also applied. In the training process, every batch of images is transformed online with the combination of the above strategy, so as to achieve the purpose of data enhancement. It will save physical storage space and promote the training speed comparing with the traditional off-line data enhancement mode.

Transfer learning: Transfer learning can take advantage of some basic characteristics (e.g., color and edge features) of source dataset, which is beneficial to the classification performance of the target one. In this paper, we will transfer the pre-trained model in ImageNet (including more than 1.2 million natural images and 1000 different categories) datasets to pathological image classification tasks. The specific operation is to freeze the parameters of the shallow layers, and to train the parameters of the high-level layers.

3.3 Multi-task loss

CNN builds a highly nonlinear mapping between input and output through cascaded convolution layers. This hierarchical representation can extract simple and complex features of the network, and different tasks can share the same features. Therefore, CNN is suitable for multi-task learning. Taking this into account, we design a multi-task learning architecture [49, 50], and the steps are as follows.

First, the input image pairs are generated from the training dataset, and the basic unit includes $(x_i, x_j, y_i, y_j, y_{ij})$. Among them, x_i and x_j represent the input images; y_i and y_j are their corresponding labels. If x_i and x_j belong to the same class, the attribute value of

y_{ij} is 1, if x_i and x_j do not belong to the same class, y_{ij} is 0. Second, Euclidean distance between f_i and f_j , which are the extracted features of network with the x_i and x_j as inputs, is calculated. Third, according to the softmax with loss function and the formula of the contrastive loss, the two types of losses can be obtained respectively. The two losses are combined through the weights as the final loss, which is used for the training of the proposed model. Finally, the Nadam [9] optimization method is used for training.

The fine-grained classification features of pathological images are learned by two supervised signals (tasks). The first one is the multi-class recognition signal. In order to divide the pathological images into different categories (for example, 8 categories), a probability distribution of 8 categories is obtained through connecting the 8-way softmax layer after CNN. The network is trained by minimizing the cross-entropy loss, as shown in formula (1).

$$L_{softmax} = L(x, y, \theta) = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{j=1}^k 1\{y_i = j\} \log \frac{e^{\theta_j^T x_i}}{\sum_{j=1}^k e^{\theta_j^T x_i}} \right] \tag{1}$$

where $1(y_i = j)$ is the indicative function, and the rule of value is: $1\{\text{expression is true}\}=1$, and $1\{\text{expression is false}\}=0$. N is the number of images, and k is the number of image categories. θ represents the parameters of the softmax classifier.

The second is the verification signal, which encourages that distance between the features of the same class images is as small as possible and distance between the features of the different types of images is as far as possible. The verification signal can effectively reduce the variance of the extracted features of similar pathological images, and preserve the variance of the extracted features of different pathological images, which will make the model to be more discriminative. Inspired by the literature [14, 36], the following loss function is adopted to constrain the extraction of the features, as shown in formula (2).

$$L_{contrastive} = L(f_i, f_j, y_{ij}, m) = y_{ij}(f_i - f_j)^2 + (1 - y_{ij}) \max(0, m - (f_i - f_j))^2 \tag{2}$$

where f_i , f_j , and y_{ij} have been described above, and m is a learning parameter, which usually set to 1.

Despite its merits in learning feature representation, there still exist several disadvantages for minimizing formula (2) in recognition tasks. For example, given a dataset with N images, the number of all possible pairs of images is N^2 . Each pair of images contains much less information than the classification constraint that provides specific label among k classes, which can lead to a slow convergence. In addition, in the absence of explicit classification constraint, the accuracy of only similar constraint may be inferior to that of traditional CNNs using softmax, especially for fine-grained problems with subtle differences of subclasses. Given the limitations of training with the contrastive loss solely, we constitute these two kinds of losses via a multi-task learning strategy:

$$C = \lambda_s L_{softmax}^1 + \lambda_s L_{softmax}^2 + (1 - 2\lambda_s) L_{contrastive} \tag{3}$$

where $L_{softmax}^1$ and $L_{softmax}^2$ are the softmax with loss of the two inputs, respectively. $L_{contrastive}$ is the contrast loss. λ_s is the weight to control the trade-off between three different losses. Since softmax with loss may contain more information than contrastive loss in each iteration, it is should assign a higher weight to softmax. Actually, in [47], the authors claimed that the performance was not sensitive to small variations to the weight of soft-

max with loss, i.e., within 0.8% difference in a range of [0.55, 0.85]. Guiding by literature [47], we set λ_s as 0.35 in our experiments. Therefore, softmax with loss and contrastive loss accounts for 70% and 30% of the total loss, respectively.

4 Experiments and results

4.1 Datasets description

We evaluated our proposed models based on the experimental results from three different datasets: (1) BreakHis. (2) Grading of invasive breast carcinoma. (3) Lymphoma sub-type classification.

- (1) BreakHis. The dataset consists of 7909 BC histopathology images acquired on 82 patients with different magnification (40x, 100x, 200x, and 400x). This database was built in collaboration with the P&D Laboratory— Pathological Anatomy and Cytopathology, Parana, Brazil (<http://www.prevencaoediagnose.com.br/>). It contains four types of benign breast tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA); and four types of BC: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). The size of images is 700x460, and the mode is RGB three channels (24 bits color, each channel is 8 bits). Table 1 is the specific distribution of the benign and malignant tumor subclass images with different magnification [33].
- (2) Grading of invasive breast carcinoma. This dataset contains cases of breast carcinoma histological specimens received in General Hospital of Thessaloniki, Greece [8]. It consists of 300 annotated images with resolution 1280x960 corresponding to 21 different patients with invasive ductal carcinoma of the breast of grades 1-3 (grade 1: 107, grade 2: 102 and grade 3: 91 images). The image frames are from tumor regions captured through a Nikon digital camera attached to a compound microscope with x40 magnification objective lens, as shown in Fig. 5.
- (3) Lymphoma sub-type classification. The dataset has been prepared by different pathologists from different laboratories to create a real-world type cohort which contains a larger degree of stain and scanning variances [16], as shown in Fig. 6. It consists of 374 images with resolution 1388x1040 and is broken into three subtypes: 113 for the Chronic Lymphocytic Leukemia (CLL), 139 for the Follicular Lymphoma (FL) and 122 for the Mantle Cell Lymphoma (MCL).

Table 1 The specific distribution of the benign and malignant tumor subclass images with different magnification

Magnification	Benign				Total	Malignant				
	A	F	TA	PT		DC	LC	MC	PC	Total
40x	114	253	109	149	598	864	156	205	145	1370
100x	113	260	121	150	614	903	170	222	142	1437
200x	111	264	108	140	594	896	163	196	135	1390
400x	106	237	115	130	562	788	137	169	138	1232
Total	444	1014	453	569	2368	3451	626	792	560	5429
# Patients	4	10	3	7	24	38	5	9	6	58

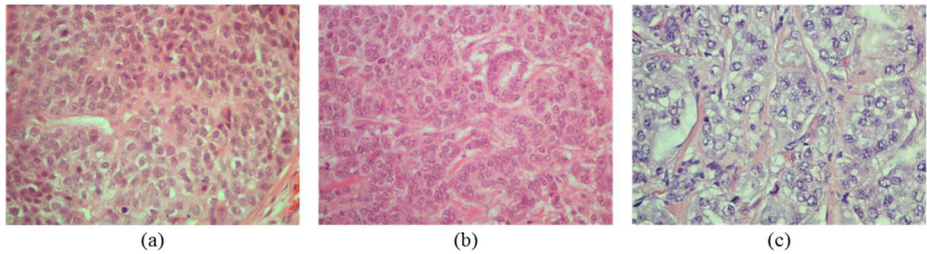


Fig. 5 Grading of invasive breast carcinoma. **a** grade1, **b** grade2, **c** grade3

4.2 Performance metrics

There is only image level information in two of the three experimental datasets, so we calculate the recognition rate from the image level. N_{all} represents the number of validation and test set of pathological images, N_r is the number of pathological images that are correctly classified. So the image level recognition rate can be expressed as:

$$Recognition\ Rate = \frac{N_r}{N_{all}} \quad (4)$$

4.3 Experimental results and analysis

The study is implemented with python 2.7 on a workstation with Intel(R) Xeon(R) E5-2650 v2 CPU, 32 GB memory and the model of GPU is GTX1080. Comparison experiments are conducted between our proposed methods and the recently published state-of-the-art models or approaches.

(1) Experiment on BreaKHis dataset

The BreaKHis dataset is randomly divided into three parts: 20% validation set, 20% test set and the other is training set, meanwhile, none of which overlap with others. The experiment results between our proposed method and the recently state-of-the-arts are shown in Table 2. From Table 2, we can draw several conclusions: 1) the same method obtains the similar performance from the validation set and the test set, indicating that the model has good generalization ability. 2) In general, the performance of multi-task CNN is better than single task CNN, and fine-tuning multi-task CNN is superior to multi-task CNN training from scratch. Xception network is trained from scratch with the pathological images, and the classification results are

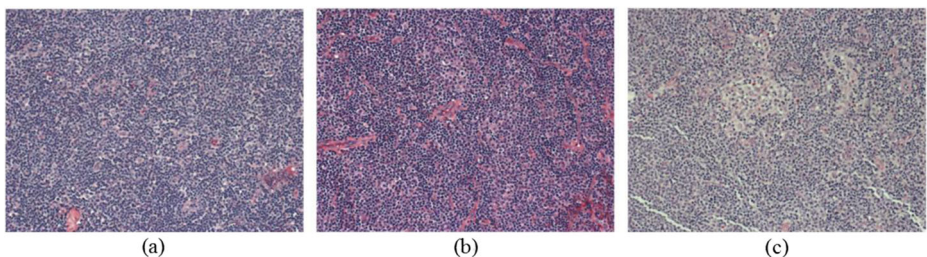


Fig. 6 Lymphoma sub-type Classification. **a** CLL, **b** FL, **c** MCL

Table 2 Comparison of the proposed fine-grained classification method against seven state of the art approaches based on of BreakHis images

Methods	Magnification factors				
	40x	100x	200x	400x	all
ST (valid)	92.42±0.39	93.29±0.81	92.29±0.39	90.81±1.42	92.25±0.26
ST (test)	94.11±0.52	91.77±0.83	92.12±0.51	90.25±2.11	92.08±0.28
MT (valid)	92.89±1.18	91.65±0.73	93.03±0.6	91.42±0.54	92.25±0.3
MT (test)	94.69±0.8	93.8±0.63	92.66±0.68	92.24±1.11	93.36±0.19
MT SCR Xception (valid)	86.69±3.79	83.12±4.61	84.32±4.95	86.54±2.04	85.08±3.85
MT SCR Xception (test)	87.95±2.79	83.67±4.56	85.04±3.63	85.22±2.11	85.50±3.17
MT FT Xception (valid)	95.26±0.49	93.37±1.08	93.09±1.45	91.65±1.66	93.36±0.83
MT FT Xception (test)	94.8±0.33	94.03±1.19	93.85±0.24	90.71±0.71	93.43±0.47
CSDCNN + Raw [15]	89.4±5.4	90.8±2.5	88.6±4.7	87.6±4.1	–
CSDCNN+ Aug [15]	92.8±2.1	93.9±1.9	93.7±2.2	92.9±1.8	–
Dimitropoulos et al. [8]	91.8	92.1	91.4	90.2	91.38
Gupta et al. [12]	86.96	88.92	90.32	86.85	–
Spanhol et al. IJCNN16 [34]	89.6±6.5	85.0±4.8	82.8±2.1	80.2±3.4	–
Spanhol et al. TBE16 [33]	81.6±3.0	79.9±5.4	85.1±3.1	82.3±3.8	–
Song et al. ISBI17 [30]	87.0±2.6	86.2±3.7	85.2±2.1	82.9±3.7	–
Song et al. MICCAI17 [31]	87.7±2.4	87.6±3.9	86.5±2.4	83.9±3.6	–

Abbreviations: ST=single task, MT=multi-task, SCR Xception=training from scratch based on Xception model, FT Xception=fine tuning the Xception model training from ImageNet, Raw=raw dataset, Aug=augmented dataset

Bold symbols represent the maximum values of each column in the tables

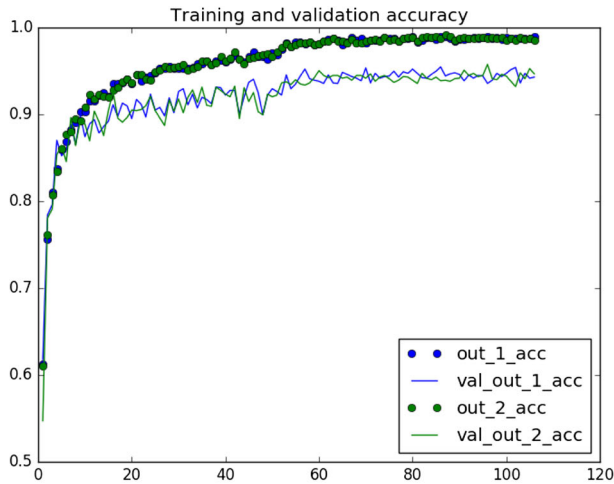
not ideal. The reason may be that the size of the data is so small and the model is not well matched, resulting in overfitting or getting stuck in poor local minima. 3) The results of our method are superior to that of literature [8, 12, 30, 31, 33, 34], although results of these articles provided are binary classification. It is widely known that fine-grained classification (multi-class classification) is more difficult than the binary classification. 4) CSDCNN [15] was an excellent multi-classification method, which obtained very good performance.

The performance of our proposed method (fine-tuning multi-task CNN) is 3% better than that of CSDCNN+Raw [15] in all magnification factors. Most of the results of our proposed method (fine-tuning multi-task CNN) are better than that of CSDCNN+Aug [15], and the performance is slightly worse than that of CSDCNN+Aug [15] when magnification is 400X. However, CSDCNN+Aug [15] has expanded 14 times for the training datasets, and we simply augment the data online, which greatly save the storage overhead.

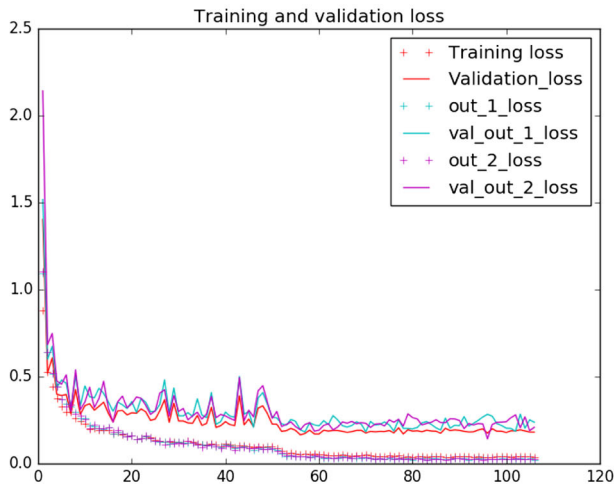
From Fig. 7, we can see that the training was stopped due to the absence of further improvement in validation loss and accuracy after less than 110 epochs. Actually, it begins to obtain a relatively good performance while the epoch is about 60.

(2) Experiment on grading of invasive breast carcinoma dataset

The experiment results between our proposed method and the recently state-of-the-arts are shown in Table 3. From Table 3, we can draw almost the same conclusions as the experiments of BreakHis dataset: 1) the same method obtains the similar



(a)



(b)

Fig. 7 Performance curves when MT FT Xception (multi-task fine tuning the Xception model training from ImageNet). **a** Training and validation accuracy against training steps. out_1_acc and out_2_acc represent the training accuracies of two outputs according to the input images pairs, respectively; val_out_1_acc and val_out_2_acc are the validation accuracies of two outputs according to the input images pairs, respectively. **b** Training and validation loss against training steps. Training loss and validation loss are the overall loss when training and validation phases, which are obtained by formula (3). out_1_loss and out_2_loss represent the training loss of two outputs according to the input images pairs, respectively; val_out_1_loss and val_out_2_loss are the validation loss of two outputs according to the input images pairs, respectively

performance from the validation set and the test set, indicating that the model has good generalization ability. 2) In general, the performance of multi-task CNN is better than single task CNN, and fine-tuning multi-task CNN is superior to multi-task CNN training from scratch. In [8], experiments were implemented with different patch sizes and patching strategies which contained overlapping, non-overlapping and random

Table 3 Comparison of the proposed method against grassmann manifold approach based on grading of invasive breast carcinoma dataset

Methods	Grading of invasive breast carcinoma			
	grade1	grade2	grade3	all
ST (valid)	78.79±2.14	98.41±2.24	92.59±2.62	89.62±2.04
ST (test)	81.82±11.13	96.83±2.24	92.98±2.48	90.32±5.27
MT (valid)	87.88±2.14	98.41±2.24	87.04±2.62	91.26±0.77
MT (test)	84.85±4.29	95.24±3.89	92.98±6.56	90.86±2.01
MT SCR Xception (valid)	78.79±2.14	100.0±0.0	94.44±4.54	90.71±2.04
MT SCR Xception (test)	77.27±9.82	96.83±2.24	98.25±2.48	90.32±4.75
MT FT Xception (valid)	89.39±2.14	98.41±2.24	96.30±5.24	94.54±1.55
MT FT Xception (test)	86.36±7.42	100.0±0.0	92.98±4.96	93.01±1.52
Dimitropoulos et al. 8x8, overlapping [8]	–	–	–	95.8
Dimitropoulos et al. 16x16, overlapping [8]	–	–	–	91.5
Dimitropoulos et al. 32x32, random [8]	–	–	–	85.1
Dimitropoulos et al. 64x64, random [8]	–	–	–	82.3

Bold symbols represent the maximum values of each column in the tables

patches. Image patches of size 8x8 provide the best classification rate (95.8% for overlapping). Other best classification rate of different patch sizes and patching strategies are shown in Table 3. Results show that patches of size 8x8 contain sufficient dynamics and appearance information for the classification of histological images, while the strategy of overlapping patches (with 50% overlap between patches) result in 151,376 Grassmannian points in each histological image.

(3) Experiment on lymphoma sub-type classification dataset

The experiment results between our proposed method and the recently state-of-the-arts are shown in Table 4. For a fair comparison, we divide IICBU dataset into training set and testing set, 75% is the training set, and 25% is the testing set. The experimental

Table 4 Comparison the proposed method against three state of the art approaches based on lymphoma sub-type classification dataset

Methods	Lymphoma sub-type			
	CLL	FL	MCL	all
ST (test)	94.25±3.25	95.18±2.65	90.32±4.56	93.3±3.45
MT (test)	97.7±3.25	96.19±3.56	93.55±6.97	95.79±1.49
MT SCR Xception (test)	90.8±10.66	87.51±5.85	89.25±8.05	89.09±7.5
MT FT Xception (test)	96.55±0.0	94.26±4.02	95.7±1.52	95.43±1.98
Codella et al. [7]	–	–	–	95.5
Song et al. MICCAI17 [31]	–	–	–	96.5±2.7
Janowczyk et al. [16]	–	–	–	96.58±0.01

Bold symbols represent the maximum values of each column in the tables

results are the mean and variance of the 5 random data set experiments, and the performance of our proposed method is almost the same with that of [7] and slightly worse than that of [31]. Song et al. [31] needed to combine FV encoding and multilayer neural network, in addition, their inputs were multiscale and the linear-kernel SVM was adopted as the classifier, which was more complex than our proposed end-to-end DCNN pipeline. At the same time, the performance of lymphoma sub-type classification in [16] was slightly better than our proposed method, but the paper achieved the performance by means of a five steps pipeline ((a) extract patches from all images separated into the 3 sub-types, (b) patches were split into a 5-fold training and testing sets, and 825 k patches were used for training (c) create 5 sets of leveldb training and testing databases, (d) training of DL classifier and (e) use final model to generate the output, a voting scheme per subtype was used where votes were aggregated based on the deep learnings output per patch. The class with the highest number of votes became the designated class for the entire image), which is much more complex than our models.

5 Conclusions

In this paper, we proposed a fine-grained classification and grading model for pathological images. In order to further improve the accuracy of classification, multi-class recognition task and verification task of image pair are combined in the representation learning process; in addition, a prior knowledge is embedded in the process of feature extraction, which will effectively overcome the intractable problem that small inter-class variance and large intra-class variance in pathological image. At the same time, the priori information that pathological images with different magnifications belong to the same subclass is embedded in the feature extraction process, which contributes to less sensitive with image magnification. Both qualitative and quantitative experimental results on BreaKHis, grading of invasive breast carcinoma and lymphoma sub-type classification dataset show that our method obtains the promising performance, and it is superior to several state of the art approaches.

Acknowledgements The authors would like to thank Spanhol et al. [33], Dimitropoulos et al. [8], and Dr. Andrew Janowczyk et al. [16] for publishing the datasets. We would like to express our gratitude to anonymous reviewers and editor for their helpful comments. This research was supported in part by the National Natural Science Foundation of China (Grant Nos. 21365008, 61462018, 61762026 and 61562013), and Natural Science Foundation of Guangxi Province (No. 2017GXNSFDA198025 and 2017GXNSFAA198278). The authors declare that they have no conflict of interest.

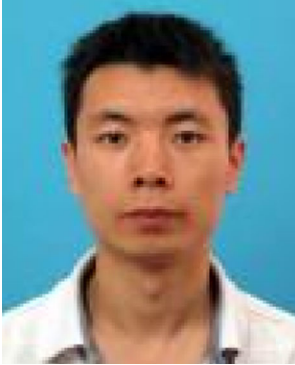
Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Akbar S, Peikari M, Salama S, Nofech-Mozes S, Martel A (2017) Transitioning between convolutional and fully connected layers in neural networks. In: Deep learning in medical image analysis and multimodal learning for clinical decision support (DLMIA)
2. Bayramoglu N, Kannala J, Heikkilä J (2016) Deep learning for magnification independent breast cancer histopathology image classification. In: International conference on pattern recognition (ICPR), pp 2441–2446

3. Bell S, Bala K (2015) Learning visual similarity for product design with convolutional neural networks. *ACM Trans Graph* 34(4):98:1–98:10
4. Chen H, Dou Q, Wang X et al (2016) Mitosis detection in breast cancer histology images via deep cascaded networks. In: Proceedings of the thirtieth AAAI conference on artificial intelligence (AAAI), pp 1160–1166
5. Chollet F (2016) Xception: deep learning with depthwise separable convolutions. arXiv:1610.02357
6. Chopra S, Hadsell R, LeCun Y (2005) Learning a similarity metric discriminatively, with application to face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 539–546
7. Codella N, Moradi M, Matasar M, Sveda-Mahmood T, Smith JR (2016) Lymphoma diagnosis in histopathology using a multi-stage visual learning approach. In: Proceedings of the SPIE, medical imaging 2016: digital pathology
8. Dimitropoulos K, Barmpoutis P, Zioga C, Kamas A, Patsiaoura K, Grammalidis N (2017) Grading of invasive breast carcinoma through grassmannian VLAD encoding. *Plos One* 12(9):e0185110
9. Dozat T (2015) Incorporating nesterov momentum into adam. Technical report, http://cs229.stanford.edu/proj2015/054_report.pdf
10. Garud H, Karri SPK, Sheet D et al (2017) High-magnification multi-views based classification of breast fine needle aspiration cytology cell samples using fusion of decisions from deep convolutional networks. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 828–833
11. Gupta V, Bhavsar A (2017) An integrated multi-scale model for breast cancer histopathological image classification with joint colour-texture features. In: 17th international conference on computer analysis of images and patterns (CAIP), pp 354–366
12. Gupta V, Bhavsar A (2017) Breast cancer histopathological image classification: is magnification important? In: IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 769–776
13. Gupta V, Singh A, Sharma K et al (2017) Automated classification for breast cancer histopathology images: is stain normalization important? In: Computer assisted and robotic endoscopy and clinical image-based procedures (CLIP) 10550, pp 160–169
14. Hadsell R, Chopra S, LeCun Y (2006) Dimensionality reduction by learning an invariant mapping. In: Computer vision and pattern recognition (CVPR), pp 1735–1742
15. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S (2017) Breast cancer multi-classification from histopathological images with structured deep learning model. *Sci Rep* 7(1):4172
16. Janowczyk A, Madabhushi A (2016) Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform* 7(1):29
17. Kowal M, Filipczuk P, Obuchowicz A et al (2013) Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Comput Biol Med* 43(10):1563–1572
18. Lan R, Zhou Y (2016) Quaternion-michelson descriptor for color image classification. *IEEE Trans Image Process* 25(11):5281–5292
19. Lan R, Zhou Y, Tang Y (2016) Quaternionic local ranking binary pattern: a local descriptor of color images. *IEEE Trans Image Process* 25(2):566–579
20. Lan R, Zhou Y, Tang Y (2017) Quaternionic weber local descriptor of color images. *IEEE Trans Circuits Syst Video Techn* 27(2):261–274
21. Lan R, Zhong S, Liu Z et al (2018) A simple texture feature for retrieval of medical images. *Multimed Tools Appl* 77(9):10853–10866
22. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
23. Li P, Wang D, Wang L, Lu H (2018) Deep visual tracking: review and experimental comparison. *Pattern Recogn* 76:323–338
24. Li Y, Lu H, Li KC et al (2018) Non-uniform de-scattering and de-blurring of underwater images. *Mob Netw Appl* 23(4):352–362
25. Lu H, Zhang L, Serikawa S (2012) Maximum local energy: an effective approach for multisensor image fusion in beyond wavelet transform domain. *Comput Math Appl* 64(5):996–1003
26. Lu H, Li B, Zhu J et al (2017) Wound intensity correction and segmentation with convolutional neural networks. *Concurr Comput Pract Experience* 29(6):3927
27. Lu H, Li Y, Chen M et al (2018) Brain intelligence: go beyond artificial intelligence. *Mob Netw Appl* 23(2):1–8
28. Pan X, Li L, Yang H et al (2017) Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing* 229:88–99
29. Pan X, Yang D, Li L et al (2018) Cell detection in pathology and microscopy images with multi-scale fully convolutional neural networks. *World Wide Web-internet & Web Information Systems* 21(6):1721–1743

30. Song Y, Zou J, Chang H et al (2017) Adapting fisher vectors for histopathology image classification. In: International symposium on biomedical imaging (ISBI), pp 600–603
31. Song Y, Chang H, Huang H, Cai W (2017) Supervised intra-embedding of fisher vectors for histopathology image classification. In: Medical image computing and computer-assisted intervention (MICCAI)
32. Spanhol FA, Cavalin P, Oliveira LS, Petitjean C, Heutte L (2017) Deep features for breast cancer histopathological image classification. In: IEEE international conference on systems, man, and cybernetics (IEEE SMC), pp 1868–1873
33. Spanhol FA, Oliveira LS, Petitjean C et al (2016) A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng* 63(7):1455–1462
34. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) Breast cancer histopathological image classification using convolutional neural networks. In: 2016 international joint conference on neural networks, pp 2560–2567
35. Stewart B, Wild CP (2014) World cancer report 2014. International Agency for Research on Cancer, World Health Organization
36. Sun Y, Chen Y, Wang X et al (2014) Deep learning face representation by joint identification-verification. In: International conference on neural information processing systems (ICNIPS), pp 1988–1996
37. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al (2015) Going deeper with convolutions. In: Computer vision and pattern recognition (CVPR)
38. Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 1701–1708
39. Wang D, Lu H, Li X (2011) Two dimensional principal components of natural images and its application. *Neurocomputing* 74(17):2745–2753
40. Wang P, Hu X, Li Y, Liu Q, Zhu X (2016) Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Process* 122:1–13
41. Wang D, Khosla A, Gargeya R et al (2016) Deep learning for identifying metastatic breast cancer. arXiv:1606.05718
42. Wei B, Han Z, He X, Yin Y (2017) Deep learning model based breast cancer histopathological image classification. In: 2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA)
43. Xiao T, Xu Y, Yang K et al (2015) The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 842–850
44. Zhang S, Metaxas D (2016) Large-scale medical image analytics: recent methodologies, applications and future directions. *Med Image Anal* 33:98–101
45. Zhang Y, Zhang B, Coenen F et al (2014) One-class kernel subspace ensemble for medical image classification. *EURASIP J Adv Signal Process* 2014:17. <https://doi.org/10.1186/1687-6180-2014-17>
46. Zhang X, Su H, Yang L, Zhang S (2015) Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
47. Zhang X, Zhou F, Lin Y, Zhang S (2016) Embedding label structures for fine-grained feature representation. In: 29th IEEE conference on computer vision and pattern recognition (CVPR), pp 1114–1123
48. Zhi W, Yueng HW, Chen Z et al (2017) Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images. In: 24th international conference on neural information processing (ICNIP)
49. Zhu X, Li X, Zhang S (2016) Block-row sparse multiview multilabel learning for image classification. *IEEE Trans Cybern* 46(2):450–461
50. Zhu X, Suk HI, Lee SW et al (2016) Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Trans Bio-Med Eng* 63(3):607–618



Lingqiao Li is a Ph.D. candidate with School of Automation, Beijing University of Posts and Telecommunications, China. Currently he is assistant researcher at the Guilin University of Electronic and Technology, China and his research interests include machine learning, spectrum analysis.



Xipeng Pan is a Ph.D. candidate with School of Automation, Beijing University of Posts and Telecommunications, China. His research interests include machine learning, deep learning, medical image processing. He is a member of CCF.



Huihua Yang received his Ph.D. degree from East China University of Science and Technology, China in 2005. He was a postdoctoral research fellow of Tsinghua University from 2005 to 2007. Currently, he is a professor of School of Automation, Beijing University of Posts and Telecommunications, China. His research interests include machine learning, spectrum analysis, and optimization. Dr. Yang has published more than 40 papers and serves as Director of China Instrument and Control Society (CICS), Vice Director of NIR Division of CICS, and is a senior Member of CCF, and a member of ACM.



Zhenbing Liu received his Ph.D. degree in Computer Science from Huazhong University of Science and Technology, China in 2010. He was a visiting research fellow at Pennsylvania University in 2015. Currently, he is a professor of Guilin University of Electronic Technology, China. His research interests include machine learning and medical image processing. He has published more than 30 papers.

Affiliations

Lingqiao Li^{1,2} · Xipeng Pan¹  · Huihua Yang^{1,2} · Zhenbing Liu² · Yubei He³ · Zhongming Li¹ · Yongxian Fan² · Zhiwei Cao¹ · Longhao Zhang¹

✉ Zhenbing Liu
zbliu@guet.edu.cn

Lingqiao Li
54pe@163.com

Xipeng Pan
pxp201@bupt.edu.cn

- ¹ School of Automation, Beijing University of Posts and Telecommunications, No. 10, Xitucheng Road, Haidian District, Beijing, China
- ² School of Computer Science and Information Security, Guilin University of Electronic Technology, No. 1 Jinji Road, Qixing District, Guilin, China
- ³ School of Computing and Information Systems, University of Melbourne, Melbourne, Australia