Histopathological whole slide image analysis using context-based CBIR

Yushan Zheng, Zhiguo Jiang, *Member, IEEE,* Haopeng Zhang*, *Member, IEEE,* Fengying Xie, Yibing Ma, Huaqiang Shi and Yu Zhao

***Abstract*—Histopathological image classification (HIC) and content-based histopathological image retrieval (CBHIR) are two promising applications for histopathological whole slide image (WSI) analysis. HIC can efficiently predict the type of lesion involved in a histopathological image. In general, HIC can aid pathologists in locating high-risk cancer regions from a WSI by providing a cancerous probability map for the WSI. In contrast, CBHIR was developed to allow searches for regions with similar content for a region of interest (ROI) from a database consisting of historical cases. Sets of cases with similar content are accessible to pathologists, which can provide more valuable references for diagnosis. A drawback of the recent CBHIR framework is that a query ROI needs to be manually selected from a WSI. An automatic CBHIR approach for a WSI-wise analysis needs to be developed. In this paper, we propose a novel aided-diagnosis framework of breast cancer using whole slide images, which shares the advantages of both HIC and CBHIR. In our framework, CBHIR is automatically processed throughout the WSI, based on which a probability map regarding the malignancy of breast tumors is calculated. Through the probability map, the malignant regions in WSIs can be easily recognized. Furthermore, the retrieval results corresponding to each sub-region of the WSIs are recorded during the automatic analysis and are available to pathologists during their diagnosis. Our method was validated on fully annotated WSI datasets of breast tumors. The experimental results certify the effectiveness of the proposed method.**

***Index Terms*—Whole slide image analysis, CBIR, breast cancer, contextual information**

I. Introduction

With the development of digital pathology, histological sections can be scanned by pathologists using micro-scanners during their rest time and stored as digital whole slide images (WSIs). The time between scanning and diagnosis is a valuable resource for computer-aided diagnosis (CAD). After or during the scanning period, the WSIs can be analyzed using a reliable artificial intelligent algorithm, which can promote the diagnostic accuracy and relieve the workload of the pathologists. Across various applications of histopathological image analysis, image classification [1] and content-based image retrieval (CBIR) [2] are important challenges.

Histopathological image classification (HIC) aims to directly predict different types of lesions (for example, the classification of benign and malignant tumors). In recent years, many new approaches to histopathological image classification [3], [1] have been developed. Recently, Srinivas et al. and Vu et al. [4], [5] proposed using a sparsity model to encode cellular patches, and classified histopathological images by fusing the predictions of these cellular patches. In [6], Zhang et al. applied a hashing-based method to identify individual cells in lung images, and then classified the images into two subtypes of lung cancer. With the development of computer science and digital pathology, an increasing number of studies are being conducted to analyze digital whole slide images (WSIs). Xu et al. [7] and Kandemir et al. [8] proposed dividing WSIs into square blocks, and classifying WSIs using classical visual features combined with multi-instance learning models. Mercan et al. [9] utilized the scanning information of pathologists in clinical diagnosis to localize the diagnostically relevant regions of interest (ROIs) in an unknown WSI. Recent years, deep-learning models, including convolutional neural networks (CNN) [10], [11], and auto-encoders [12], [13] have been introduced into histopathological WSI analysis, yielding a more effective CAD performance. These methods divide a WSI into square blocks and segment the WSI by classifying each block. However, the appearance of meaningful objects in a WSI is diverse. Dividing a WSI using square blocks does not characterize the objects. More recently, Bejnordi et al. [14] presented an automated ductal carcinoma in situ (DCIS) detection framework for breast cancer, in which the WSI is first divided into irregular regions and then classified as DCIS or benign/normal tissue. This method has been demonstrated to be directly applicable to diagnosis using a WSI.

Differing from image classification that aims to directly recognize tumors, content-based histopathological image retrieval (CBHIR) [2], [15] searches for regions with similar content for the queried ROI from a database. In addition, for digital pathology platforms (e.g. MoticGallery[1]) assembling historical cases, the diagnoses of experts for cases are also accessible to users, which can provide more comprehensive information for pathologists. In particular, retrieval-based applications can help pathologists during training. Early research on CBHIR was focused on sub-images of WSIs [16], [17]. In

This work was supported by the National Natural Science Foundation of China (No. 61771031, 61371134, 61471016, and 61501009) and project of Motic-BUAA Image Technology Research and Development Center. *Asterisk indicates corresponding author.*

Yibing Ma, Zhiguo Jiang, Haopeng Zhang, Fengying Xie and Yushan Zheng are with Image Processing Center, School of Astronautics, Beihang University, Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, and Beijing Key Laboratory of Digital Media, Beijing, 100191, China. (e-mail: yszheng, jiangzg, zhanghaopeng, xfy_73, mayibing@buaa.edu.cn)

Huaqiang Shi and Yu Zhao are with Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen 361101, China. (e-mail: shihq, zhaoy@motic.com)

Huaqiang Shi is chief physician, department of pathology, Peoples Liberation Army Air Force General Hospital, Beijing 100142, China.

[1]http://med.motic.com/MoticGallery/[accessible 2018-01-16]

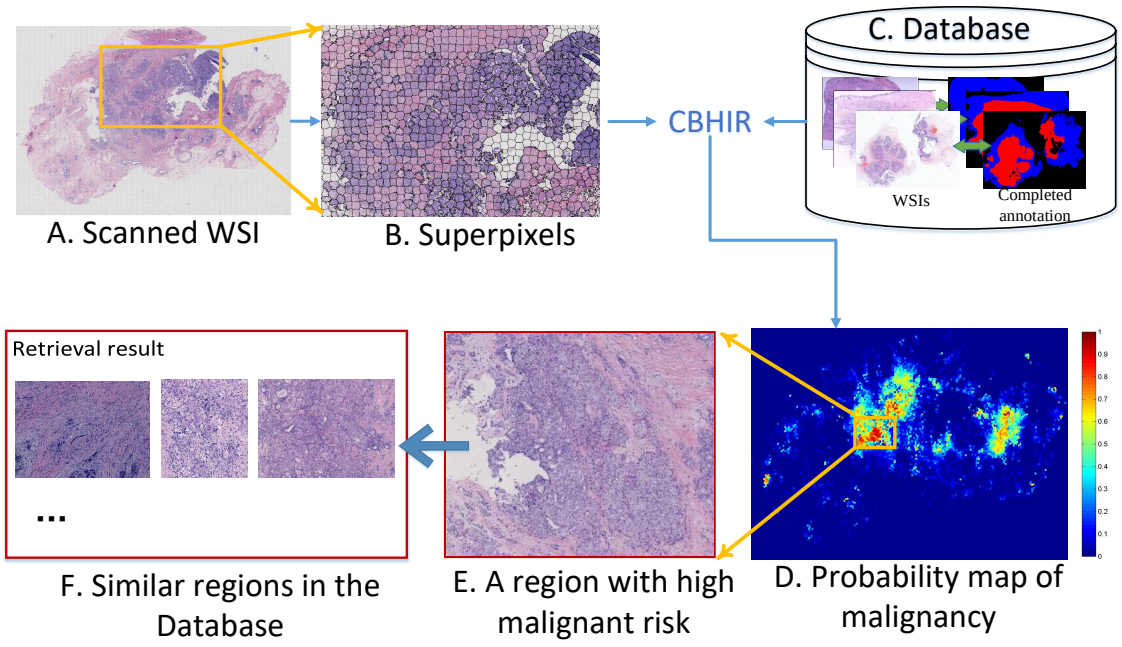


Fig. 1. Flowchart of the proposed WSI analysis framework, where A is a WSI, B shows the super-pixels in a part of this WSI, and C is a database that contains diagnosed WSIs and an annotation of malignant (displayed in red) and benign (displayed in blue) regions. Through the CBHIR technique, all super-pixels in an unknown WSI can be diagnosed and a probability map of the malignancy (D) is generated. In the probability map, red denotes that the probability of the malignancy is 1, blue denotes the probability is 0, and the probability gradually increases from blue to red. E presents a region with a high risk of cancer proposed by the framework. F shows several similar regions retrieved from the database in the CBHIR procedure, which are available to pathologists after the analysis.

particular, semantic analysis models [18], [19] were introduced in histopathological image analysis and have been proven effective for CBHIR [20], [21], [22], [23]. In recent years, Zhang et al. [24], [25] proposed the retrieval of images with similar content from a labeled database. By quantifying the labels of the returned images, the lesion of the query image can be predicted. This research provides an alternative approach to predict lesions from histopathological images. To achieve the retrieval from a database consisting of WSIs, Ma et al. [26] proposed storing the WSIs into the database following the sliding-window (SW) paradigm. During retrieval, the similarities between all windows in the database and the query image are calculated. This work provides a preliminary strategy to retrieval in a WSI-database; however, the SW paradigm causes a large amount of redundancy in computation. To solve this problem, more efficient WSI-retrieval methods [27], [28], [29] based on a selective search and multiple binary encoding have been proposed.

The majority of HIC approaches can be applied to a WSI analysis, in which the WSI is segmented into blocks or super-pixels and then predicted using the HIC algorithm. Furthermore, a probability map indicating relatively important regions [30], [31] can be generated by predicting all blocks or super-pixels in the WSI. Checking the probability map before diagnosis, pathologists can quickly locate the important regions. The diagnosis of a reliable HIC-based system can provide a verification of the diagnosis for pathologists. In general, HIC-based approaches only provide a prediction of lesions, but have difficulty providing the reason or reference of this prediction. In contrast, CBHIR can provide more information in aided-diagnosis by returning similar regions from a database consisting of diagnosed cases. The appearance and the historical diagnosis of these retrieved regions can be regarded as bases of the aided-diagnosis. Hence, CBHIR is more effective than HIC in improving the capability of a pathologist. Nevertheless, in the present studies [26], [2], the query ROI is provided by pathologists. This means the pathologists need to manually locate the key regions in a WSI and then use the CBHIR application to analyze these ROIs. It cannot relieve the workload of pathologists.

In this paper, we leverage the characteristics of HIC and CBHIR, and propose a novel WSI analysis framework for breast cancer diagnosis, which retains the advantages of both HIC and CBHIR. The analysis of a WSI is achieved using a designed CBHIR model. Differing from the research on WSI-database retrieval [27], [28], [29], the WSIs in the database were entirely annotated by pathologists. Hence, the status of sub-regions in an unknown WSI can be quantitatively analyzed using the strategy proposed in Zhang et al. [24]. Similar to an HIC-based system, a probability map is generated to aid pathologists in locating important ROIs in the WSI. Because the probability of each sub-region is obtained using the CBHIR framework, the regions with similar content for each sub-region are recorded during the automatic analysis and are available to pathologists whenever scanning a WSI. Different from the present CBHIR framework where the retrieval is processed online, the entire process of the proposed framework is offline. Therefore, no additional computation is required when pathologists intend to visit the retrieval results of each region in the WSI. To achieve a more reliable aided-diagnosis,

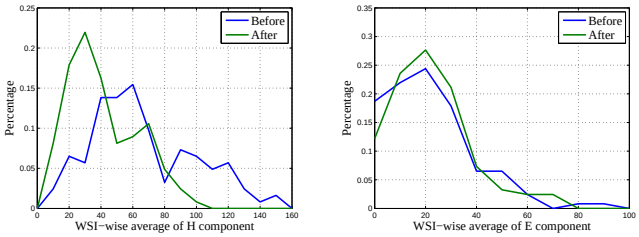


Fig. 2. Allocation for WSI-wise average of staining components before and after the normalization.

a novel feature extraction pipeline considering multiple magnifications of histopathological images is proposed based on deep neural networks. In addition, contextual information of multiple scales for a sub-region is considered in the feature extraction and retrieval stages because the allocation of objects surrounding a sub-region is significant for diagnosis. The proposed framework is evaluated through experiments of sub-region retrieval, classification, and WSI segmentation tasks on breast tumor databases. In addition, the results demonstrated that the proposed framework is effective in histopathological image analysis and is superior to the compared state-of-the-art methods developed for histopathological image classification and retrieval.

The remainder of this paper is organized as follows. Section II introduces the proposed method. The experiment is presented in Section III. Finally, Section IV summarizes the present contributions and suggests directions for future work.

II. Methodology

A. Overview

The main pipeline of the proposed method is shown in Fig. 1. Because the WSI is diagnosed through a CBHIR approach, a database of completely annotated WSIs needs to be established beforehand. Through super-pixel segmentation, feature extraction, and binarization steps, the malignant/non-malignant (including benign and normal) labels, and binary codes of super-pixels for a WSI are stored into the database. For an unknown WSI, it is first segmented into super-pixels and then encoded. By searching the similar regions for each super-pixel from the database, a probability map (Fig. 1D) regarding the malignancy of tumors is generated. In addition, the retrieval results (Fig. 1F) for the sub-regions at different scales are recorded during the retrieval, which are available after analysis.

B. Representation of WSIs

The encoding of WSIs is the basis of CBHIR. With the proposed method, a WSI is encoded through four steps: pre-processing, super-pixel segmentation, feature extraction, and binarization.

1) Pre-processing: The WSIs considered in this work are hematoxylin-eosin (H&E) stained sections, where hematoxylin mainly stains the nuclei, and eosin stains the cytoplasm/stroma. The color of HE-stained WSIs is a combination of the two stains, while the digital WSIs are generally stored in

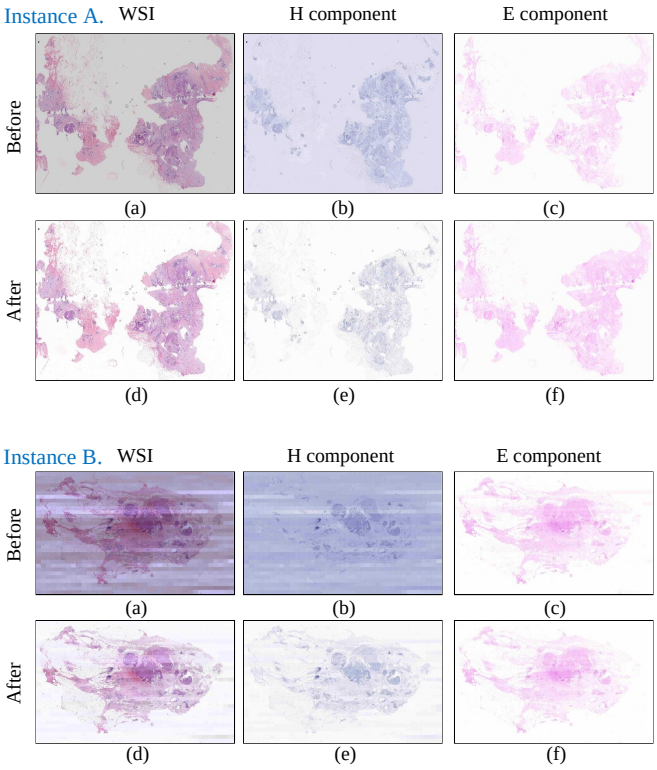


Fig. 3. Two instances of normalization, where (a) in each instance shows the original WSI, (b) and (c) display the H & E components separated from (a) by color deconvolution, as well as (d), (e), and (f) are the images after normalization for (a), (b), and (c), respectively.

RGB channels. To directly process the staining information, we utilized the color deconvolution (CD) technique [32] to separate the H and E components from RGB-stored WSIs. With CD, the parameter matrix is estimated using WSIs in an ideal imaging situation, and is constant in terms of application. However, the quality of digital WSIs varies under different scanning situations, which decreases the robustness of CD. To achieve a robust staining separation performance, we propose normalizing the luminance and saturation in the hue-saturation-value (HSV) space prior to the CD. Letting $(h_k, s_k, v_k), h_k, s_k, v_k \in [0, 1]$ denote the hue, saturation, and value of the k-th pixel in an HSV space, the normalization is defined through the following equations:

$$\begin{aligned} \hat{h}_k &= h_k \\ \hat{s}_k &= \begin{cases} 0 & , s_k < \bar{s}_{back} \\ (s_k - \bar{s}_{back})/(1 - \bar{s}_{back}) & , s_k \geq \bar{s}_{back} \end{cases} \\ \hat{v}_k &= \begin{cases} 1 & , v_k > \bar{v}_{back} \\ v_k/\bar{v}_{back} & , v_k \leq \bar{v}_{back} \end{cases} \end{aligned} \quad (1)$$

where $(\hat{h}_k, \hat{s}_k, \hat{v}_k)$ is the result of normalization, and $\bar{s}_{back}$ and $\bar{v}_{back}$ are the mean values of the background regions in the saturation and value channels, respectively. The background is defined as pixels that have the 5% lowest in the saturation channel, which are generally occupied by blank regions surrounding the tissue. The normalization is applied to all the WSIs in the experimental database. Letting $\overline{H}$ and $\overline{E}$ denote the average value of H/E component in a WSI, the allocations of $\overline{H}$ and $\overline{E}$ in the database are presented

in Fig. 2. After the normalization, $\overline{H}$ and $\overline{E}$ of the WSIs are more uniform than those before the normalization. Fig. 3 shows the normalization performance of two digital WSIs with different imaging quality. With the normalization, the signal in H channel is more discriminative, and the background is clearer than before. The following analysis is based on the normalized data of the HE-staining-space.

2) Super-pixel: To achieve fine analysis, histopathological images are usually segmented into irregular sub-regions [33], [14]. In this paper, a sub-region is defined by a super-pixel, which is segmented using a linear spectral clustering (LSC) method [34]. The segmentation is processed in the HE-staining-space, which directly represents the allocation of nuclei and stroma in the WSI. Furthermore, the H channel is smoothed prior to segmentation using a Gaussian filter to relieve the sawtooth-shaped border between super-pixels caused by sparsely distributed nuclei. And the E channel is not smoothed, for it affects little to the performance of segmentation in the experiment. Fig. 4 compares the super-pixels segmented by LSC in the RGB and HE space[2]. Clearly, a super-pixel segmented in the HE space grasps the meaningful objects (e.g., epithelium) more effectively, and has a more regular border than a super-pixel segmented in the RGB space.

3) Context definition: To utilize the contextual information, the spatial relationship among super-pixels is defined. Let $p_i, i = 1, 2, \ldots, N_s$ denote the i-th super-pixel included in a WSI, N_s denote the number of super-pixels in the WSI, and $\mathbf{A} \in \mathbb{R}^{N_s \times N_s}$ be the adjacency matrix where $a_{ij} = 1$ indicates that p_j is adjacent to p_i, and $a_{ij} = 0$ otherwise. Then, the super-pixels next to p_k are described through the following collection:

$$R(p_k) = \{p_j | a_{kj} = 1, a_{kj} \in \mathbf{A}, j = 1, 2, \ldots, N_s\}. \quad (2)$$

Based on $R(p_k)$, the different scales of regions centered on p_k are defined by

$$C_k^{(n)} = \begin{cases} \{p_k\}, & n = 0 \\ C_k^{(0)} \cup R(p_k), & n = 1 \\ \bigcup\limits_{p_j \in C_k^{(n-1)}} R(p_j), & n > 1 \end{cases}$$

where n denotes the scale of the region surrounding p_k. In particular, $C_k^{(0)}$ is defined to represent p_k itself. Fig. 5 illustrates the regions defined by $R(p_k)$ and $C_k^{(n)}$ with $n = 0, 1, 3, 7$, which are four regions considered in the feature extraction and retrieval stages.

4) Feature extraction: The scale-invariant feature transform (SIFT) [35] based bag of features (BoF) [36] representation has proven to be effective in CBHIR [24]. Referring to [24], the points detected by SIFT are highly correlated with nuclei centers, and are thereby effective in locating crucial regions in histopathological images. Nevertheless, descriptors of SIFT are designed for nature images, rather than histopathological images. To obtain more discriminative representations for histopathological images, we prefer to replace SIFT descriptors with a designed neural network to extract features from the SIFT points. A previous study of DNN-based feature extraction from key points of the histopathological ROIs was reported in [31]. In this paper, we extend this for super-pixels at different magnifications. Fig. 6 illustrates a flowchart of the feature extraction for a certain super-pixel. Because pathologists generally analyze a WSI under lenses at different magnification for diagnosis, we propose extracting features from digital WSIs at four different magnifications. Correspondingly, the SIFT points are divided into four groups according to the scale, and assigned to the four magnifications, based upon which the patterns of the histopathological images are extracted. Moreover, the four regions of contextual information are considered in the feature extraction stage. The relationship between magnification and context region are listed in Table I, and illustrated in Fig. 6(b). More details on feature extraction are presented in Supplemental material A. (Supplementary materials are available in the supplementary files /multimedia tab.)

TABLE I
REGIONS AND MAGNIFICATIONS CONSIDERED IN THE FOUR CONTEXT REGIONS.

Context index	Context region	Magnification of lens	Resolution
$l = 0$	$C_k^{(0)}$	20×	1.2μm/pixel
$l = 1$	$C_k^{(1)}$	10×	2.4μm/pixel
$l = 2$	$C_k^{(3)}$	5×	4.8μm/pixel
$l = 3$	$C_k^{(7)}$	2×	12μm/pixel

5) Binarization: The binarization of features is the basis of efficient retrieval from large-scale database. Therefore, the features extracted above are converted into binary codes to improve the retrieval efficiency and reduce the storage of WSI features in the database. Letting $\mathbf{x_i} \in \mathbf{R}^d$ denote the feature vector of the i-th super-pixel, the binarization of $\mathbf{x}_i$ can be commonly represented as

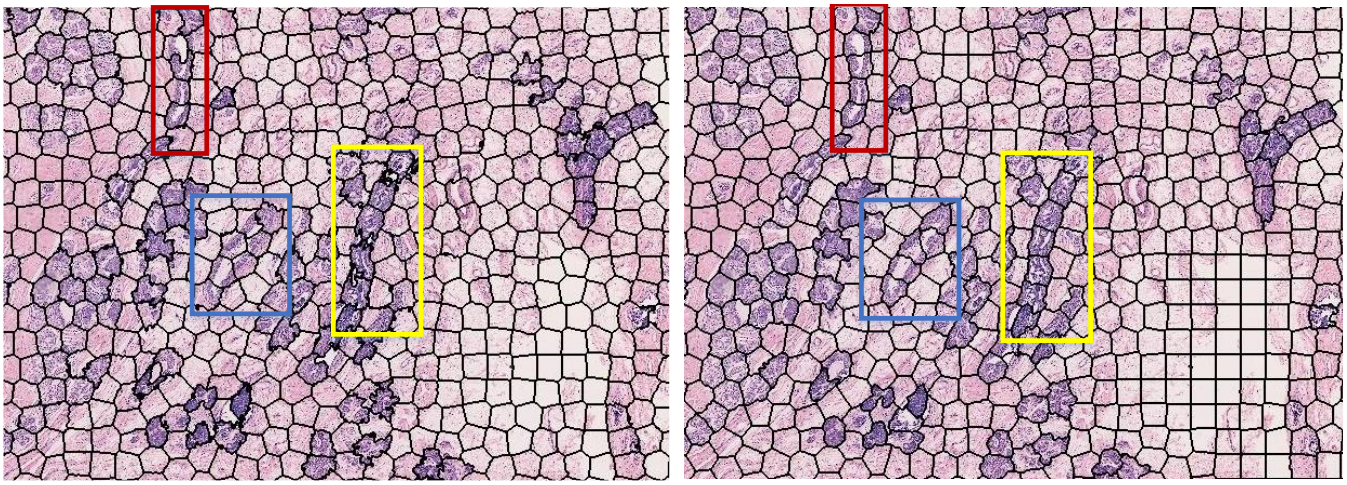
$$\mathbf{b}_i = \mathbf{h}(\mathbf{x}_i), \quad (3)$$

where $\mathbf{b}_i = (b_{i1}, b_{i2}, ..., b_{iK})$ indicates the binary code of $\mathbf{x}_i$, $\mathbf{h}(\cdot) = \{h_1(\cdot), h_2(\cdot), \ldots, h_K(\cdot)\}$ denotes a set of binary functions with $h_j : \mathbf{R}^d \longmapsto \{1, -1\}$[1], and K is the function number, namely, the bit number of the binary code $\mathbf{b}_i$. For each super-pixel, the features extracted from the four magnifications are separately converted into binary codes. Consequently, each WSI can be represented by a table that records the index of super-pixels and the corresponding binary codes in the four magnifications. It is a memory-saving representation. Supposing that the total length of a binary representation for a super-pixel N is set to 192-bits and that the WSI is segmented into about 10 K super-pixels (the average for the database used in the experiment), the amount of memory required to store a WSI is about 240 Kb.

C. Aided-diagnosis based on CBHIR

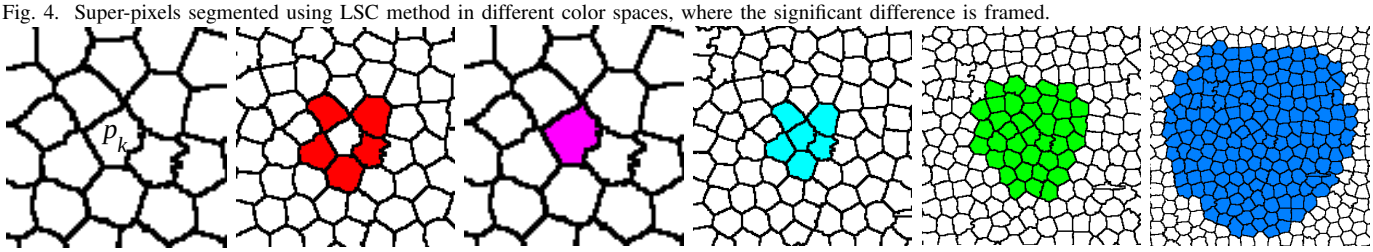
In this paper, a probability of malignancy for each super-pixel is given through a context-based CBHIR approach. The details of the proposed method are provided in this section.

[2]The segmentation is completed in the HE space and the result is displayed on the RGB-colored image.



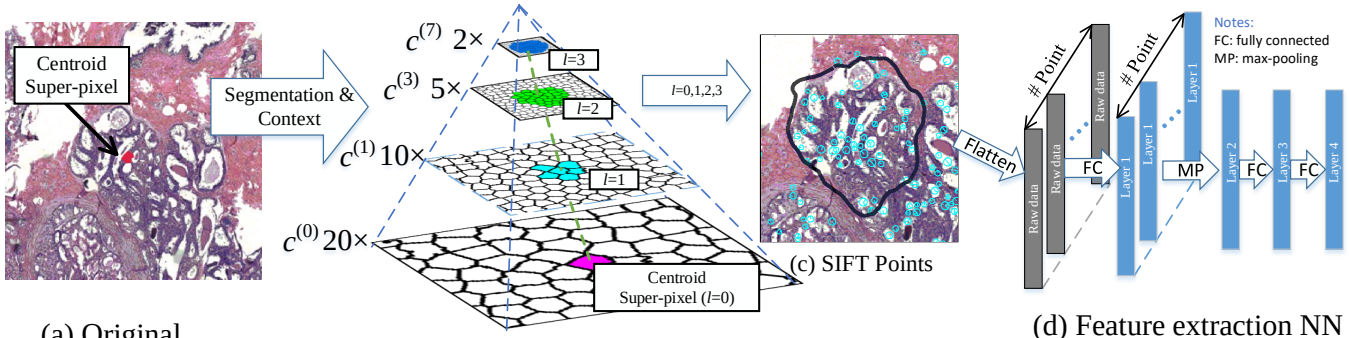
(a) Segmentation in RGB space. (b) Segmentation in HE space.

Fig. 4. Super-pixels segmented using LSC method in different color spaces, where the significant difference is framed.



(a) p_k (b) $R(p_k)$ (c) $C_k^{(0)}$ (d) $C_k^{(1)}$ (e) $C_k^{(3)}$ (f) $C_k^{(7)}$

Fig. 5. Definition of contextual regions for super-pixel p_k, where $R(p_k)$ consists of super-pixels adjacent to p_k, $C_k^{(0)}$ is defined as p_k itself, and $C_k^{(n)}, n = 0, 1, 3, 7$ are the four context regions considered in our analysis.



(a) Original (b) 4 level context regions (d) Feature extraction NN

Fig. 6. Flowchart of feature extraction, where (a) shows a super-pixel in the region, (b) displays in a pyramid the context regions of the super-pixel, (c) shows a context region and the SIFT points located in it, and (d) illustrates the structure of neural networks (containing three fully connected (FC) layers and a max-pooling (MP) layer) used in this paper.

1) Context-based CBHIR: Based on the binary codes of super-pixels, the similarity of super-pixels p_i and p_j can be defined as

$$s(p_i, p_j) = \mathbf{b}_i^{\mathrm{T}} \mathbf{b}_j,$$

which is an efficient computation of Hamming distance between $\mathbf{b}_i$ and $\mathbf{b}_j$ [37]. In this paper, we extend this similarity measurement with the contextual information of the two super-pixels, defining a context-based similarity measurement:

$$S(p_i, p_j) = \sum_{l=0}^{L} \lambda_l s^{(l)}(p_i, p_j), \tag{4}$$

where $s^{(l)}(p_i, p_j)$ denotes the similarity between p_i and p_j in context region l (defined in Table I), L defines the level of contextual information considered in the measurement, and λ_l is the weight. $s^{(l)}(p_i, p_j)$ is defined as

$$s^{(l)}(p_i, p_j) = \mathbf{b}_i^{(l)\mathrm{T}} \mathbf{b}_j^{(l)}, \tag{5}$$

where $\mathbf{b}_i^{(l)}$ denotes the binary codes generated from the features of the context region l.

For a certain super-pixel in an unknown WSI, the similarities between the super-pixel and all super-pixels in the

database can be measured using Eq.4. After ranking, the top-M-similar super-pixels in the database can be retrieved.

2) Aided-diagnosis: Repeating the context-based CBHIR processing throughout an unknown WSI, the retrieval results for all super-pixels are obtained. Because the WSIs in the database are completely annotated, the probability of malignancy for each super-pixel can be calculated by quantifying the amount of malignant return M_{mlg} in the top-M-similar super-pixels. Specifically, the probability of malignancy for the i-th super-pixel is defined as

$$P(mlg|p_i) = M_{i,mlg}/M,$$

where $M_{i,mlg}$ represents the amount of malignant return in the top-M-similar super-pixels for i-th super-pixel. Clearly, $P(mlg|p_i) \in [0, 1]$. Then, a probability map regarding malignancy is generated by $P(mlg|p_i)$ for all the super-pixels in the unknown WSI. When a doctor opens a digital WSI, the probability map is shown, based upon which the regions with high-malignant-risk can be recognized easily. Because an aided-diagnosis is achieved by CBHIR, the retrieval results for each super-pixel are recorded and made available to the doctor during diagnosis.

III. Experiment on Motic Dataset

A. Experimental setting

The dataset was supplied by Motic[3]. A total of 145 WSIs from 145 patients with epithelial breast tumors were used in the experiment, and 95 WSIs were used as the training set and the remainder used for testing. All regions with malignant tumors in the 145 WSIs were annotated by trained pathologists from the general air force hospital of PLA, China. In the dataset, 83 WSIs were annotated as showing a malignant tumor, and 62 WSIs were annotated as showing a benign tumor. According to the annotation, the super-pixels in these WSIs were labeled as malignancy or non-malignancy. The size of the super-pixels in the experiments is around 2,500 pixels under a 2× lens. The parameters of the proposed model were determined using a super-pixel-level retrieval task for the training set, and the performance of the WSI-wise analysis was evaluated for the testing set.

B. Parameter selection

Regarding the super-pixels of the training WSI as individual samples, we conducted experiments to determine the parameters involved in the proposed context-based CBHIR framework. In total, 19,641 malignant and 18,560 non-malignant super-pixels were randomly selected from the 95 training WSIs, based upon which, a five-fold cross-validation was employed. For each fold, the super-pixels of one-fifth of the WSIs were regarded as the query ROIs and remainder were the retrieval database. A feature extraction network, a binarization model, and a context-based retrieval model need to be determined during the training stage, the details of which are presented in this section.

[3]Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen 361101, China

1) Feature extraction network: The feature extraction neural network consists of three fully connected layers and a max-pooling layer. The activation function for the fully connected layers is a sigmoid function, which is defined as $\sigma(t) = 1/(1+e^{-t})$. The three fully connected layers (layer 1, 3, and 4 in Fig. 6(d)) of the neural network were pre-trained using sparse auto-encoders [38], and then fine-tuned according to the labels of the super-pixels. The number of nodes in the four contextual levels were experimentally determined (the details of which are given in Supplemental material A.). Finally, the setting of each layer is shown in Table II.

TABLE II
Number of nodes of the feature extraction networks for four context regions.

Context index	Layer 1	Layer 2	Layer 3	Layer 4
$l = 0$	260	260	260	150
$l = 1$	260	260	200	150
$l = 2$	260	260	120	150
$l = 3$	280	280	180	150

2) The binarization methods: Five data-dependent binarization methods were validated as effective for histopathological image retrieval [2], [27], [28]. They are

- *Thresholded PCA (tPCA)* [39]: the feature $\mathbf{x}$ is projected to low-dimensional codes using principal component analysis (PCA) and then converted into binary codes using 0 as the threshold. Letting $\mathbf{w}_k$ denote the k-th projection vector of PCA, the k-th binary function of tPCA can be represented as $h_k(\mathbf{x}) = sgn(\mathbf{w}_k^{\mathrm{T}}\mathbf{x})$.
- *Iterative quantization (ITQ)* [40]: The projection matrix of tPCA is further optimized by minimizing the quantization error between low-dimensional codes and the binary codes. And the binary function is the same as tPCA.
- *Binary autoencoder (BA)* [41]: The feature $\mathbf{x}$ is converted to low-dimensional codes through non-linear functions trained using auto-encoders. Letting σ denote a non-linear function, vector $\mathbf{w}_k$ and b_k be the weight and bias for the k-th non-linear function, the binary function of BA can be represented as $h_k(\mathbf{x}) = sgn(\sigma(\mathbf{w}_k^{\mathrm{T}}\mathbf{x} + b_k))$.
- *Binary factor analysis (BFA)* [41]: A approximate solution of BA, where the function σ is linear.
- *Supervised hashing with kernels (KSH)* [37]: The feature $\mathbf{x}$ is converted to higher dimensional space using kernel functions and then projected there into binary codes. In addition, pairwise labels between features are established and used to improve the discrimination of the binary codes.

These five methods are considered candidates in our framework. To evaluate the retrieval performance for a query super-pixel, the average precision (AP) of the top-M-returned results was calculated. And the mean average precision (MAP) of all the testing super-pixels is used as the metric, which is defined as

$$\mathrm{MAP} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{M}, \tag{6}$$

where M is the number of returned super-pixels in retrieval, N is the number of query super-pixels, and T_i denotes the number

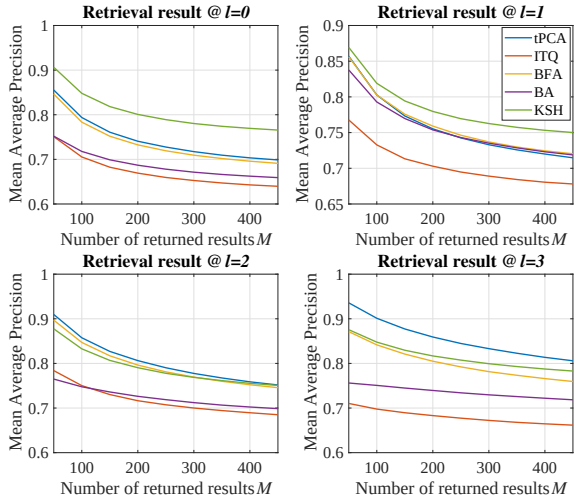


Fig. 7. Retrieval performance of five different binarization methods under similarity defined by $s^{(l)}(p_i, p_j)$ (Eq.5).

of returned super-pixels that share the same label with the i-th super-pixel. For each level of context, the binarization model was selected from the five binarization methods according to the retrieval performance. The MAPs of retrieval for the different magnifications (defined by l in Eq.5) were separately calculated. The results for different return number M are illustrated in Fig. 7. For retrieval under $s^{(0)}(p_i, p_j)$ and $s^{(1)}(p_i, p_j)$, KSH delivers the best performance, whereas for $s^{(2)}(p_i, p_j)$ and $s^{(3)}(p_i, p_j)$, tPCA performs the best. Therefore, KSH was used for the features extracted under the $20\times$ and $10\times$ lenses, and tPCA was used for features extracted under $5\times$ and $2\times$ lenses.

3) Context-based retrieval: To verify the effectiveness of the contextual information, the retrieval performance obtained under different contextual levels ($L = 0, 1, 2, 3$) was compared. In the context-based similarity measurement $S(p_i, p_j)$ (defined in Eq. 4), the parameter λ_l (defined in Eq. 2) controls the weight of information from the l-th context region in the similarity measurement. To reach an appropriate combination of contextual information, λ_l was optimized according to MAP through a cross-validation in the training set, and then $S(p_i, p_j)$ was determined. Specifically, λ_l was determined in greedy manner. The procedure for context level L is presented in Algorithm 1. The ratio of optimized λ_l is illustrated in Fig. 8, and the corresponding optimal MAPs are shown in Fig. 9. The retrieval accuracy increases as L enlarges. This proves that the combination of multiple levels of contextual information can improve the performance of CBHIR. Further, the ratios of $\lambda_l, l = 0, 1, 2, 3$ are proportionate to each other, which demonstrates that the four levels of contextual information are all necessary for an accurate retrieval.

C. Evaluation for WSI analysis

Using the optimized models, we conducted experiments to evaluate the performance of the proposed method for a WSI-analysis. All super-pixels in the unknown WSI were predicted through context-based CBHIR. As stated in II-C2, a probability map of malignancy for a testing WSI can be generated. Regarding the proposed retrieval model as a K-nearest neighbor (KNN) classifier, each super-pixel can be classified as malignant or non-malignant. Specifically, a super-pixel is classified as a malignant sample if $P(mlg|p_i) > 0.5$; otherwise, it is classified as a non-malignant sample. Then, the entire WSI is segmented by classifying all super-pixels in it.

A MAP with $M = 100$ is used to evaluate the performance of the super-pixel-wise retrieval. Three metrics, sensitivity (true-positive-rate (TPR)), specificity (true-negative-rate (TNR)), and accuracy, were used to evaluate the performance

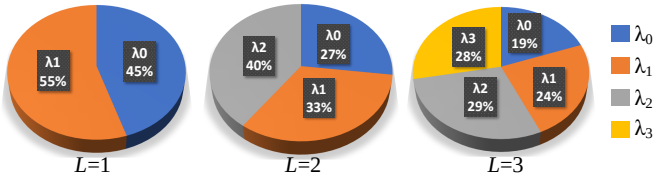


Fig. 8. Optimal weights $\lambda_l, l = 0, 1, 2, 3$ (Eq. 4) of the four levels of contextual information determined in the cross-validation.

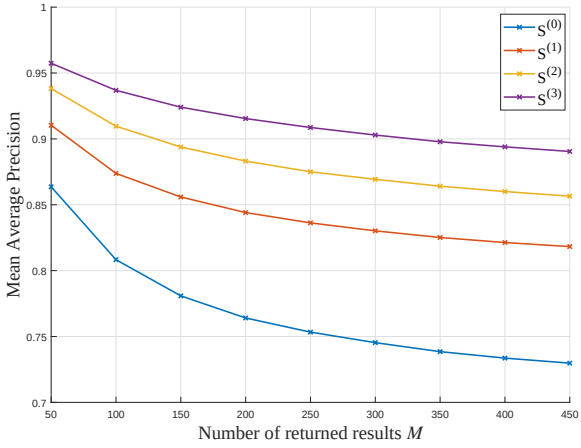


Fig. 9. Performance of retrieval with optimal percentage of contextual information.

Data: L : The level of context ($L > 0$)
Result: S : The context-based similarity measurement
$\lambda_0 \leftarrow 1, S = s^{(0)}$;
for $l = 1$ *to* L **do**
  Find λ_l via line search to maximize MAP under the similarity measurement $S + \lambda_l s^{(l)}$;
  $\lambda_l = argmax_\lambda \text{MAP}(S + \lambda s^{(l)})$;
  for $i = 0$ *to* l **do**
    $\lambda_i \leftarrow \lambda_i / \sum_{i=0}^{l} \lambda_i$;
    $S \leftarrow \sum_{i=0}^{l} \lambda_i s^{(i)}$;
  end
end
return S;

Algorithm 1: Optimization of the similarity measurement $S(p_i, p_j)$ with level L, where $S(p_i, p_j)$ abbreviates as S, $s^{(0)}(p_i, p_j)$ as $s^{(0)}$, and MAP(S) represents the MAP of retrieval under similarity measurement $S(p_i, p_j)$.

of super-pixel-wise classification. These metrics are defined through the following equations:

$$\begin{aligned} Sensitivity &= \frac{TP}{TP+FN}, \\ Specificity &= \frac{TN}{TN+FP}, \\ Accuracy &= \frac{TP+TN}{TP+TN+FP+FN}. \end{aligned} \tag{7}$$

where TP denotes the number of correctly classified malignant super-pixels, TN denotes the number of correctly identified benign super-pixels, FP and FN are the number of incorrectly classified malignant and benign super-pixels, respectively. In addition, the WSI-wise performance is also considered, which is quantified through the following equations:

$$\begin{aligned} Sensitivity &= \frac{1}{N_w}\sum_{w=1}^{N_w}\frac{TP_w}{TP_w+FN_w}, \\ Specificity &= \frac{1}{N_w}\sum_{w=1}^{N_w}\frac{TN_w}{TN_w+FP_w}, \\ Accuracy &= \frac{1}{N_w}\sum_{w=1}^{N_w}\frac{TP_w+TN_w}{TP_w+TN_w+FN_s+FP_w}. \end{aligned} \tag{8}$$

where TP_w, TN_w, FP_w, and FN_w represent TP, TN, FP, and FN in the w-th WSI, respectively, and N_w is the number of testing WSIs.

1) Comparative methods: The features used in the proposed method are extracted using a neural network (NN) from the HE-space, and the aided-diagnosis approach applied is context-based CBIR. Hence, the proposed method is abbreviated as HE-NN-C-CBIR. In this section, the state-of-the-art histopathological image classification and retrieval frameworks are compared. They are

- *DFDL* (Vu et al. [5]): The patches centered on the SIFT points are classified using the DFDL algorithm. In addition, each super-pixel is predicted by the majority vote of the prediction of patches located in the super-pixel.
- *DCNN* (Xu et al. [11]): The super-pixels are cropped by their bounding boxes, then resized to 28×28, and finally classified using a deep convolutional neural network.
- *SIFT-BoF-CBIR* (Zhang et al. [24]): SIFT descriptors are extracted from histopathological images and are quantified using a bag of features (BoF) model to represent super-pixels. The super-pixels are then classified through the CBIR method using the KSH [37] model.
- *LDA-SH-CBIR* (Ma et al. [28]): The super-pixels are represented as binary codes using latent-Dirichlet-allocation-based (LDA-based) [19] supervised hashing (SH). The super-pixels are then classified through the CBIR method.

In the four methods, DCNN was processed with images under a $2\times$ lens and the other methods were processed with images under a $20\times$ lens. In addition, to verify the proposed method, three degraded models of HE-NN-C-CBIR are implemented:

- *SIFT-BoF-C-CBIR*: SIFT-based BoF representations are used to describe the super-pixels. The other parts of the model are the same as HE-NN-C-CBIR.
- *O-HE-NN-C-CBIR*: The histopathological images used are not normalized. The other steps are the same as HE-NN-C-CBIR.
- *HE-NN-CBIR*: Although the feature extraction and binarization steps are the same as HE-NN-C-CBIR, the contextual information is not considered in the CBIR procedure. Specifically, the similarity is measured using Eq. 4 with $L = 0$.

All the methods are implemented using MATLAB 2017 on a PC with 12 cores of 2.10 GHz and evaluated in our database. The models of the comparative methods are optimized on the training set using the same five-fold cross-validation as the proposed method. In addition, the comparison is conducted on the testing set using the optimized models. For a fair comparison, the MAPs of the retrieval-based methods are all calculated using $M = 100$ (defined in Eq. 6).

2) Accuracy for Retrieval and Classification: Table III shows the performance of the compared methods. Note that the WSIs that contain no malignant (benign) regions are not considered in the calculation of the WSI-wise sensitivity (specificity). Therefore, the WSI-wise sensitivity reported in Table III is calculated from 38 WSIs, and the WSI-wise specificity is from 44 WSIs.

Overall, HE-NN-C-CBIR achieves the best performance with a MAP of 90.5% in super-pixel-wise retrieval, an accuracy of 94.9% in super-pixel-wise classification, and an accuracy of 94.1% in WSI-wise segmentation. This proves that the proposed context-based CBHIR framework is effective for WSI-wise breast lesion analysis.

The retrieval and classification performance of HE-NN-CBIR is better than that of SIFT-BoF-CBIR [24]. This certifies that the designed feature extraction network is more effective in histopathological image representation than SIFT-BoF features. This can also be concluded from the comparison between HE-NN-C-CBIR and SIFT-BoF-C-CBIR. The features in O-HE-NN-C-CBIR are extracted from the HE-space separated from the WSIs without normalization, where some WSIs suffer from poor imaging quality. Utilizing the WSI normalization approach, the retrieval accuracy of HE-NN-C-CBIR improves 2.8%, and the WSI-segmentation accuracy increases 3.0%. This shows that the normalization approach is effective and necessary in a WSI analysis. HE-NN-C-CBIR achieves an accuracy of 94.9% in the super-pixel-wise classification, which is 6.7% higher than that of HE-NN-CBIR. Such improvement is also significant for models using SIFT-BoF representations (see the results of rows 3 and 6 in Table III). This certifies that the patterns from different scales and different magnifications surrounding a sub-region are complementary in the identification of malignant regions from WSIs.

The performance of DCNN [11] and LDA-SH-CBIR [26] are competitive in comparison with HE-NN-CBIR (the single-scale version of the proposed method) but cannot surpass HE-NN-C-CBIR. The main reason is that the features used in DCNN and LDA-SH-CBIR are extracted from a single

TABLE III
PERFORMANCE OF SUPER-PIXEL-WISE RETRIEVAL, CLASSIFICATION, AND WSI-WISE SEGMENTATION OF THE COMPARED METHODS, WHERE NO. 1 AND NO. 2 ARE HIC-BASED APPROACHES AND THE REMAINDER ARE CBIR-BASED METHODS.

No.	Methods	SP-wise retrieval	SP-wise classification			WSI-wise segmentation			Context
		MAP	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	
1	*DCNN*[11]	-	0.888	0.848	0.867	0.874	0.804	0.865	
2	*DFDL*[5]	-	0.857	0.777	0.802	0.833	0.712	0.789	
3	*SIFT-BoF-CBIR*[24]	0.756	0.800	0.827	0.813	0.763	0.835	0.802	
4	*LDA-SH-CBIR*[28]	0.847	0.911	0.882	0.896	0.834	0.781	0.879	
5	*HE-NN-CBIR*	0.836	0.889	0.876	0.882	0.812	0.781	0.870	
6	*SIFT-BoF-C-CBIR*	0.861	0.897	0.918	0.908	0.847	**0.886**	0.891	√
7	*O-HE-NN-C-CBIR*	0.877	0.935	0.907	0.921	0.882	0.801	0.911	√
8	*HE-NN-C-CBIR*	**0.905**	**0.960**	**0.939**	**0.949**	**0.921**	0.845	**0.941**	√

TABLE IV
AVERAGE RUNNING TIME PER WSI.

Methods	Time Consuming (min)				
	Seg.	Encode	Retrieve	Others	Total
DCNN [11]	0.72	2.11	-	-	2.83
DFDL [5]		8.13	-	-	8.85
SIFT-BoF-CBIR [24]		8.31	5.36	-	14.39
LDA-SH-CBIR [28]		26.60	5.36	-	32.68
HE-NN-CBIR		4.07	5.36	0.09	10.24
HE-NN-C-CBIR		6.54	6.95	0.41	14.62

magnification and single scale of the super-pixels (2× lens for DCNN and 20× lens for LDA-SH-CBIR). The patterns surrounding a super-pixel are not applied in prediction. HE-NN-C-CBIR considered the surrounding regions of super-pixels in the feature extraction and prediction stages, thereby achieving higher classification and retrieval accuracies.

3) Computational Complexity: For the compared methods, the average running time used to analyze a WSI is presented in Table IV. All methods are based on super-pixels segmented using LSC, and thereby take the same amount of time, namely, 0.72 min. The time cost for feature extraction and binarization is attributed to *Encode*, *Retrieve* includes the computations and sorting of the similarities, and *Others* includes the time for pre-processing and context construction. Fig. 10 presents the joint performance of WSI-wise segmentation and time consumption. DCNN cost the least amount of time because it was conducted on WSIs under a 2× lens. For methods based on images under a 20× lens, LDA-SH-CBIR achieved the best accuracy but cost more than 30 min. Compared with HE-NN-CBIR, HE-NN-C-CBIR obtained an accuracy 6.7% higher, at an additional cost of 42.7% in running time.

4) Visual Results: Fig. 11 illustrates segmentation performances of several typical WSIs generated using the compared state-of-the-art methods, in which the performances of the WSI analysis are intuitively presented. Overall, the results using the proposed method, HE-NN-C-CBIR, are the most similar with the ground truth, which is consistent with the conclusion reached from the numerical results. Segmentation using the models by Ma et al. [26], and Xu et al. [11], effectively highlights the malignant regions in WSIs, but contains a large amount of false-positive regions. And results from Zhang et al. [24] and HE-NN-CBIR have evident false-negative regions. In contrast, considering the contextual information, HE-NN-C-CBIR filtered the majority of isolated false-identified regions

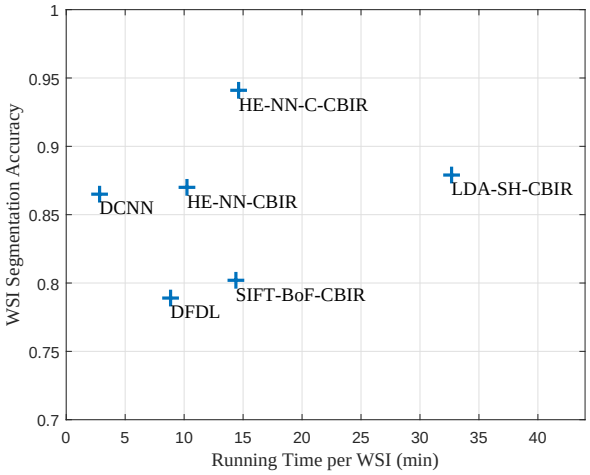


Fig. 10. Joint performance of WSI-wise segmentation and running time.

and achieved a more accurate segmentation. Fig. 12 shows the probability maps that generated using our method with $L = 0, 1, 2, 3$. With L enlarged, the probability map becomes more accurate and the malignant regions are more discriminative to the benign regions. Therefore, the probability maps generated using HE-NN-C-CBIR are more reliable, allowing pathologists to locate diagnostically relevant regions.

IV. RESULTS ON CAMELYON 16 DATASET

With specific post-processing to the segmentation output, the proposed framework can be used as cancerous detection task. We conducted the cancer detection experiments on Camelyon 16 database.

The dataset was from the Camelyon 16[4] challenge of cancer metastasis detection in lymph node. It contains a total of 400 whole-slide images (WSIs) of sentinel lymph node (240 Normal and 160 containing metastases), where 270 WSIs are used for training and the remainder are used for testing. All the regions with cancer in these WSIs are annotated by pathologists.

In the database, an average of 33,647 super-pixels were segmented from each WSI. However, there were many blank regions in these WSIs. To reduce the computation, these blank regions were filtered by a threshold after the normalization

[4]https://camelyon16.grand-challenge.org/[accessible 2018-01-16]

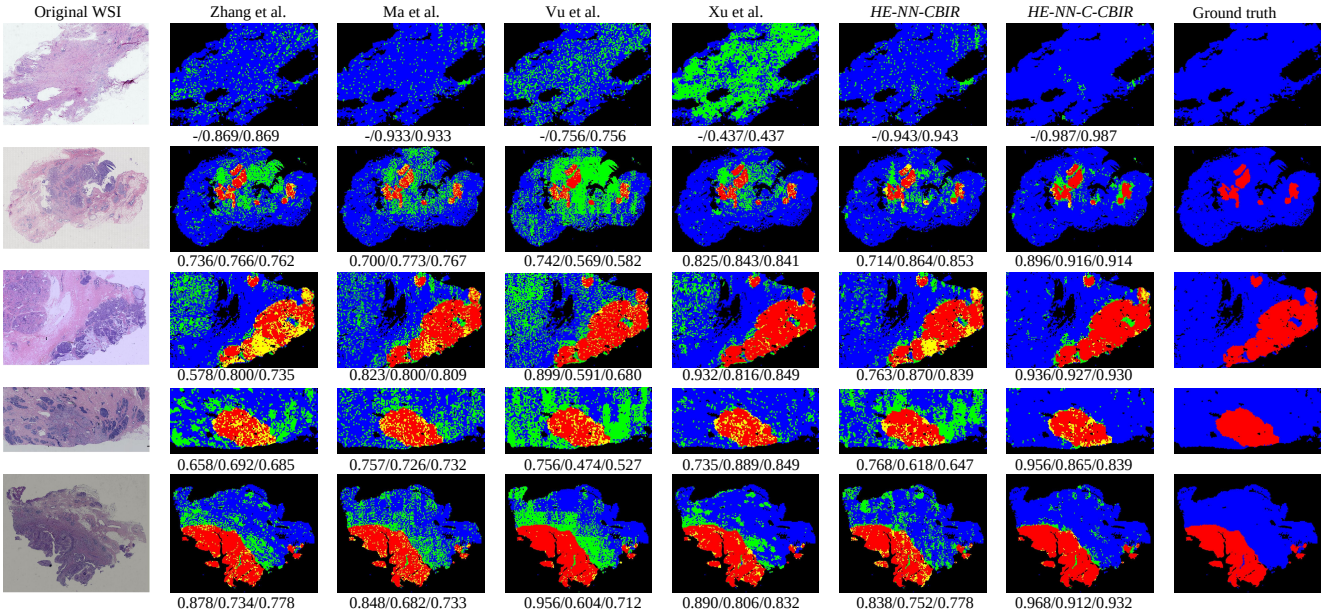


Fig. 11. Visual performance of different methods for identifying malignant regions, where true-positive (TP) regions are displayed in red, true-negative (TN) regions are displayed in blue, false-positive (FP) regions are shown in green, false-negative (FN) regions are in yellow, and the sensitivity/specificity/accuracy of WSI-wise segmentation is presented under each result. In addition, the ground truth for each WSI is given on the last column, where red denotes malignant regions, blue denotes benign regions, and black denotes background.

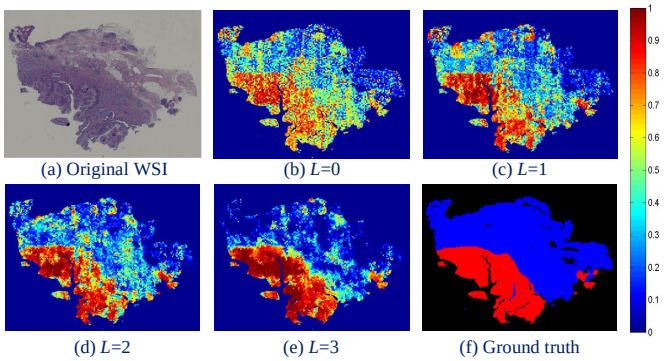


Fig. 12. Probability maps generated by HE-NN-C-CBIR with $L = 0, 1, 2, 3$, where red represents $P(mlg|p_i) = 1$, blue represents $P(mlg|p_i) = 0$, and the map between color and $P(mlg|p_i)$ is given on the right.

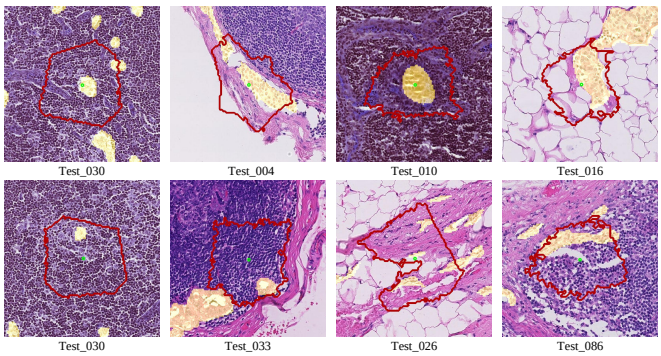


Fig. 13. Visualization of true-positive super-pixels classified on Camelyon 16 dataset, where a super-pixel is displayed with a red contour, the center of the super-pixel is drawn in green, and the ground truth for cancer detection is displayed in yellow. According to the metric of Camelyon 16, the super-pixels in the first row are regarded as true-positive, and super-pixels in the second row are regarded as false-positive.

step. Specifically, a super-pixel was regarded as background if its mean in the value channel (of HSV-space) $\overline{v}$ is above 200. Then, the average number of valid super-pixels for a WSI was reduced to 3,226. The optimization of the framework on Camelyon 16 dataset followed the same paradigm as that on our database. The binarization methods for the four context regions were all determined as KSH. The combination weights λ_l of context-based similarity measurement were optimized to $(0.36, 0.24, 0.28, 0.12)$. In the testing stage, the background super-pixels were directly classified as the negative. And the other super-pixels were classified using the optimized model. As a result, the sensitivity, specificity, and accuracy of super-pixel-wise classification were 0.953, 0.986, and 0.983 in the testing set. And the running time per WSI was 9.95 min, in which feature extraction cost 5.21 min, and retrieval cost 4.25 min.

The performance of cancer detection in Camelyon 16 is evaluated using free-response receiver operating characteristic curve (FROC), which is defined as the plot of sensitivity versus the average number of false-positive per WSI. And the mean sensitivity at average false-positive numbers 0.25, 0.5, 1, 2, 4, and 8 per WSI are calculated as the score. To compute sensitivity, a detected tumor region needs to be represented as a point with a cancerous probability. The point will be regarded as true-positive when it is within a annotation of tumor, or it will be considered as false-positive, otherwise.

In our framework, the unit of detection is super-pixel. Then, we simply extracted the center of positive super-pixels as the detected results. As a result, the score of detection for our framework is 0.567, which is between the sores of ranks 7 and

TABLE V
COMPARISON OF DETECTION SCORE ON CAMELYON 16 DATASET.

Method	#False-positive						Score	Rank
	0.25	0.5	1	2	4	8		
Harvard Medical School and MIT, Method 2	0.773	0.778	0.813	0.827	0.827	0.827	0.807	1
Radboud University Medical Center (DIAG), Netherlands	0.493	0.524	0.569	0.600	0.631	0.631	0.575	7
The Chinese University of Hong Kong (CU lab) - Method 1	0.404	0.471	0.493	0.582	0.631	0.684	0.544	8
HE-NN-C-CBIR	0.357	0.574	0.604	0.617	0.622	0.628	0.567	-

8 on the leader board of detection. The results[5] of methods that are typical to our method are given in Table V.

For our method, the sensitivity of detection is much lower than the sensitivity of super-pixel-wise classification on Camelyon 16 dataset. The reduction of sensitivity is from the procedure of converting super-pixels to detected points. In this dataset, there are a certain number of tumor regions that are much smaller than the super-pixels used in our framework (as shown in Fig, 13), so that the center of a true-positive super-pixel may be out of the ground truth (the second row of Fig. 13). As a consequence, these super-pixels were regarded as false-positive, and the tumor regions within the super-pixels were regarded as false-negative in the evaluation stage. Anyway, these super-pixels successfully cover the tumor regions. Therefore, it is still effective for pathologists in locating key regions in a WSI.

V. DISCUSSION

The performance of the framework was evaluated on breast datasets. However, the entire framework contains few hand-crafted components for breast cancer, and procedures including super-pixel segmentation, feature extraction, binarization, and diagnosis stages are common to all histopathological images. Therefore, the framework can be extended to other types of cancer essentially without modification. Certainly, these stages can be improved or redesigned to refine the framework for specific lesions.

The probability map of the aided-diagnosis is quantified from the labels of the retrieved super-pixels. In our dataset, the tissues in our database are labeled as binary (benign or malignant). As an extension, the framework can also be applied for multi-class tasks if the database provides corresponding annotations, e.g., the level of progress and the sub-categories of the cancer. Specifically, a normalized histogram can substitute the probability defined in Section II-C2 to estimate the probabilities of multiple classes, which can be defined as

$$P(c|p_i) = M_{i,c}/M, c = 1, 2, ...C,$$

where p_i denotes the i-th super-pixel in the unknown WSI, $P(c|p_i)$ represents the probability that p_i belongs to the c-th class, $M_{i,c}$ denotes the number of returned images that belongs to the c-th class, and C is the number of classes involved in the database. Then, the probability map for each class can be separately generated based on $P(c|p_i)$.

[5]Only the results that are typical to our method are given, the complete leader board can be found at https://camelyon16.grand-challenge.org/results/[accessible 2018-01-16]

The calculation and sorting of similarities required about 6.95 min and 5.21 min in average on the two datasets, respectively, although the retrieval was completed using binary codes. The development of a more efficient retrieval scheme for WSI retrieval could improve the efficiency of the entire framework and make it applicable to a larger WSI database.

VI. CONCLUSION

In this paper, we proposed a novel histopathological WSI analysis framework for breast cancer. The contribution of this work mainly includes the following three aspects. First, we fused whole slide histopathological image classification and retrieval into an integrated framework, through which a probability map, and similar regions in historical cases for key regions in the WSI are simultaneously obtained. Next, a feature extraction approach involving multiple magnifications of sub-regions for WSIs was proposed and certified as effective for histopathological image classification and retrieval. Then, contextual information is considered in predicting the super-pixels, which contributes to a better performance. Referring to the results obtained using the proposed framework, the malignant regions can be easily recognized by pathologists, meanwhile the regions with similar content provided through retrieval can aid the pathologist in reaching a more reasonable diagnosis. Further work will focus on studying more discriminative features for histopathological images and a more efficient retrieval scheme for large-scale WSI databases.

REFERENCES

[1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[2] S. Zhang and D. Metaxas, "Large-scale medical image analytics: Recent methodologies, applications and future directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[3] J. S. Duncan and N. Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 85–106, 2000.

[4] U. Srinivas, H. S. Mousavi, V. Monga, A. Hattel, and B. Jayarao, "Simultaneous sparsity model for histopathological image representation and classification," *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1163–1179, 2014.

[5] T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. K. A. Rao, "Histopathological image classification using discriminative feature-oriented dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 738–751, 2016.

[6] X. Zhang, F. Xing, H. Su, L. Yang, and S. Zhang, "High-throughput histopathological image analysis via robust cell segmentation and hashing," *Medical Image Analysis*, vol. 26, no. 1, pp. 306–315, 2015.

[7] Y. Xu, J.-Y. Zhu, E. I-Chao, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Medical Image Analysis*, vol. 18, no. 3, pp. 591–604, 2014.

[8] M. Kandemir and F. A. Hamprecht, "Computer-aided diagnosis from weak supervision: A benchmarking study," *Computerized Medical Imaging and Graphics*, vol. 42, pp. 44–50, 2015.

[9] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. T. Brunye, and J. G. Elmore, "Localization of diagnostically relevant regions of interest in whole slide images: a comparative study," *Journal of Digital Imaging*, vol. 29, no. 4, pp. 496–506, 2016.

[10] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and 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.

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

[12] H. Chang, N. Nayak, P. T. Spellman, and 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.

[13] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, and F. A. Gonzalez, "An unsupervised feature learning framework for basal cell carcinoma image analysis," *Artificial Intelligence in Medicine*, vol. 64, no. 2, pp. 131–145, 2015.

[14] B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. A. van der Laak, "Automated detection of dcis in whole-slide h&e stained breast histopathology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2141–2150, 2016.

[15] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, "Scalable histopathological image analysis via supervised hashing with multiple features," *Medical Image Analysis*, vol. 34, pp. 3–12, 2016.

[16] G. Zhou and L. Jiang, "Content-based cell pathology image retrieval by combining different features," *Medical Imaging Pacs and Imaging Informatics*, vol. 5371, pp. 326–333, 2004.

[17] N. Mehta, R. S. Alomari, and V. Chaudhary, "Content based sub-image retrieval system for high resolution pathology images using salient interest points," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, 2009, pp. 3719–3722.

[18] T. Hofmann, "Probabilistic latent semantic analysis," in *Fifteenth Conference on Uncertainty in Artificial Intelligence*, 1999, pp. 289–296.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[20] J. C. Caicedo and E. Izquierdo, "Combining low-level features for improved classification and retrieval of histology images," *Ibai Publishing*, vol. 2, no. 1, pp. 68–82, 2010.

[21] J. Shi, Y. Ma, Z. Jiang, H. Feng, J. Chen, and Y. Zhao, "Pathological image retrieval for breast cancer with plsa model," in *Seventh International Conference on Image and Graphics*, 2013, pp. 634–638.

[22] Y. Ma, S. Jun, Z. Jiang, and F. Hao, "Plsa-based pathological image retrieval for breast cancer with color deconvolution," *Proceedings of SPIE*, vol. 8920, no. 8, pp. 89 200L–89 200L–7, 2013.

[23] Y. Zheng, Z. Jiang, J. Shi, and 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.

[24] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 496–506, 2015.

[25] X. Zhang, H. Dou, T. Ju, J. Xu, and S. Zhang, "Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, 2016.

[26] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, and Y. Zhao, "Breast histopathological image retrieval based on latent dirichlet allocation," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 1114–1123, 2017.

[27] Y. Zheng, Z. Jiang, Y. Ma, H. Zhang, F. Xie, H. Shi, and Y. Zhao, "Content-based histopathological image retrieval for whole slide image database using binary codes," in *SPIE Medical Imaging*, 2017, p. 1014013.

[28] Y. Ma, Z. Jiang, H. Zhang, F. Xie, Y. Zheng, H. Shi, and Y. Zhao, "Proposing regions from histopathological whole slide image for retrieval using selective search," in *IEEE International Symposium of Biomedical imaging*, 2017.

[29] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, and Y. Zhao, "Size-scalable content-based histopathological image retrieval from database that consists of wsis," *IEEE Journal of Biomedical and Health Informatics*, Epub 2017 Jul 04.

[30] J. N. Kather, A. Marx, C. C. Reyes-Aldasoro, L. R. Schad, F. G. Zoellner, and C.-A. Weis, "Continuous representation of tumor microvessel density and detection of angiogenic hotspots in histological whole-slide images," *Oncotarget*, vol. 6, no. 22, pp. 19 163–19 176, 2015.

[31] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, and Y. Zhao, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognition*, vol. 71, pp. 14–25, 2017.

[32] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology*, vol. 23, no. 4, pp. 291–299, 2001.

[33] C. C. Reyes-Aldasoro, L. J. Williams, S. Akerman, C. Kanthou, and G. M. Tozer, "An automatic algorithm for the segmentation and morphological analysis of microvessels in immunostained histological tumour sections," *Journal of Microscopy*, vol. 242, no. 3, pp. 262–278, 2011.

[34] J. Chen, Z. Li, and B. Huang, "Linear spectral clustering superpixel," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3317–3330, 2017.

[35] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 60, pp. 91–110, 2004.

[36] F. F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 524–531.

[37] W. Liu, J. Wang, R. Ji, and Y. G. Jiang, "Supervised hashing with kernels," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2074–2081.

[38] A. Coates, A. Y. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *International conference on artificial intelligence and statistics*, 2011, pp. 215–223.

[39] J. Wang, S. Kumar, and S. F. Chang, "Semi-supervised hashing for scalable image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3424–3431.

[40] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval." in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2916–2929.

[41] M. A. Carreira-Perpiñán and R. Raziperchikolaei, "Hashing with binary autoencoders," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 557–566.