Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification

Yushan Zheng [a,b], Zhiguo Jiang [a,b], Fengying Xie [a,b,*], Haopeng Zhang [a,b], Yibing Ma [a,b], Huaqiang Shi [c,d], Yu Zhao [c]

[a] *Image Processing Center, School of Astronautics, Beihang University, Beijing, 100191, China*
[b] *Beijing Key Laboratory of Digital Media, Beijing, 100191, China*
[c] *Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen, 361101, China*
[d] *General Hospital of the Air Force, PLA, Beijing, 100036, China*

A R T I C L E   I N F O

A B S T R A C T

Feature extraction is a crucial and challenging aspect in the computer-aided diagnosis of breast cancer with histopathological images. In recent years, many machine learning methods have been introduced to extract features from histopathological images. In this study, a novel nucleus-guided feature extraction framework based on convolutional neural network is proposed for histopathological images. The nuclei are first detected from images, and then used to train a designed convolutional neural network with three hierarchy structures. Through the trained network, image-level features including the pattern and spatial distribution of the nuclei are extracted. The proposed features are evaluated through the classification experiment on a histopathological image database of breast lesions. The experimental results show that the extracted features effectively represent histopathological images, and the proposed framework achieves a better classification performance for breast lesions than the compared state-of-the-art methods.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer is the second most commonly occurring cancer for females. According to an estimation from the American Cancer Society in 2017 [1], around 30% of the new cancer cases in American women consist of breast cancer. New technologies [2–4] for breast cancer diagnosis have been developed in recent years. However, final diagnosis still currently relies on biopsies [5]. Owing to the development in digital pathology, the whole slide image (WSI) of a histopathological section can be captured within a few minutes by a micro scanner, and stored in a data server. These cancer images provide valuable data sources for researches in computer science and related areas, such as image processing, pattern recognition, and data mining. As a result, many approaches to histopathological image analysis [6–8] have been developed and applied to practical clinical diagnoses, relieving the workload of pathologists and assisting pathologists in making more reliable and consistent diagnoses.

Across various applications of histopathological image analysis, image classification and content-based image retrieval (CBIR) are always important challenges. Histopathological image classification aims to directly predict the types of lesions (for example, the classification of benign and malignant tumors). A reliable classification system can provide a verification of the diagnosis of the doctor. Meanwhile, CBIR can search and return cases with similar content to the query image. Using information regarding diagnoses of similar cases for reference, doctors can reach a more reliable diagnosis for the query image. For these applications, feature extraction plays a crucial role. However, histopathological images may contain hundreds of structures [9] and the appearance within the same lesion is varied [10], which makes feature extraction a challenging task.

In this paper, a novel nuclei-guided feature extraction method based on convolutional neural network is proposed for histopathological images stained by hematoxylin and eosin (HE). The nuclei in the image are first detected. Considering the location information of the nuclei, a fine-designed neural network is trained to extract features regarding the pattern and distribution of the nuclei. Through classification experiments on a breast-lesion database, the

---

* Corresponding author at: Image Processing Center, School of Astronautics, Beihang University, Beijing, 100191, China.

E-mail addresses: yszheng@buaa.edu.cn (Y. Zheng), jiangzg@buaa.edu.cn (Z. Jiang), xfy_73@buaa.edu.cn (F. Xie), zhanghaopeng@buaa.edu.cn (H. Zhang), mayibing@buaa.edu.cn (Y. Ma), shihq@motic.com (H. Shi), zhaoy@motic.com (Y. Zhao).

proposed features are validated to be effective for histopathological images.

The remainder of this paper is organized as follows. Section 2 reviews relevant work regarding histopathological image analysis. Section 3 introduces the proposed feature extraction method. The experiment and discussion are presented in Sections 4 and 5. Finally, Section 6 summarizes the present contributions and suggests directions for future work.

2. Related work

Many feature extraction methods have been proposed for histopathological images. They can be broadly classified into two categories: statistics based method and learning based method.

Inspired by the success of natural image analysis, some researchers [11–14] have employed classical feature descriptors, such as color histogram, scale-invariant feature transform (SIFT) [15], histogram of oriented gradient (HOG) [16], and local binary pattern (LBP) [17], to depict histopathological images. Although the characteristics of cancer in histopathological images are quite different from those in natural images, classical features have achieved considerable success in histopathological image analysis. Simultaneously, several researchers [18–20] have tended to describe images in terms of pathological aspects. In [19], the statistics regarding the shape, size, distribution of the nuclei are computed to give the histopathological image representation. The statistics can describe the general status of the nuclei, but may weaken the response of key nuclei that are few but important for diagnosis. Further, to describe meaningful objects in the histopathological images (such as glands and ducts), graph-based features [21,22] were introduced to histopathological image analysis. To establish more comprehensive representations, mixed features [23–29] have been proposed, and achieve a better performance. In [27], Kandemir et al. represented the histopathological images by mixing color histogram, SIFT, LBP and a set of well-designed nuclear features, and analyzed the necessities of patterns from color, texture and nuclear statistics.

With the development of machine learning, several feature learning methods, including auto-encoders [30–32], restricted Boltzmann machines [33,34], sparse representation [35–37], and convolutional neural network (CNN) [38–40], have been introduced to mine patterns from the raw pixel data of histopathological images. In [33], sparse features were extracted from patches using a restricted Boltzmann machine, and then quantified into the image-level feature through a max-pooling process. In [30], sparse auto-encoders are applied to learn a set of feature extractors for histopathological images. Then, the feature maps are obtained through convolutional operations on the feature extractors and the image. Furthermore, in [31], a stacked feature representation was appended to the former framework. In [39], Xu et al. applied a deep convolutional neural network to analyze and classify the epithelial and stromal regions in histopathological images. In these approaches, the features are extracted patch by patch throughout the histopathological image, which ignored the characteristics of histopathological images and resulted in a high computational complexity. In contrast, Srinivas et al. [36] proposed a sparsity model to encode cellular patches that selected by a blob detection approach, and then classified histopathological images by fusing the predictions of these cellular patches. As a continuation of this work, Vu et al. [37] proposed a class-specific dictionary learning method for the sparsity model. Using this dictionary, more discriminative sparse representations of image patches were extracted. This patch-selected scheme of [36,37] sharply reduces the running time, but often failed to identify the malignant samples when the malignant area is minority in the image.

In this paper, we propose a novel nuclei-guided convolutional neural network to learn features from histopathological images.

Unlike in a normal CNN, the activations on the first convolutional layer are limited to the key patches, which are selected by an efficient nucleus detection algorithm. Compared with existing patch-based methods [36,37,41,42], in which the image-level prediction is obtained by fusing the predictions of key patches, we generate the patch-level and image-level features and the image classification criteria simultaneously during an end-to-end learning process. Through a classification experiment on a breast-lesion database, the proposed features are verified to be both effective and efficient for histopathological image representation and classification.

3. Method

According to pathologists [5,43], the appearance of the cell nucleus and its surrounding cytoplasm, as well as the distribution of nuclei, are important indicators for cancer diagnosis. Therefore, both the appearance and distribution of nuclei are considered. The proposed framework consists of an offline training stage and an online encoding stage, where the former can be divided into three steps: nuclei detection, pre-training, and fine-tuning.

Fig. 1 presents a flowchart of the training stage. the cell nuclei are first detected. Then, a fine-designed neural network is constructed under the guide of nucleus locations to extract image-level features. The neural network consists of three basic structures from the bottom to the top: the patch-level, block-level, and image-level structure. The patch-level structure contains a convolutional layer and a non-linear layer. Both the block-level and image-level structures consist of a pooling layer, a full-connecting layer, and a non-linear layer.

Plenty of labeled histopathological images are required to train an effective deep neural network if the biases and weights of the network are randomly initialized. However, it is a difficult task to produce a complete annotation for a whole slide image (WSI), even for an experienced pathologist. Therefore, we proposed to first pre-train the network with abundant unlabeled regions, and then fine-tune it using the labeled images.

3.1. Nucleus detection

In this paper, the nuclei are located using a blob detection method. To remove noises caused by stroma, the nucleus component $\mathbf{H}$ (i.e., the optical density of the hematoxylin component in the HE-stained histopathological image) is first extracted using the color deconvolution algorithm [44]. Then, a Gaussian filter is employed to highlight the cell nuclei in $\mathbf{H}$, thus providing a probability map of cell nuclei:

$$\mathbf{H}_{nuc} = \mathbf{H} * \mathbf{G}(\xi), \tag{1}$$

where $*$ denotes the convolution operation, $\mathbf{G}(\xi)$ is a Gaussian with mean 0 and scale ξ. In this paper, ξ is determined experimentally as 6.5, and the size of the Gaussian filter is 13×13, which can cover most of the nuclei under a $20 \times$ lens.

By searching for local maxima in $\mathbf{H}_{nuc}$, the nuclei are coarsely detected. At the same time, some noise will also be detected. Considering that these noises have weak responses in $\mathbf{H}_{nuc}$, a threshold is used to filter them out. Fig. 2 illustrates an instance of nuclei detection, where (a) is the original image and (e) shows the detection result. It can be seen that the nuclei are located effectively. In next section, patches centered on nuclei are sampled (see the red box in Fig. 2(e)), and used to mine the histopathological features.

3.2. Pre-training

Auto-encoder(AE) can remove redundancies and noises, and mine the discriminative pattern for unlabeled data, which has
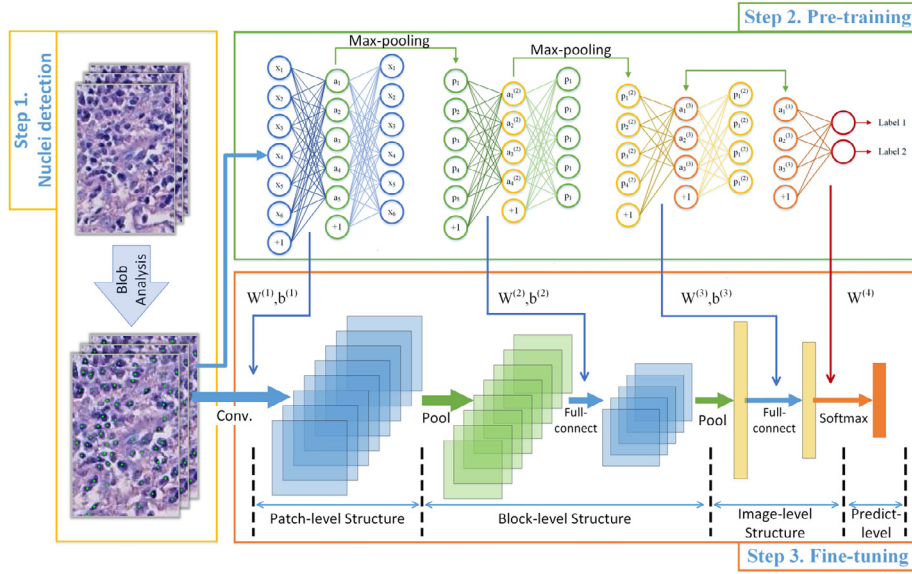
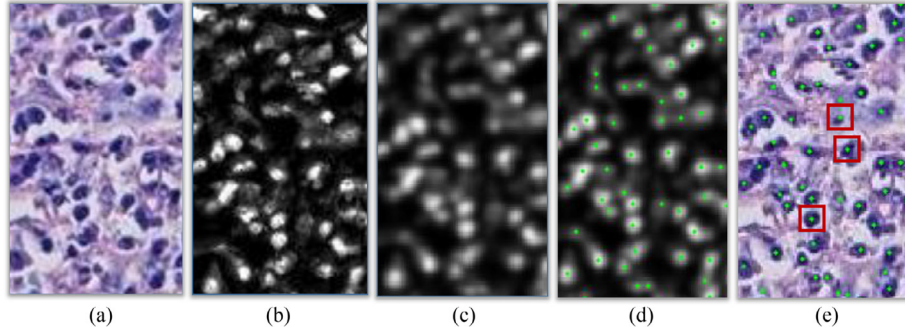**Fig. 1.** Flowchart of the training stage.

(a) (b) (c) (d) (e)

Fig. 2. An instance of nuclei detection. Here, (a) is the original image, (b) is the hematoxylin component separated by color deconvolution, (c) is the result obtained by Gaussian filtering, (d) shows the positions of local maxima, and (e) is the detection result, in which some instances of nucleus patches are displayed in red boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proven effective in pre-training network parameters [45,46]. However, for histopathological images, the difference among nuclei appearance is much more important for the diagnosis of breast lesions than the difference between nuclei and stroma. If AE is conducted on entire image, the discrimination among nuclei appearance in the extracted features will be weakened. Therefore, in our method, only nucleus locations are considered in the training of AE.

3.2.1. Patch-level structure

A convolutional layer is first employed to extract basic features from the raw pixel data. In this paper, only nuclei are considered for constructing the feature extraction network, thus the computation for stromal regions are unnecessary. As shown in Fig. 2(e), the nuclei are spatially sparse and scarcely overlap with each other. Therefore, we propose to take nuclei as independent samples and use full-connecting operation to replace the convolutional operation, which can simplify the network and reduce the computational complexity. Specifically, the nucleus-centered patches having the same size as convolution kernel (weight matrix) are extracted from the image, and the pixel data of each patch is flattened as a column vector. Let $\mathbf{x}_i$ denote the column vector of the ith patch, the encoding can be represented as

$$\mathbf{a}_i = \sigma(\mathbf{W}^{\mathrm{T}}\mathbf{x}_i + \mathbf{b}), \tag{2}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$ and $\mathbf{b} = [b_1, b_2, \ldots, b_K]^{\mathrm{T}}$ are the weights and bias, $\mathbf{w}_k$ is the flattened weights vector of the kth weight matrix, K is the neuron number of the hidden layer, and σ is the activation function. In this paper, σ denotes the sigmoid function $\sigma(t) = 1/(1 + e^{-t})$. To fit the input of sigmoid function, the raw pixel data of the patches is normalized. In this paper, the raw pixel data is truncated to ± 3 times standard deviations and then squashed to [0, 1]. The corresponding decoding is

$$\hat{\mathbf{x}}_i = \sigma(\tilde{\mathbf{W}}^{\mathrm{T}}\mathbf{a}_i + \tilde{\mathbf{b}}), \tag{3}$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{b}}$ are the weights and bias of the decoding stage, respectively. $\hat{\mathbf{x}}_i$ is the reconstruction of the patch data $\mathbf{x}_i$. Then, the objective function to be minimized is defined as the mean squared error (MSE):

$$J = \frac{1}{2n}\sum_{i=1}^{n}(\hat{\mathbf{x}}_i - \mathbf{x}_i)^2 + \frac{\lambda}{2}(\|\mathbf{W}\|_F^2 + \|\tilde{\mathbf{W}}\|_F^2), \tag{4}$$

where n denotes the number of nuclei used in pre-training stage and λ is the weight of regularization. In general, an additional constraint is required in Eq. (4) to ensure a robust performance, and in this paper the sparse constraint is used to generate the cost function of the sparse auto-encoder (SAE) [45,47].:

$$J_{sparse} = J + \beta \sum_{j=1}^{d} KL(\rho \| \hat{\rho}_j), \tag{5}$$
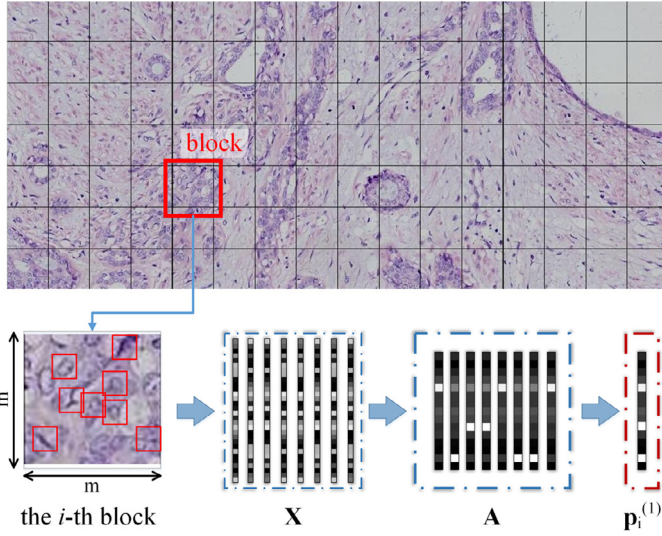
Fig. 3. Block distribution in the image. The nucleus pixels of the ith block are represented as $\mathbf{X}$, with which the activations of each nucleus are calculated by Eq. (2) and collected in the matrix $\mathbf{A}$. Then, the block-level representation $\mathbf{p}_i^{(1)}$ can be obtained by max-pooling.

where β is the weight of the sparse constraint, and $KL(\rho\|\hat{\rho}_j)$ is the Kullback-Leibler divergence between ρ and $\hat{\rho}_j$, which is defined by

$$KL(\rho\|\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j},$$

where $\hat{\rho}_j$ denotes the average value of the activations on the jth hidden unit over all the training patches, $\hat{\rho}_j = 1/n \sum_{i=1}^{n} a_{ij}$ and ρ is a constant that defines the prospective average activation. For the sparse auto-encoder, ρ constrains the degree of the sparsity of the hidden representations. The smaller ρ is, the more the semantics will be mined [47].

Finally, the optimal parameters can be obtained by solving the equation

$$(\mathbf{W}^*, \mathbf{b}^*, \tilde{\mathbf{W}}^*, \tilde{\mathbf{b}}^*) = \arg \min_{\mathbf{W}, \mathbf{b}, \tilde{\mathbf{W}}, \tilde{\mathbf{b}}} J_{sparse}. \quad (6)$$

Through Eq. (6), the hidden representations of nuclei patches ($\mathbf{a}_j$) are optimized and used for the encoding of block-level structure.

3.2.2. Block-level structure

The image-level features can be quantified by simply pooling all of the activations of nucleus patches within the image. However, this strategy disregards the spatial distribution of nuclei. To preserve the distribution information, the block-level structure is inserted between the nucleus-level and image-level structures. As shown in Fig. 3, the image is divided into square blocks of $m \times m$ pixels, in which the nucleus activations are quantified.

Let $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N)$ be the set of nucleus activations that are located in the ith block. Then, the quantified representation is defined as the max-pooling result of $\mathbf{A}$:

$$\mathbf{p}_i^{(1)} = (p_{i1}^{(1)}, p_{i2}^{(1)}, \ldots, p_{iK}^{(1)})^T = (\|\mathbf{A}_1\|_\infty, \|\mathbf{A}_2\|_\infty, \ldots, \|\mathbf{A}_K\|_\infty)^T, \quad (7)$$

where $\mathbf{A}_k$ denotes the kth row of $\mathbf{A}$, and $\|\cdot\|_\infty$ is the infinite norm. To simplify the expression, the max-pooling operation is represented as $\mathbf{p}_i^{(1)} = maxpool(\mathbf{A})$.

Let $n^{(1)}$ be the number of blocks for training. Then, the representations of the first pooling layer are defined as $\mathbf{P}^{(1)} = (\mathbf{p}_1^{(1)}, \mathbf{p}_2^{(1)}, \ldots, \mathbf{p}_{n^{(1)}}^{(1)})$. Regarding $\mathbf{P}^{(1)}$ as the input, another SAE can be trained. The activations of the second sparse AE are defined as

$$\mathbf{A}^{(2)} = (\mathbf{a}_1^{(2)}, \mathbf{a}_2^{(2)}, \ldots, \mathbf{a}_{n^{(1)}}^{(2)}),$$

where $\mathbf{a}_i^{(2)} = \sigma(\mathbf{W}^{(2)T}\mathbf{p}_i^{(1)} + \mathbf{b}^{(2)})$ is the hidden activation of the ith block. Here, $\mathbf{W}^{(2)}$ and $\mathbf{b}^{(2)}$ are the encoding weights and bias of the second SAE, respectively. Similarly to the patch-level structure, the columns in $\mathbf{A}^{(2)}$ can be regarded as the block-level features.

3.2.3. Image-level structure

By quantifying the block-level activations using Eq. (7) within each image, we obtain a set of image-level representations $\mathbf{P}^{(2)} = (\mathbf{p}_1^{(2)}, \mathbf{p}_2^{(2)}, \ldots, \mathbf{p}_{n^{(2)}}^{(2)})$, where $n^{(2)}$ is the count of the images used for pre-training.

Finally, the third SAE is trained to generate the image-level activations

$$\mathbf{A}^{(3)} = (\mathbf{a}_1^{(3)}, \mathbf{a}_2^{(3)}, \ldots, \mathbf{a}_{n^{(2)}}^{(3)}),$$

where $\mathbf{a}_i^{(3)} = \sigma(\mathbf{W}^{(3)T}\mathbf{p}_i^{(2)} + \mathbf{b}^{(3)})$ is the image-level feature for the ith histopathological image.

3.3. Fine-tuning

Following pre-training, the three basic structures are stacked to construct the integrated network, which can extract image-level features of the histopathological image. Since the weights and biases in the three structures are trained separately in greedy manner, they are not globally optimal for the stacked neural network. Therefore, a softmax layer is connected to the final hidden layer of the network, utilizing the supervised information to further optimize the network parameters. Then, the whole network is fine-tuned with the labeled histopathological images using the L-BFGS [48] algorithm. For the end-to-end supervised learning stage, more discriminative patterns for the descriptions of the histopathological images can be mined.

3.4. Encoding

Given a histopathological image, the nuclei are first detected, and then the patches of the nuclei are fed into the fine-tuned network to obtain the image-level features. Algorithm 1 presents the pseudo-code of the encoding. The proposed features are output from the last hidden layer (before the softmax layer). By putting these features into a classifier, the image can be classified.

Algorithm 1: Steps of encoding.

Data: the nucleus patches data $\mathbf{X}$ in the image
Result: the image-level feature $\mathbf{a}^{(3)}$
$\mathbf{W}^{(1)}, \mathbf{b}^{(1)} \leftarrow$ weights and bias of the patch-level structure;
$\mathbf{W}^{(2)}, \mathbf{b}^{(2)} \leftarrow$ weights and bias of the block-level structure;
$\mathbf{W}^{(3)}, \mathbf{b}^{(3)} \leftarrow$ weights and bias of the image-level structure;
$B \leftarrow$ number of blocks in the image;
for $i = 1$ *to* B **do**
    $N \leftarrow$ number of nucleus-patches in the ith block;
    for $j = 1$ *to* N **do**
        $\mathbf{x}_j \leftarrow$ the pixel data of the jth nucleus-patch in the ith block;
        $\mathbf{a}_j^{(1)} = \sigma(\mathbf{W}^{(1)T}\mathbf{x}_j + \mathbf{b}^{(1)})$;
    end
    $\mathbf{A}^{(1)} = (\mathbf{a}_1^{(1)}, \mathbf{a}_2^{(1)}, \ldots, \mathbf{a}_N^{(1)})$;
    $\mathbf{p}_i^{(1)} = maxpool(\mathbf{A}^{(1)})$;
    $\mathbf{a}_i^{(2)} = \sigma(\mathbf{W}^{(2)T}\mathbf{p}_i^{(1)} + \mathbf{b}^{(2)})$;
end
$\mathbf{A}^{(2)} = (\mathbf{a}_1^{(2)}, \mathbf{a}_2^{(2)}, \ldots, \mathbf{a}_B^{(2)})$;
$\mathbf{p}^{(2)} = maxpool(\mathbf{A}^{(2)})$;
$\mathbf{a}^{(3)} = \sigma(\mathbf{W}^{(3)T}\mathbf{p}^{(2)} + \mathbf{b}^{(3)})$;
return $\mathbf{a}^{(3)}$

Table 1

Details of the database. The first column gives the full names of breast lesions, the second column gives their abbreviations, the third column shows the labels of benign (B) or malignant (M) breast lesions, and the final column shows the numbers of the images.

Name	Abbr.	B\M	Number
invasive carcinoma of no special type	ICNST	M	300
invasive lobular carcinoma	ILC	M	300
tubular carcinoma	TC	M	274
carcinoma with medullary features	CMF	M	300
carcinoma with apocrine differentiation	CAD	M	300
glycogen-rich clear cell carcinoma	GRCC	M	300
ductal carcinoma *in situ*	DCIS	M	300
intraductal papilloma	IP	B	300
intraductal papillary carcinoma	IPC	M	300
lobular carcinoma *in situ*	LCIS	M	300
fibroadenoma	F.	B	300
phyllodes tumor	PT	B	300
hamartoma	H.	B	300
nonneoplastic lesion	NL	–	300
healthy tissue	HT	–	296
total	–	–	4470

4. Experiment

In this paper, a novel nucleus-guided feature extraction framework based on CNN is proposed for histopathological images. The proposed algorithm is implemented in Matlab 2013a on the PC with a 12 Intel Core Processor (2.10 GHz) and a GPU of Nvidia Tesla k40. The implementation of the whole network is based on the UFLDL tutorial.[1]

The performance of the proposed method is evaluated using a fine-annotated histopathological image database of breast lesions. Experiments are conducted to determine the parameters for the proposed neural network. And several state-of-the-art frameworks [27,31,33,37] for size-scalable histopathological image analysis are compared.

4.1. Experimental setting

The experimental images used in this study are supplied by Motic (Xiamen) Medical Diagnostic Systems Co. Ltd. There are 715 HE-stained whole slide images (WSIs) of breast lesions under a 20× lens (the spatial resolution is 1.2 μm/pixel), where 175 WSIs have been annotated by pathologists and the remainder 540 WSIs are not annotated. To ensure the reliability, the 175 WSIs have been independently annotated by two pathologists, and the final annotations are judged by a third pathologist. Using the 175 WSIs, two datasets of labeled images are established:

- The 2-class dataset: 2013 and 2096 labeled images with 256K–1536K pixels are respectively sampled from the annotated regions of malignant and benign breast tumors, generating a 2-class dataset.
- The 15-class dataset: Images in the 2-class dataset are fine-classified into 13 sub-categories of breast tumors according to the world health organization (WHO) standard [10]. In addition, images of the non-neoplastic lesions and healthy tissue are sampled as two further categories. Consequently, a 15-class dataset containing 4470 labeled images is established. The details of the 15-class dataset are listed in Table 1, and a representative image of each category is presented in Fig. 4.

For both the 2-class and 15-class datasets, a quarter of samples are used for testing, and the remainder for training.

For the 2-class dataset, three metrics including sensitivity, specificity, accuracy of classification are used to evaluate the algorithm performance, which are defined by equations:

$$Sensitivity = \frac{TP}{P}, Specificity = \frac{TN}{N}, Accuracy = \frac{TP+TN}{P+N}. \quad (8)$$

where *TP* denotes the number of correctly identified malignant samples, *TN* denotes the number of correctly identified benign samples, *P* and *N* are the number of malignant and benign samples, respectively. Similarly, for the 15-class dataset, the sensitivity and total classification accuracy are used. Let S_i denote the number of samples belonging to the *i*th class lesion and T_i be the number of correctly identified samples in the *i*th class lesion, the sensitivity for the *i*th class lesion is defined as T_i/S_i and the total accuracy is calculated by equation $\sum_{i=1}^{15} T_i / \sum_{i=1}^{15} S_i$.

4.2. Training details

In the pre-training stage, the samples used are unlabeled. Here, 14,352 images obtained through randomly sampling the 540 unlabeled WSIs are used. Millions of nucleus patches are extracted from these images using the nuclei detection method described in Section 3.1, in which one million nucleus patches are randomly selected to train the first SAE. Then, the quantified representation ($\mathbf{p}_i^{(1)}$) from all of the blocks in the 14,352 images are used to train the second SAE. The input for the third SAE consists of the pooling results ($\mathbf{p}_i^{(2)}$) of the 14,352 images.

In the fine-tuning stage, the labeled samples are used. For the 2-class and 15-class datasets, the number of the output nodes of the softmax layer are two and 15, respectively, corresponding to the labels in each dataset. To improve the robustness of the proposed neural network and prevent from over-fitting, the images are rotated and flipped to produce data augmentation in the fine-tuning stage.

4.3. Parameter setting

The size of the nucleus patch is set to 11 × 11, which can cover a nucleus and contain appropriate amount of stroma surrounding the nucleus. And three channels in RGB color space are used. Hence, the dimension of the nucleus data $\mathbf{x}_i$ is 11 × 11 × 3 = 363. Seven parameters including $K^{(1)}$, $K^{(2)}$, $K^{(3)}$, ρ, m, λ, and β need to be determined, where $K^{(1)}$, $K^{(2)}$ and $K^{(3)}$ are the number of hidden nodes in the three auto-encoders, ρ is the prospective sparsity mentioned in Section 3.2.1, m is the size of block mentioned in Section 3.2.2, and λ and β are the weights for the regularization term (Eq. (4)) and the sparsity term (Eq. (5)), respectively. When determining one of the seven parameters, the other parameters are set as constant.

$K^{(1)}$, $K^{(2)}$ and $K^{(3)}$ are determined in the pre-training stage with the 14,352 unlabeled images. According to [47], too few hidden nodes in SAE cannot cover all of the patterns appearing in the original data and result in a high reconstruction error (Eq. (6)). While, too many hidden nodes will result in redundancy, and increase the risk of over-fitting. In general, the reconstruction error decreases as the number of nodes increases. And when the number of nodes is enough to cover the patterns appearing in the data, the reconstruction error changes little as the node number increasing. Fig. 5 presents the reconstruction errors with various numbers of hidden nodes for the three auto-encoders. We determined $K^{(1)}$, $K^{(2)}$ and $K^{(3)}$ when there are three consecutive values whose corresponding reconstruction errors decrease less than 10^{-3}. Consequently, $K^{(1)}$, $K^{(2)}$ and $K^{(3)}$ are set to 300, 230 and 210.

The parameter ρ, m, λ, and β are chosen in a large range using a 10-fold cross validation [49] on the training set. A high classification accuracy indicates a good algorithm performance. Then,
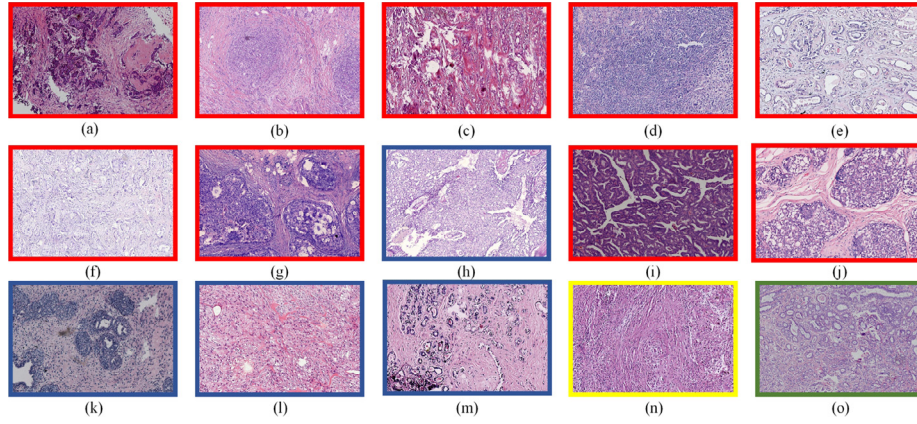
Fig. 4. Images in the database, where (a)-(g), (i) and (j) are malignant tumors, (h), (k), (l) and (m) are benign tumors. The names of the categories are (a) invasive carcinoma of no special type, (b) invasive lobular carcinoma, (c) tubular carcinoma, (d) carcinoma with medullary features, (e) carcinoma with apocrine differentiation, (f) glycogen-rich clear cell carcinoma, (g) ductal carcinoma in situ, (h) intraductal papilloma, (i) intraductal papillary carcinoma, (j) lobular carcinoma in situ, (k) fibroadenoma, (l) phyllodes tumor, (m) hamartoma, (n) nonneoplastic lesion, and (o) healthy tissue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
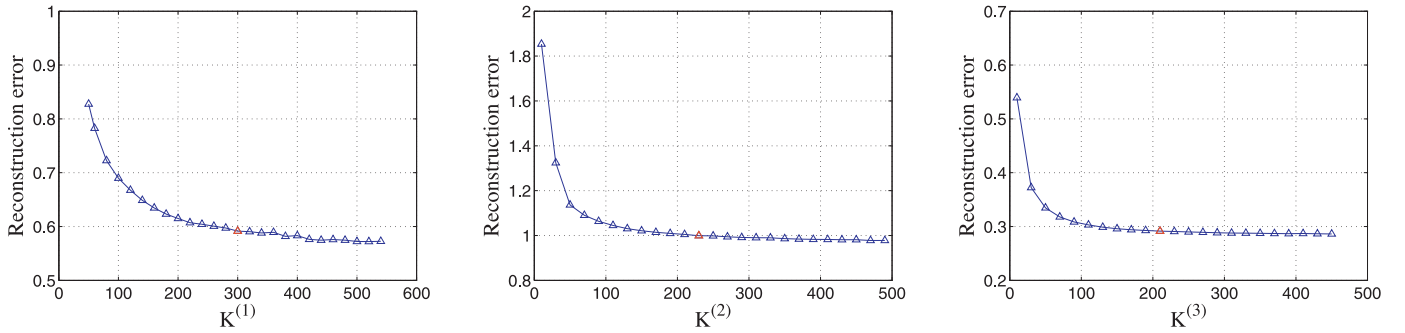
Fig. 5. Reconstruction errors with different numbers of hidden nodes in each auto-encoder, where the optimal number for each layer is indicated in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
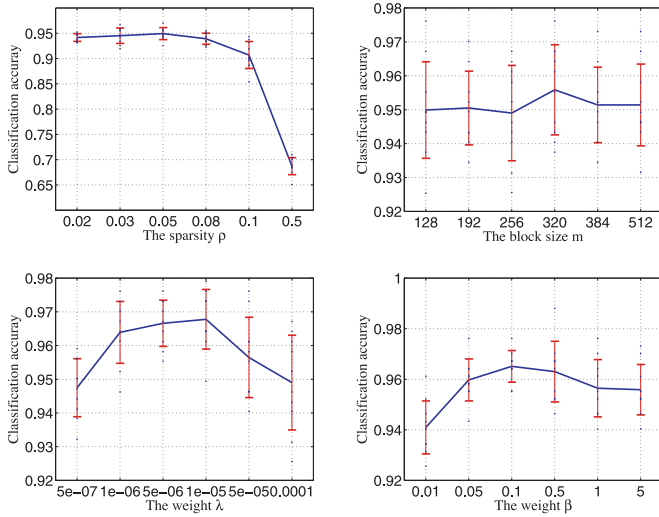
Fig. 6. Classification accuracy varied according to parameters for the 2-class dataset.

Table 2
The optimal parameters for the two datasets.

Dataset	ρ	m	λ	β
2-class	0.05	320	1×10^{-5}	0.1
15-class	0.1	256	1×10^{-6}	0.1

4.4. Verification of the designed network structure

The proposed histopathological image classification framework is abbreviated as N-CNN. To validate the effectiveness of the network structure of N-CNN, three degraded frameworks are compared:

- rand-CNN: The network structure is the same as N-CNN, but the patches used on the first layer are randomly selected.
- N-CNN$^-$: The block-level structure is omitted, and the nucleus activations are used to directly generate the input of the image-level structure.
- N-CNN^0: The pre-trained network of N-CNN. That is, the supervised information only impacts the softmax layer, and other hidden layers are not fine-tuned.

Tables 3 and 4 present the classification accuracies of N-CNN and the three degraded models on the testing sets, where all the features used for comparison are extracted with the optimal parameters determined in the training sets and classified using softmax classifiers.

As presented in Tables 3 and 4, classification accuracies obtained by N-CNN are much higher than those by rand-CNN.

the performance for different ρ, m, λ, and β are presented in Fig. 6 and the optimal parameters are listed in Table 2. As for the 15-class dataset, the same experimental protocol as on the 2-class dataset was adopted, and the final parameter settings are also given in Table 2.

Based on the optimal parameters determined above, our final network structure is shown in Fig. 7.
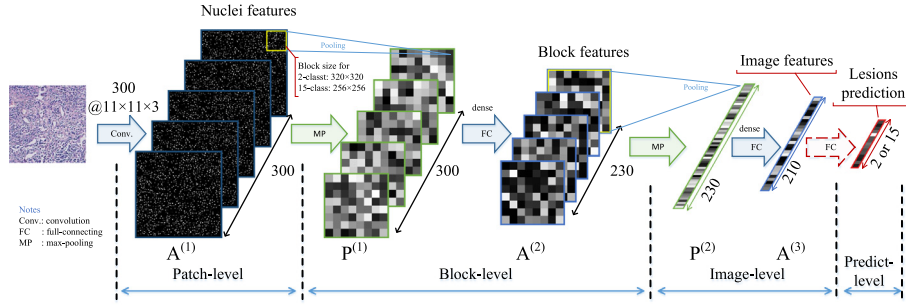
Fig. 7. The final structure of the proposed network.

Table 3
Classification performance of different network structures on the 2-class dataset.

| Method	Sensitivity	Specificity	Accuracy
rand-CNN	0.917	0.915	0.916
N-CNN^0	0.868	0.939	0.906
N-CNN$^-$	0.950	0.948	0.949
N-CNN	**0.955**	**0.964**	**0.959**

Table 4
Classification performance of different network structures on the 15-class dataset.

Method	rand-CNN	N-CNN^0	N-CNN$^-$	N-CNN
ICNST	0.986	0.986	0.986	**1.000**
ILC	0.733	0.600	0.813	**0.947**
IC	0.933	0.960	**0.973**	**0.973**
CMF	0.941	0.897	0.971	**1.000**
CAD	0.933	0.973	0.973	**1.000**
GRCC	0.840	0.787	0.907	**0.960**
DCIS	**1.000**	0.973	**1.000**	**1.000**
IP	0.747	0.720	0.800	**0.907**
IPC	0.813	0.707	0.947	**0.960**
LCIS	0.947	**0.973**	0.947	**0.973**
F.	0.867	0.827	0.893	**0.960**
PT	0.773	0.747	0.840	**0.933**
H.	0.680	0.747	0.787	**0.880**
NL	0.987	**1.000**	**1.000**	**1.000**
HT	0.853	0.813	0.880	**0.960**
Accuracy	0.869	0.847	0.914	**0.964**

(a) N-CNN
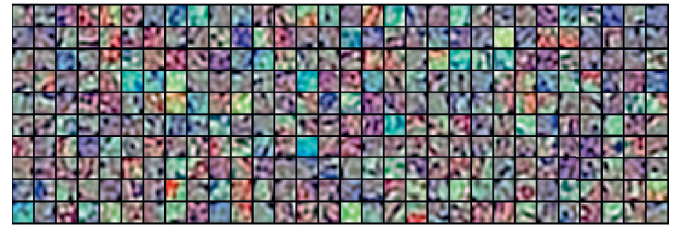
(b) rand-CNN

Fig. 8. Comparison of the neurons of N-CNN and rand-CNN in the patch-level structure.

Table 5
Classification accuracy for various number of layers between image-level representation and the softmax layer.

#Layer	1	2	3	4
2-class	0.959	0.955	0.957	0.544
15-class	0.964	0.965	0.955	0.843

Especially for 15-class dataset, the difference between them is 9.5%. Rand-CNN extracted features using random-sampled patches, which did not consider nucleus patterns. Relatively, N-CNN considered more nucleus patterns and filter the noise caused by stroma. Fig. 8 visualizes the difference in neurons between the N-CNN and rand-CNN. It can be seen that the neurons trained using N-CNN are more regular and can reflect the nuclei pattern such as shape and size better than the neurons trained with random-selected patches, therefore extracted histopathological images more effectively. Tables 3 and 4 also give the results of N-CNN$^-$ and N-CNN^0. N-CNN$^-$ obtained image-level features by directly pooling the patch-level features, which neglected the local distribution of the nuclei. N-CNN^0 did not utilize the supervised information to fine-tune the neurons of the three basic structures of the proposed network. In contrast, our model considered the distribution of nuclei and fully utilized the supervised information, achieving a better performance. Overall, comparing with the three degraded models, our designed model has the best classification accuracy, which means the structure of our designed model is better than the three degraded models.

Furthermore, more layers containing full-connecting and nonlinear units can be employed between the image-level representa-

tion and the softmax layer. In this paper, a single layer is adopted. The classification accuracy using different number of layers between the image-level representation layer and softmax layer are presented in Table 5. It can be seen that, the classification precision changes little when layers increase from one to three and dramatically decreases when the fourth layer is inserted. Therefore, a single layer is appropriate for the proposed network.

4.5. The robustness of the proposed feature

The classifier used in N-CNN is a softmax classifier. To evaluate the robustness of the proposed feature, another two general classifiers were validated, generating two further classification frameworks:

- N-CNN-SVM: The activations of the penultimate layer of N-CNN (the image-level features $\mathbf{A}^{(3)}$ mentioned in Section 3.2.3) are used as the input of a linear SVM.

Table 6
Classification performance using different classifiers on the 2-class dataset.

Method	Sensitivity	Specificity	Accuracy
N-CNN-kNN	**0.983**	0.950	**0.966**
N-CNN-SVM	0.956	0.941	0.958
N-CNN	0.955	**0.964**	0.959

Table 7
Classification performance using different classifiers on the 15-class dataset.

Method	N-CNN-kNN	N-CNN-SVM	N-CNN
ICNST	**1.000**	1.000	1.000
ILC	0.853	**0.947**	0.947
IC	**0.973**	0.973	0.973
CMF	0.985	**1.000**	1.000
CAD	0.987	0.987	**1.000**
GRCC	0.907	**0.960**	0.960
DCIS	**1.000**	1.000	1.000
IP	0.867	0.867	**0.907**
IPC	0.973	**1.000**	0.960
LCIS	**0.973**	0.973	0.973
F.	0.973	**0.987**	0.960
PT	0.907	0.893	**0.933**
H.	0.787	0.880	**0.880**
NL	**1.000**	1.000	1.000
HT	**0.973**	0.973	0.960
Accuracy	0.944	0.963	**0.964**

Table 8
Classification performance of different frameworks on the 2-class dataset.

| Method	Sensitivity	Specificity	Accuracy
RBM [33]	0.746	0.708	0.726
miGraph [27]	0.862	0.893	0.879
DFDL [37]	0.904	0.917	0.911
CNN-BOF [31]	**0.973**	0.935	0.953
N-CNN	0.955	**0.964**	**0.959**

- N-CNN-KNN: The activations of the penultimate layer of N-CNN are used to generate a KNN classifier.

Tables 6 and 7 show the classification performance using different classifiers on the two experimental datasets. It can be seen that N-CNN-kNN and N-CNN-SVM achieve comparable performances with N-CNN, and all of the three have high classification accuracies. It indicates that the image-level features extracted by the proposed network are robust and appropriate for describing histopathological images. In addition, N-CNN-kNN achieved the highest accuracy on the 2-class dataset and 0.7% more than N-CNN. On the 15-class dataset, N-CNN obtained the best accuracy and 2% higher than N-CNN-kNN. Therefore, N-CNN is better and selected as the final classification framework.

4.6. Comparison of different classification frameworks

Fig. 9 shows the receiver operating characteristic (ROC) curves and the corresponding area under curve (AUC) values for different classification frameworks. The closer to the top-left corner the ROC curve, the better the method, and the higher the AUC value, the better the algorithm performance. It can be seen that our method achieved the best performance with an AUC of 0.863.

Tables 8 and 9 give the classification results for the 2-class and 15-class datasets respectively. Obviously, our method has the highest classification accuracy with 95.9% on the 2-class dataset and 96.4% on the 15-class dataset. miGraph [27] is a classification framework based on statistics-based features, which are not
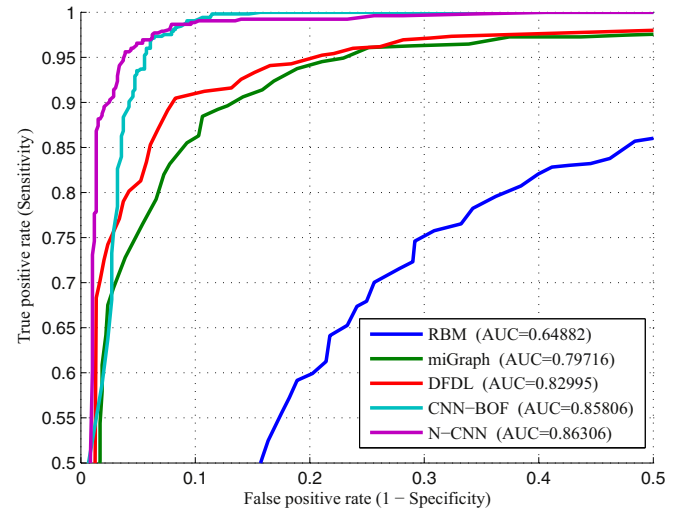
Fig. 9. Classification performance of different network structures on the 2-class dataset.

discriminative enough to represent complex histopathological images. RBM [33] and CNN-BOF [31] learned features from the entire image and did not consider the characteristics of histopathological images. DFDL [37] classified the image by quantifying the prediction results of the key patches for an image, which ignored the interrelation and the spatial information of these patches. Relatively, N-CNN generates the image-level prediction step-by-step from the patch-level, block-level, and image-level features through an integrated network, which considered both patterns and distribution of the key patches, and therefore obtained better classification performance.

The computational complexity is also important for classification methods. Table 10 shows the average time costs of different methods on the 15-class dataset. Because the majority of the cost is from the feature extraction, thus the time consume on the 2-class dataset is almost the same with that on the 15-class dataset. Therefore, the time consume on the 2-class dataset is not given here. From Table 10, it can be seen that, RBM [33] is the fastest with 0.12 s among the five compared methods, but it has the lowest classification accuracy (see Tables 8 and 9). CNN-BOF [31] encodes all of the feasible patches within the image, which requires a computational time of 7.82 s. Compared with CNN-BOF, N-CNN takes a short time to search the key patches from all the feasible patches to encode, thus sharply reduces the computation.

Therefore, with the best classification accuracy and the second short time cost, our proposed method greatly outperforms other compared methods.

5. Discussion

In this paper, the nucleus features including the pattern and spatial distribution are extracted from histopathological images using a designed CNN combined with the guide of nuclei, which has good classification performance for breast lesions. Regarding the proposed method, we have the following discussion:

5.1. Differences between normal CNNs and the proposed network

The first layer of the proposed network is a convolutional layer. In this paper, we extracted the first layer features (patch-level features) through a full connection operation on the nucleus-centered patches. In fact, we can also calculate the feature maps through convolution operations between the image and the weights matrices, and zero non-nucleus-centered positions of the feature maps

Table 9
Classification performance of different frameworks on the 15-class dataset.

Method	RBM [33]	miGraph [27]	DFDL [37]	CNN-BOF [31]	N-CNN
ICNST	0.878	0.986	0.986	**1.000**	**1.000**
ILC	0.480	0.733	0.760	0.720	**0.947**
IC	0.853	0.933	0.960	0.947	**0.973**
CMF	0.721	0.941	0.941	0.926	**1.000**
CAD	0.880	0.933	0.973	0.973	**1.000**
GRCC	0.693	0.840	**0.960**	0.893	**0.960**
DCIS	**1.000**	**1.000**	**1.000**	0.987	**1.000**
IP	0.573	0.747	**0.907**	0.880	**0.907**
IPC	0.520	0.813	0.840	0.947	**0.960**
LCIS	0.880	0.947	0.960	0.960	**0.973**
F.	0.693	0.867	0.933	**0.987**	0.960
PT	0.440	0.773	0.787	0.920	**0.933**
H.	0.520	0.680	0.800	0.867	**0.880**
NL	0.840	0.987	0.987	**1.000**	**1.000**
HT	0.733	0.853	0.800	**0.960**	**0.960**
Accuracy	0.714	0.869	0.906	0.931	**0.964**

Table 10
Computational time for different classification frameworks (s).

RBM [33]	miGraph [27]	DFDL [37]	CNN-BOF [31]	N-CNN
0.12	11.04	1.1	7.82	0.41

to obtain the patch-level features. The full-connecting operation is a special form of the convolutional operation. Through this simplification, the final calculation in the designed network no longer needs convolutional operation and the computational complexity is greatly reduced.

CNN consists of several convolutional/non-linear layers and can end-to-end extract high-level features from images, which has been widely used in many applications. Our network is designed based on CNN. Compared with a normal CNN, the proposed network has three significant differences: Firstly, their structures are different. A normal CNN usually contains 5 convolutional layers and 3 full-connecting layers, or contains even more layers. The neurons in bottom layers generally activates on the basic elements in images, such as edges and lines. While our proposed network is designed based on pathological characteristics, and consists of the patch-level, block-level and image-level structures, which is simpler than a normal CNN. The neurons in the first layer of the designed network directly characters the nuclei, which makes the network more efficient to represent histopathological images. Secondly, the magnitude order of parameters is different. A normal CNN, e.g. AlexNet [50], has about 6×10^7 parameters, and training it usually needs more than 10^5 interactions when randomly initializing the parameters. Our proposed network only has about 2×10^5 parameters, and in our experiment, it can converge in about 1500 interactions (including the pre-training and the fine-tuning stages). Thirdly, the size of input images is different. The input size in normal CNNs is fixed, and images in different size need to be resized before inputting the network. While our network has no strict limitation to the size of input images, which is more convenient to clinical applications.

5.2. Application in WSI analysis

The proposed framework can be used for the WSI automatic analysis. A WSI has very large size. In the computer-aided diagnoses system, the analysis for a WSI follows the sliding-window paradigm. In implementation, we first divided the entire WSI into non-overlapping blocks, and then detected the nuclei and extracted block-level features throughout the WSI. Next, for each sliding window, the block-level features of all the blocks in it were pooled to extract image-level feature and the lesion was predicted. Fig. 10 is a 2-class (benign and malignant tumor) prediction instance, where Fig. 10(a) is a WSI with a size 14K × 15K pixels under a 20 × lens. We use a 1536 × 1024 sliding-window (the step length is 320 pixel) to process the WSI and predict the probability of malignant tumor for each region. Fig. 10(b) is the prediction result, where red denotes that the probability of the malignancy is 1, and blue denotes the probability is 0. The probability gradually increases from blue to red. Obviously, four regions are predicted to have high malignant probability in the WSI. Fig. 10(d) is the display for the region with high malignant probability under high resolution, where the blue line is the annotation of invasive lobular carcinoma from pathologists. Thus, using our prediction result, pathologists can fast find the informative regions from a WSI and make more accurate diagnosis.

Because the nucleus detection and block-level feature extraction are only calculated one time for the entire WSI, the redundant computation in the overlapping regions between sliding windows can be avoided. In addition, the proposed framework can automatically control the calculations for different sub-regions, because it is established on a nucleus detection method and only nucleus patches are needed to be encoded. Particularly, for uninformative regions (containing few nuclei or large blank region), the framework can recognize them by a threshold of nucleus number and avoid unnecessary calculations. In total, the processing of the WSI in Fig. 10 costs 47.9 s in Matlab, which is efficient.

Therefore, the proposed framework can aid pathologists to find the interesting region and make diagnosis in a short time, which is applicable.

6. Conclusion

The appearance and spatial distribution of cell nuclei are significant indicators for cancer diagnosis, according to which a nucleus-guided feature extraction framework based on a convolutional neural network has been proposed for histopathological image representation and classification in this paper. The nuclei are first detected from the histopathological image, and then used to guide the training of the convolutional neural network. With the nucleus-guided strategy, the network paid more attention to the difference in nucleus appearance and effectively reduced the noise and redundancy caused by stroma. Finally, discriminative features preserving both the patterns and distribution information of nuclei for histopathological images are extracted step-by-step from the patch-level, block-level and image-level structure of the integrated network. Our framework is evaluated on a breast-lesion database. The experiment results have certified that the proposed features
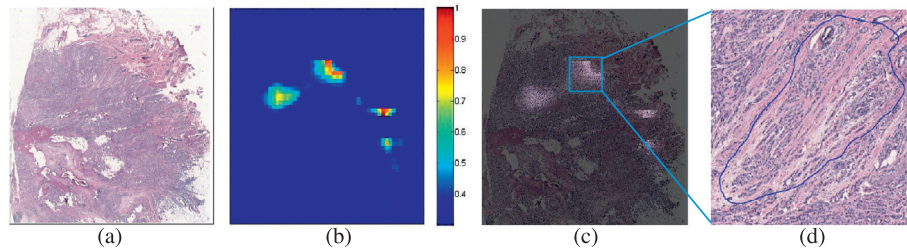
Fig. 10. WSI analysis instance using our algorithm framework, where (a) is a WSI, (b) is the probability map for malignant tumor, (c) is the fusion display of (a) & (b), and (d) is a key sub-region containing invasive lobular carcinoma.

are effective to represent histopathological images. And, the proposed framework achieved the best classification performance in classification of breast lesions compared with other state-of-the-art methods. Further work will focus on clinical applications employing the proposed features, such as whole slide image analysis and content-based histopathological image retrieval.

Acknowledgment

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, CA 67 (1) (2017) online.

[2] S. Zhang, D. Metaxas, Large-scale medical image analytics: recent methodologies, applications and future directions, Med. Image Anal. 33 (2016) 98–101.

[3] J. Wu, X. Sun, J. Wang, Y. Cui, F. Kato, H. Shirato, D.M. Ikeda, R. Li, Identifying relations between imaging phenotypes and molecular subtypes of breast cancer: model discovery and external validation, J. Magn. Reson. Imaging (2017), doi:10.1158/1078-0432.CCR-16-2415.

[4] J. Wu, Y. Cui, X. Sun, G. Cao, B. Li, D.M. Ikeda, A.W. Kurian, R. Li, Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways, Clin. Cancer Res. (2017), doi:10.1002/jmri.25661.

[5] S.L. Robbins, V. Kumar, A.K. Abbas, N. Fausto, J.C. Aster, Robbins and Cotran Pathologic Basis of Disease, Saunders/Elsevier, 2010.

[6] C. Mosquera-Lopez, S. Agaian, A. Velez-Hoyos, I. Thompson, Computer-aided prostate cancer diagnosis from digitized histopathology: a review on texture-based systems, IEEE Rev. Biomed. Eng. 8 (2014) 98–113, doi:10.1109/RBME.2014.2340401.

[7] J.S. Duncan, N. Ayache, Medical image analysis: progress over two decades and the challenges ahead, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 85–106, doi:10.1109/34.824822.

[8] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: a review, IEEE Rev. Biomed. Eng. 2 (2009) 147–171.

[9] B.E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, J.A.W.M.V. Der Laak, Automated detection of dcis in whole-slide h&e stained breast histopathology images, IEEE Trans. Med. Imaging 35 (9) (2016) 2141–2150.

[10] S.R. Lakhani, I.A. for Research on Cancer, W.H. Organization, et al., WHO Classification of Tumours of the Breast, International Agency for Research on Cancer, 2012.

[11] J.C. Caicedo, A. Cruz, F.A. Gonzalez, Histopathology image classification using bag of features and kernel functions, in: Artificial Intelligence in Medicine in Europe, Springer, 2009, pp. 126–135, doi:10.1007/978-3-642-02976-9_17.

[12] O. Sertel, J. Kong, H. Shimada, U. Catalyurek, J.H. Saltz, M.N. Gurcan, Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development, Pattern Recognit. 42 (6) (2009) 1093–1103, doi:10.1016/j.patcog.2008.08.027.

[13] E. Mercan, S. Aksoy, L.G. Shapiro, D.L. Weaver, T. Brunye, J.G. Elmore, Localization of diagnostically relevant regions of interest in whole slide images, in: International Conference on Pattern Recognition, 2014, pp. 1179–1184, doi:10.1109/ICPR.2014.212.

[14] Y. Xu, J.-Y. Zhu, I. Eric, C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Med. Image Anal. 18 (3) (2014) 591–604, doi:10.1016/j.media.2014.01.010.

[15] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (60) (2004) 91–110.

[16] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[17] T. Ojala, I. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognit. 29 (1) (1996) 51–59.

[18] M.M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, M.N. Gurcan, Computerized classification of intraductal breast lesions using histopathological images, IEEE Trans. Biomed. Eng. 58 (7) (2011) 1977–1984.

[19] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak, Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images, Comput. Biol. Med. 43 (10) (2013) 1563–1572, doi:10.1016/j.compbiomed.2013.08.003.

[20] L. Cheng, M. Mandal, Automated analysis and diagnosis of skin melanoma on whole slide histopathological images, Pattern Recognit. 48 (8) (2015) 2738–2750, doi:10.1016/j.patcog.2015.02.023.

[21] S. Doyle, A. Agner, A. Madabhushi, M. Feldman, J. Tomaszewski, Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features, in: IEEE International Symposium on Biomedical Imaging, IEEE, 2008, pp. 496–499.

[22] A.N. Basavanhally, S. Ganesan, S. Agner, J.P. Monaco, M.D. Feldman, J.E. Tomaszewski, G. Bhanot, A. Madabhushi, Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology, IEEE Trans. Biomed. Eng. 57 (3) (2010) 642–653, doi:10.1109/TBME.2009.2035305.

[23] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V.P. Kumar, D. Verbel, A. Kotsianti, O. Saidi, Multifeature prostate cancer diagnosis and gleason grading of histological images, IEEE Trans. Med. Imaging 26 (10) (2007) 1366–1378.

[24] O.S. Al-Kadi, Texture measures combination for improved meningioma classification of histopathological images, Pattern Recognit. 43 (6) (2010) 2043–2053, doi:10.1016/j.patcog.2010.01.005.

[25] E. Ozdemir, C. Gunduz-Demir, A hybrid classification model for digital pathology using structural and statistical pattern recognition, IEEE Trans. Med. Imaging 32 (2) (2013) 474–483.

[26] Y. Zheng, Z. Jiang, J. Shi, Y. Ma, Retrieval of pathology image for breast cancer using plsa model based on texture and pathological features, in: IEEE International Conference on Image Processing, 2014, pp. 2304–2308, doi:10.1109/ICIP.2014.7025467.

[27] M. Kandemir, F.A. Hamprecht, Computer-aided diagnosis from weak supervision: a benchmarking study, Comput. Med. Imaging Graph. 42 (2015) 44–50, doi:10.1016/j.compmedimag.2014.11.010.

[28] X. Zhang, H. Dou, T. Ju, J. Xu, S. Zhang, Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis, IEEE J. Biomed. Health Inform. 20 (5) (2015) 1377–1383.

[29] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, Y. Zhao, Breast histopathological image retrieval based on latent Dirichlet allocation, IEEE J. Biomed. Health Inform. (2016), doi:10.1109/JBHI.2016.2611615.

[30] A.A. Cruz-Roa, J.E.A. Ovalle, A. Madabhushi, F.A.G. Osorio, A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 403–410.

[31] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, F.A. González, An unsupervised feature learning framework for basal cell carcinoma image analysis, Artif. Intell. Med. 64 (2) (2015) 131–145.

[32] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2016) 119–130, doi:10.1109/TMI.2015.2458702.

[33] N. Nayak, H. Chang, A. Borowsky, P. Spellman, B. Parvin, Classification of tumor histopathology via sparse feature learning, in: International Symposium on Biomedical Imaging, IEEE, 2013, pp. 410–413, doi:10.1109/ISBI.2013.6556782.

[34] H. Chang, N. Nayak, P.T. Spellman, B. Parvin, Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 91–98.

[35] Y. Zhou, H. Chang, K. Barner, P. Spellman, B. Parvin, Classification of histology sections via multispectral convolutional sparse coding, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, pp. 3081–3088.

[36] U. Srinivas, H.S. Mousavi, V. Monga, A. Hattel, B. Jayarao, Simultaneous sparsity model for histopathological image representation and classification, Med. Imaging IEEE Trans. 33 (5) (2014) 1163–1179.

[37] T. Vu, H. Mousavi, V. Monga, G. Rao, A. Rao, Histopathological image classification using discriminative feature-oriented dictionary learning, Med. Imaging IEEE Trans. 35 (3) (2016) 738–751.

[38] C. Malon, M. Miller, H.C. Burger, E. Cosatto, H.P. Graf, Identifying histological elements with convolutional neural networks, in: Proceedings of the 5th International Conference on Soft Computing as Transdisciplinary Science and Technology, ACM, 2008, pp. 450–456.

[39] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.

[40] L. Hou, D. Samaras, T.M. Kurc, Y. Gao, J.E. Davis, J.H. Saltz, Patch-based convolutional neural network for whole slide tissue image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2424–2433.

[41] X. Zhang, H. Su, L. Yang, S. Zhang, Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015a, pp. 5361–5368, doi:10.1109/CVPR.2015.7299174.

[42] X. Zhang, F. Xing, H. Su, L. Yang, S. Zhang, High-throughput histopathological image analysis via robust cell segmentation and hashing, Med. Image Anal. 26 (1) (2015b) 306–315.

[43] D. Zink, A.H. Fischer, J.A. Nickerson, Nuclear structure in cancer cells, Nat. Rev. Cancer 4 (9) (2004) 677–687.

[44] A.C. Ruifrok, D.A. Johnston, Quantification of histochemical staining by color deconvolution, Anal. Quant. Cytol. Histol. 23 (4) (2001) 291–299.

[45] A. Coates, A.Y. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning, in: International Conference on Artificial Intelligence and Statistics, 2011, pp. 215–223.

[46] J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, Stacked convolutional autoencoders for hierarchical feature extraction, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 52–59, doi:10.1007/978-3-642-21735-7_7.

[47] A. Ng, Sparse autoencoder, CS294A Lect. notes 72 (2011) 1–19.

[48] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. 16 (5) (1995) 1190–1208.

[49] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai, 1995, pp. 1137–1145.

[50] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.

Yushan Zheng received the B.E. and M.S. degree from Beihang University in 2012 and 2015 respectively. He is currently a Ph.D. candidate majored in Pattern Recognition and Intelligent System at Beihang University. His research interests include medical image processing, analysis, classification and retrieval.

Zhiguo Jiang received the B.E., M.S., and Ph.D. degrees from the Beihang University in 1987, 1990, and 2005, respectively. He is currently a professor in Image Processing Center, School of Astronautics, Beihang University. His research interests include medical image processing and classification, remotely sensed image processing, target detection and recognition.

Fengying Xie received the Ph.D. Degree in Pattern Recognition and Intelligent System from Beihang University in 2009. She is now a professor in Image Processing Center, School of Astronautics, Beihang University. Her research interests include biomedical image and remote sensing image processing, image quality assessment, image segmentation and classification.

Haopeng Zhang received his B.S. and Ph.D. degrees from Beihang University in 2008 and 2014, respectively, and is currently with the Image Processing Center, School of Astronautics, Beihang University. His main research interests are multi-view object recognition, 3D object recognition and pose estimation, etc.

Yibing Ma received the B.E. degree from Beihang University in 2011. He is currently a Ph.D. candidate majored in Pattern Recognition and Intelligent System at Beihang University. His research interests include medical image processing, analysis, classification and retrieval.

Huaqiang Shi is a deputy chief physician of the General Hospital of the Air Force of PLA. He is also the medical advisor of Motic (Xiamen) Medical Diagnostic Systems Co. Ltd. His research interests are diagnosis of breast cancer and the clinical application of image analysis technology.

Yu Zhao is the Software R&D Manager of Motic (Xiamen) Medical Diagnostic Systems Co. Ltd. His research interests are clinical application of image analysis technology, telemedicine system and content-based medical image retrieval.