Fusing Heterogeneous Features From Stacked Sparse Autoencoder for Histopathological Image Analysis

Xiaofan Zhang, *Student Member, IEEE*, Hang Dou, Tao Ju, Jun Xu, and Shaoting Zhang, *Member, IEEE*

***Abstract*—In the analysis of histopathological images, both holistic (e.g., architecture features) and local appearance features demonstrate excellent performance, while their accuracy may vary dramatically when providing different inputs. This motivates us to investigate how to fuse results from these features to enhance the accuracy. Particularly, we employ content-based image retrieval approaches to discover morphologically relevant images for image-guided diagnosis, using holistic and local features, both of which are generated from the cell detection results by a stacked sparse autoencoder. Because of the dramatically different characteristics and representations of these heterogeneous features (i.e., holistic and local), their results may not agree with each other, causing difficulties for traditional fusion methods. In this paper, we employ a graph-based query-specific fusion approach where multiple retrieval results (i.e., rank lists) are integrated and reordered based on a fused graph. The proposed method is capable of combining the strengths of local or holistic features adaptively for different inputs. We evaluate our method on a challenging clinical problem, i.e., histopathological image-guided diagnosis of intraductal breast lesions, and it achieves 91.67% classification accuracy on 120 breast tissue images from 40 patients.**

***Index Terms*—Breast lesion, feature fusion, histopathological image analysis, large-scale image retrieval, stacked sparse autoencoder (SSAE).**

I. Introduction

RECENTLY, digitized tissue histopathology for microscopic examination and automatic disease grading has become amenable to the application of computerized image analysis and computer-aided diagnosis [1]. Many methods have been proposed to tackle this important and challenging use case, by investigating object level and spatially related features [2]–[4] and employing learning-based classifiers [5]–[7] or content-based image retrieval (CBIR) [8]–[12]. In general, accurate analysis of histopathological images requires to examine cell-level information for accurate diagnosis, including individual cells (e.g., appearance [4], [13] and shapes [14]) and architecture of tissue (e.g., topology and layout of all cells [2]). These features cover both local and holistic information, all benefiting the diagnosis accuracy of histopathological images. Therefore, the complementary descriptive capability of local and holistic features naturally raises the question of how to integrate their strengths to yield more satisfactory results when classifying histopathological images. However, the characteristics, algorithmic procedures and representations of these features can be dramatically different, making them nontrivial to fuse. For example, architecture features [2] are represented as a low-dimensional vector of statistics, while local feature can be represented as high-dimensional bag-of-words (BoW) [15] and compressed as binary codes to improve the efficiency [4], [16], [17].

Generally, fusion can be carried out on the feature or rank-levels. In our context (i.e., differentiation of cancers), this means to combine different types of features in a histogram [18], [19] for learning-based classification, or to fuse the ordered results from CBIR methods [20], [21], and then, classify via majority voting, both of which are fundamental problems. Unfortunately, many existing fusion methods still have limitations for medical image analysis, especially in terms of the robustness, scalability, and generality. For example, feature-level fusion usually concatenates multiple feature vectors (e.g., histogram of color features or texture features) and produces a new feature vector that has a higher dimensionality. However, when these features are heterogeneous (e.g., having significantly different dimensions and characteristics such as low-dimensional architecture feature [2] and high-dimensional appearance feature [4] in histopathological image analysis), feature-level fusion may not be able to effectively integrate their strengths. On the other hand, rank-level fusion combines different retrieval results (i.e., a list of retrieved images), obtained from using different types of features. This approach usually needs to decide which features should have an important role in the retrieval, which is quite difficult to determine online for a specific input with a large database.

In this paper, we focus on the rank-level fusion of local and holistic features for the image-guided diagnosis of breast cancer, i.e., differentiation of the benign (the usual ductal hyperplasia) and actionable (the atypical ductal hyperplasia and ductal carcinoma in situ) cases. Particularly, we employ content-based image retrieval to discover clinically relevant instances from an image database, which can be used to infer and classify the new image. Given image ranks (i.e., retrieval results) obtained from holistic and local features, generated based on a stacked sparse autoencoder (SSAE), a data-driven and

Manuscript received June 29, 2015; accepted July 26, 2015. Date of publication July 29, 2015; date of current version September 1, 2016. This work was supported in part by Charlotte Research Institute (CRI), in part by Oak Ridge Associated Universities for the Ralph E. Powe Junior Faculty Enhancement Award, in part by the National Natural Science Foundation of China under Grant 61273259, and in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20141482.

X. Zhang and S. Zhang are with the Department of Computer Science, University of North Carolina, Charlotte, NC 28027 USA (e-mail: szhang16@uncc.edu; shaoting@cs.rutgers.edu).

H. Dou and T. Ju are with the Department of Computer Science and Engineering, Washington University, St. Louis, MO 63130 USA (e-mail: hangdou@gmail.com; taoju@cse.wustl.edu).

J. Xu is with Jiangsu Key Laboratory of Big Data Analysis Technique, Nanjing University of Information Science & Technology, Nanjing 210044, China (e-mail: xujung@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2015.2461671

graph-based method [21] is employed for accurate, robust, and efficient fusion, by evaluating the quality of each rank online. Although this graph fusion method was originally designed for natural images, it also provides an effective solution for the fusion of heterogeneous information in the domain of histopathological image analysis. We validate our framework on 120 breast tissue images from 40 patients. The experimental results demonstrate the accuracy and efficiency of our framework. A preliminary version of this study is published in [22]. In this paper, we provided comprehensive reviews, added details of our cell detection module, and conducted additional evaluations to demonstrate the efficacy of our method.

The rest of this paper is organized as follows. Section II reviews relevant work of histopathological image analysis and feature fusion. Section III presents our framework for real-time cell examination. Section IV shows the experimental results on lung microscopic tissue images. Concluding remarks are given in Section V.

II. Related Work

A. Features for Histopathological Image Analysis

There are three main categories of features for histopathological image analysis.

1) Morphology Features: Dundar *et al.* [14] proposed to segment cells using a Gaussian mixture model and a watershed algorithm, and describe individual cells by their size, shape, and nucleoli. After that, multiple-instance learning (MIL) with a support vector machine (SVM) is used to identify the stage of breast lesion. Petushi *et al.* [23] employed adaptive thresholding and morphological operations to segment cells and calculate high-density areas of different types of nuclei as features. These cells are then classified with linear discriminant analysis and forward/backward search methods.

2) Graph-Based Features: The second category includes CAD systems that use nucleus detection followed by graph-based image characterization. Basavanhally *et al.* [2] proposed to detect locations of nuclei using a combination of region growing and Markov random fields. Three graphs (i.e., Voronoi diagram, Delaunay triangulation, and minimum spanning tree) are constructed to describe the arrangement of cells. An SVM is then employed to classify the high or low presence of lymphocytic infiltration that can be used to evaluate phenotypic changes in breast cancer. There are other graph-related features for histopathological image analysis. Bilgin *et al.* [24] designed simple cell graphs, probabilistic cell graphs, and hierarchical cell graphs from the image segmentation results. Then, they calculated a set of global metrics from these cell graphs for breast tissue modeling and classification. Demir *et al.* [25] proposed a computational method based on augmented cell graphs from magnification tissue images for automated cancer diagnosis. Bilgin *et al.* [24] proposed extracellular-matrix-aware cell-graph mining for bone tissue modeling and classification, by precisely modeling the structure-function relationships.

3) Texture Features: Texture features such as SIFT [13], wavelet [3], and filter banks [26] are also employed to model phenotypic appearance in histopathological images. Caicedo *et al.* [13] proposed to detect key points and extract local descriptors with SIFT. These features are then used to obtain a BoW [27] that were then classified using an SVM with kernel functions. Similar features have been employed for large-scale image retrieval [4].

These features are either holistic or local, depicting histopathological images at different scales.

B. Feature Fusion to Classify Histopathological Images

In addition to using a single type of feature to classify histopathological images, multiple features can be combined for more comprehensive information. Feature fusion aims to combine the strengths of complementary cues such as local and holistic features. There are several attempts combining such cues either at the feature or rank level, both for natural images and medical images. Feature-level fusion usually cannot preserve the efficiency and scalability. For examples, Tabesh *et al.* [5] aggregated color, texture, and morphometric cues at the global and histological object levels for classification. Specifically, features representing different visual cues were combined in a supervised learning framework, i.e., feature selection framework that reduces the computational cost and improves the performance. Since the supervised learning is employed, this fusion process depends on the specific database, i.e., need to train different parameters for different datasets. Doyle *et al.* [7] graded the breast cancer with both graph-based and texture features. Spectral clustering is employed to reduce the dimensionality of multiple features that are used for classification. Four features are investigated, including Voronoi diagram, Delaunay triangulation, minimum spanning tree, and nuclear features, most of which represent the holistic information of cell layouts, while complementary features can be heterogeneous.

C. CBIR

Retrieval-based techniques have been widely used in different domains [28]–[30]. Particularly, CBIR has been investigated for medical image analysis such as computer-aided diagnosis [8], [9]. Since our framework fuses heterogeneous features based on CBIR, we also briefly review some relevant papers, especially in this area of histopathological image analysis. For example, Comaniciu *et al.* [31] proposed a CBIR system to support clinical decisions in pathology. Specifically, a fast color segmenter is used to extract multiple types of features such as shape, area, and texture of the nucleus. The classification accuracy was evaluated by comparing with human expert annotation on a database containing 261 digitized specimens. Song *et al.* [32] designed a hierarchical spatial matching-based image-retrieval method using spatial pyramid matching to effectively extract and represent the spatial context of pathological tissues. Schnorrenberg *et al.* [33] extended the biopsy analysis support system to include indexing and CBIR of biopsy slide images, which was tested on 57 breast-cancer cases. Zheng *et al.* [34] designed a CBIR system to retrieve images and their associated annotations from a networked microscopic pathology image database. Akakin *et al.* [35] used the multitiered approach to classify and retrieve microscopic images, enabling both multiimage query

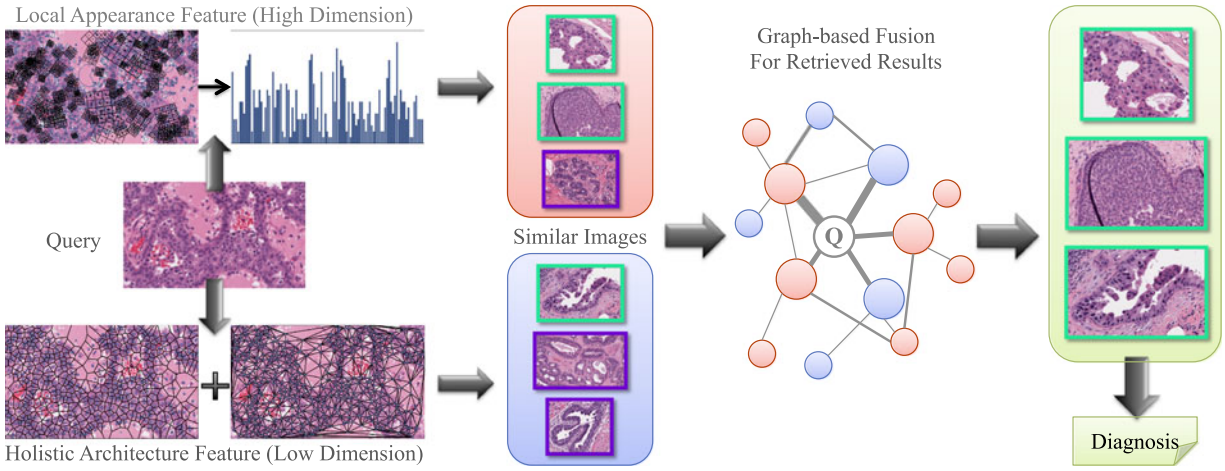

Fig. 1. Overview of the proposed framework. Both holistic architecture feature and local appearance feature are extracted and employed for image retrieval. The retrieval results are fused via the graph-based framework to improve the accuracy. Note that majority voting does not work in this example, since two ranks have no intersection.

and slide-level image retrieval. It is able to protect the semantic consistency among the retrieved images. Foran *et al.* [10] designed a scalable CBIR system for comparative analysis of tissue microarrays via high-performance computing and grid technology. Our focus is to improve traditional CBIR methods by fusing heterogeneous features.

III. Methodology

A. Overview

Fig. 1 shows the overview of our framework. From detected cells via an SSAE, we extract both holistic architecture features [2] and high-dimensional local appearance features [4], [36] (i.e., 10 000 dimensions), both of which are used for image retrieval. To ensure the computational efficiency and scalability, the high-dimensional feature is compressed as tens of hash bits [4], [16]. Combining these complementary features is an intuitive approach to improve the accuracy. However, directly combining them at the feature level may not be effective due to dramatically different representations. An alternative is to fuse them at the rank level, i.e., retrieved images. The critical issue is how to measure and compare the quality of ranks on the fly, since fusion process should favor the rank with higher quality. As the similarity scores of retrieved results may vary largely among queries and are not comparable between different ranks, a reasonable approach is to measure the *consistency* among the top candidates. Therefore, for each query image, we construct a weighted undirected graph from the retrieval results of one rank, where the retrieval quality or the relevance is modeled by the weights on the edges [21]. These weights are determined by the overlap ratio (i.e., Jaccard similarity coefficient) of two neighborhood image sets. Here, neighborhood set means the nearest neighbors of the input image, measured by the feature similarity. Then we fuse multiple graphs to one and perform a localized PageRank algorithm [37] to rerank the retrieval results according to their probability distribution. As a result, the fused retrieval results tend to be consistent among different feature representations.

B. SSAE for Automated Nuclei Detection

Since both of our holistic and local features are extracted from detected cells, we briefly introduce our cell detection method. We employ learning-based methods to locate cells, and manually fix a small portion of erroneous detections. Specifically, our detector is based on the SSAE [38]. An autoencoder is a multilayer feed-forward neural network trained to represent the input with backpropagation [39]. By applying backpropagation, the autoencoder tries to decrease the discrepancy as much as possible between input and reconstruction by learning a encoder and a decoder, which yields a set of weights W and biases b. The cost function of a sparse autoencoder (SAE) comprises three terms as follows:

$$\mathfrak{L}_{\text{SAE}}(\theta) = \frac{1}{N}\sum_{k=1}^{N}(L(x(k), d_{\hat{\theta}}(e_{\check{\theta}}(x(k))))) + \alpha\sum_{j=1}^{n} KL(\rho||\hat{\rho}_j) + \beta\|W\|_2^2 \quad (1)$$

where the first term is an average sum-of-squares error term describing the discrepancy between input $x(k)$ and reconstruction $\hat{x}(k)$ over the entire data X. Encoder $e_{\check{\theta}}(\cdot)$ maps the input $x \in R^{d_x}$ to the hidden representation $h \in R^{d_h}$, while decoder $d_{\hat{\theta}}(\cdot)$ maps the resulting hidden representation h back into the input space $\hat{x}$. Here, N is the number of entire training patches. $x(k)$ is the vector of intensity for kth patch. $d_{\hat{\theta}}(\cdot)$ and $e_{\check{\theta}}(\cdot)$ are decoder and encoder networks with parameters $\hat{\theta}$ and $\check{\theta}$, respectively. The second term is the Kullback–Leibler (KL) divergence between $\hat{\rho}_j$, the average activation (averaged over the training set) of the hidden unit j, and desired activations ρ. The third term is a weight decay term, which tends to decrease the magnitude of the weight, and helps prevent overfitting.

Our SSAE consists of two basic SAE in which the output of first layer is wired to the input of the second layer. The SSAE is trained with nuclei and nonnuclei patches from training set. The training set includes 14421 nuclei and 28032 nonnuclei patches

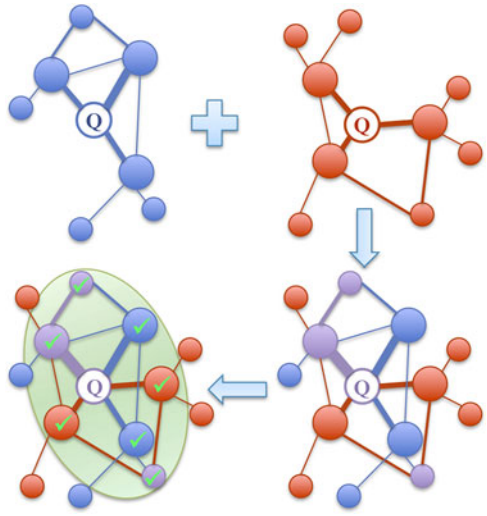

Fig. 2. Procedures of our graph fusion, including graph construction (from two ranks, represented as blue and red graphs), graph consolidation (purple to represent nodes appearing in both graphs), and subgraph selection.

(34×34 pixels for each patch) that randomly extracted from histopathological images.

After trained with nuclei and nonnuclei patches, the SSAE is employed to detect nuclei from histological images. In detection phase, a sliding window approach is used to select candidate patches, which is subsequently fed to trained an SSAE model for classifying the candidate patches being nuclei or nonnuclei patches. The detector recognizes a well-centered nuclei in its input field and rejects images containing no centered nuclei. Both architecture feature and local appearance feature are extracted from detection results.

C. Fusion of Heterogeneous Features

Based on the architecture feature and local appearance feature, *k*-nearest neighbors (kNN) algorithm can be naturally used to find similar cases of the input image. Since each feature can generate one set of results, i.e., a rank list, we conduct rank-level fusion for such heterogeneous features. This procedure includes graph construction, graph consolidation, and subgraph selection, as shown in Fig. 2.

1) Graph Construction: Given a list of ranked results (i.e., retrieved images) by one type of features, such as the architecture or appearance feature, we assume that the consensus degree among the top candidates reveals the retrieval quality. Therefore, we first build a weighted graph using the constraints derived from the consensus degree, i.e., shared kNN. Setting the query as the graph centroid, we use its kNN as the first layer of nodes in the graph, and kNN of kNN as the second layer. Note that this setting is different from traditional methods using reciprocal kNN [21], [40], since such information is usually not available for medical image analysis, i.e., query is not included in the database. Neighboring nodes are connected by edges, whose weight can be defined as the ratio of their common neighbors, i.e., Jaccard similarity, which reflects the confidence of including the connected nodes into the retrieval results. The weight between node i and i' is defined as

$$w(i,i') = J(i,i') = \frac{|N_k(i) \cap N_k(i')|}{|N_k(i) \cup N_k(i')|} \tag{2}$$

where $|\cdot|$ denotes the cardinality, $N_k(i)$ and $N_k(i')$ include the images that are the top-k retrieved candidates using i and i' as the query, respectively. The range of edge weights is from 0 to 1, with $J(i,i') = 1$ implying that these two histopathological images share exactly the same set of neighbors, in which case we assume that they are highly likely to be similar.

2) Graph Consolidation: Multiple graphs, denoted as $G^m = (V^m, E^m, w^m)$, are constructed from the retrieved results of holistic and local features. They can be fused together in a natural way, by appending new nodes or consolidating edge weights of existing nodes in the resulting graph

$$\begin{aligned} G = (V,E,w), \text{with } V &= \cup_m V^m, E = \cup_m E^m, \\ \text{and } w(i,i') &= \sum_m w^m(i,i') \end{aligned} \tag{3}$$

where $w^m(i,i') = 0$ for $(i,i') \notin E^m$. The rational of this fusion process is that though the rank lists or the similarity scores in different methods or features are not directly comparable, their Jaccard coefficients are comparable as they reflect the consistency of two nearest neighborhoods. In other words, this measure of consensus degree does not rely on the similarity scores, so it can be used and compared for different retrieval results from holistic and local features, ensuring the generality.

3) Subgraph Selection: After the candidates from both holistic and local features are fused via the graph consolidation, we need to rank them as per the relevance and select the most similar ones. This can be achieved by conducting a link analysis on the resulting graph, which is treated as a network. This is, therefore, equivalent to the PageRank problem [37] that discovers the probabilities of the nodes to be visited. Since this network is built by considering the retrieval relevance, naturally a node is more important or relevant if it has a higher probability to be visited. To compute the equilibrium state of the graph, we define the $|V| \times |V|$ transition matrix $\mathbf{P}$ as $P_{ii'} = w(i,i')/\deg(i)$ for $(i,i') \in E$, and 0 otherwise, where $\deg(i)$ means the degree or the number of neighbors for a specific node i. This matrix is row stochastic, and the summation of each row equals to one. In the *intelligent surfer model* [41], a "surfer" probabilistically moves along the edges of G to different nodes, based on the transition matrix $\mathbf{P}$. We denote p_i^t as the probability for the surfer to be at node i at a time t and $p^t = (p_i^t)$. The equilibrium state of p is obtained by the query-dependent PageRank vector as a stationary point using the power method, indicating the relevance or similarity to the query image.

Once p has converged, the histopathological images are ranked according to their probabilities in p, where a higher probability reflects a higher relevance to the query in this equilibrium state of the graph. Using fused results, i.e., a new list of histopathological images from both features, majority voting can be employed for cancer differentiation. To summarize, fusing heterogeneous features via graphs can significantly improve

the performance of each individual feature, without sacrificing the scalability and generality.

IV. Experiments

A. Experimental Setting

Histopathological images of breast tissue for this study were collected from the IU Health Pathology Lab (IUHPL) according to the protocol approved by the Institutional Review Board [14]. All the slides were imaged using a ScanScope digitizer (Aperio, Vista, CA, USA) available in the tissue archival service at IUHPL. 120 images (around 2250-K pixels for each image) were gathered from 40 patients, 3 images per patient. 20 of these patients were labeled as benign and others are actionable, based on the majority diagnosis of nine board-certified pathologists. Leave-one-patient-out validation is used to evaluate the accuracy of classification. All parameters are tuned using cross validation to optimize the final result. The experiments were conducted on a 3.40-GHz CPU with 4 cores and 16-G RAM, in a MATLAB implementation.

B. Evaluation of Individual Features

We employ two types of features, holistic and local, as the baseline methods for fusion. For holistic feature [2], the Voronoi diagram, Delaunay triangulation, minimum spanning tree are constructed and the nuclear density features are computed to model "architecture" of breast tissue, resulting in a 48-dimensional feature vector for each image. For local feature, 1500–2000 SIFT descriptors [36] are extracted from each image by detecting key points to describe the cell appearance. These descriptors are quantized into sets of cluster centers using BoW [15], in which the feature dimension equals the number of clusters. Specifically, we quantize them into high-dimensional feature vectors with length 10 000. For efficiency and scalability, we compress the high-dimensional feature into 48 binary bits with the kernelized supervised hashing (KSH) algorithm [16]. Note that this binary representation is not compatible with the holistic feature. We first evaluate the performance of image retrieval using single feature such as the holistic feature, high-dimensional local feature and compressed binary feature. kNN and an SVM are used as the baselines that have been widely employed for histopathological image analysis [2], [5], [6].

As shown in Fig. 3, both holistic and local features are able to generate reasonable results, i.e., around 80% accuracy. The only exception is that kNN fails in handling high-dimensional local feature, achieving only 74.17% accuracy. After compression with KSH, the binary codes improve the accuracy to 81.67%. In addition, using hashing representation also significantly improves the computational efficiency, i.e., thousands times faster than using original high-dimensional features, ensuring the scalability. Since both features are fairly effective but not perfect, and they should be complementary as they model different scales of information, it is natural to combine them for higher accuracy.

C. Evaluation of Feature Fusion

We compare our fusion framework with several classical methods for fusion, including both feature and rank-level approaches. For feature-level fusion, we normalize and concatenate different features into a histogram [19] and classify them with either kNN or SVM. Since the dimensions of features are largely different, it is not likely to obtain reasonable results without doing normalization. Therefore, normalization ensures that each feature contributes "equally" to the concatenated one [42]. For rank-level fusion, we combine different retrieval results via rank aggregation [20] and classify the query image with majority voting. Rank aggregation has been employed to fuse image retrieval results from similar types of features [43].

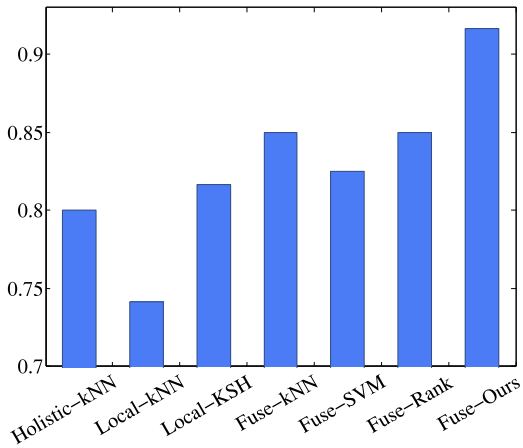

Fig. 3. Quantitative comparison of the classification accuracy. We compare the performance of each single feature, and the fusion of both holistic and local features.

As shown in Fig. 3, concatenation of feature vectors marginally improve the classification accuracy, i.e., around 1–3% better than the baseline, due to the dramatically different characteristics of heterogeneous features. On the other hand, rank aggregation also merely improves the accuracy by 3%, since there may be no intersection among the top candidates retrieved by the local and holistic features. Our graph fusion method determines online which features should play a major role in the retrieval, in an unsupervised scheme. As a result, our fusion of heterogeneous features significantly improves the accuracy by around 10%, i.e., achieving 91.67% overall accuracy on this challenging problem. In addition, since this fusion process is applied on the retrieved results, i.e., a small subset of the whole dataset, it is very efficient and only takes milliseconds, ensuring promising scalability. Fig. 4 shows some retrieval results using our framework.

D. Discussions

In this section, we discuss the parameters and implementation issues of our system. Our fusion method only has one important parameter, i.e., k for constructing the graphs. As shown in Fig. 5, the accuracy is related to this parameter. For example, choosing a very small value for k (e.g., 3) indicates strong constraints of including nodes in the graph. Therefore, the resulting graphs usually do not have enough nodes. In other words, our graph fusion and reranking method cannot find enough candidates to select from. On the other hand, choosing a large value for k (e.g., 25) looses the constraints, so the graphs may have many

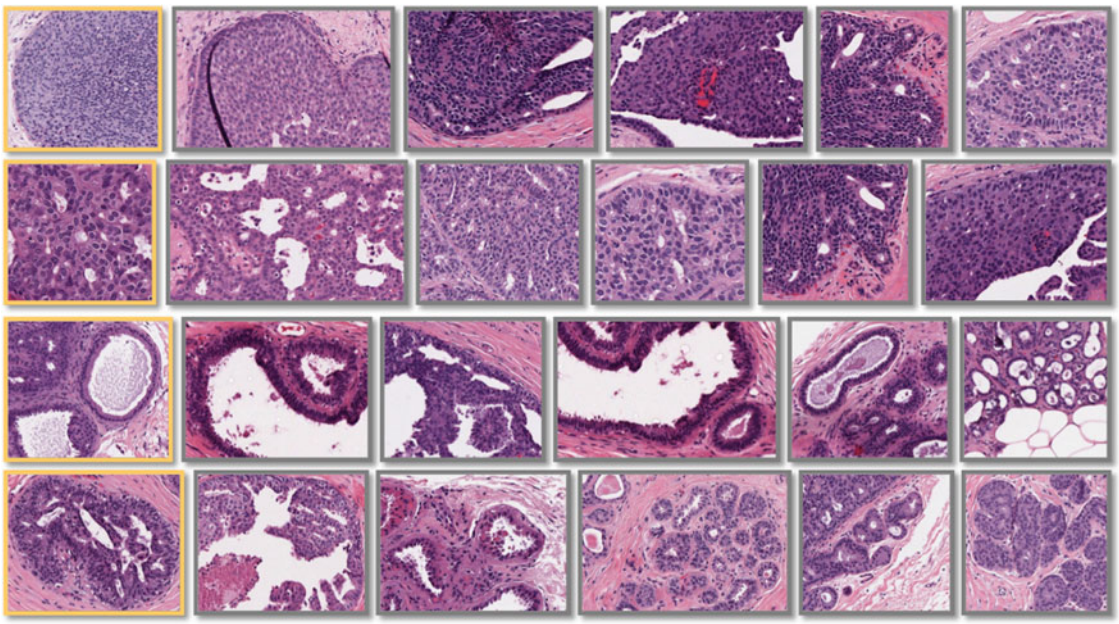

Fig. 4. Retrieval results using our fusion framework. The first image in each row is the query, and the remaining ones are retrieval results. Top two rows are actionable cases, and bottom two rows are benign.

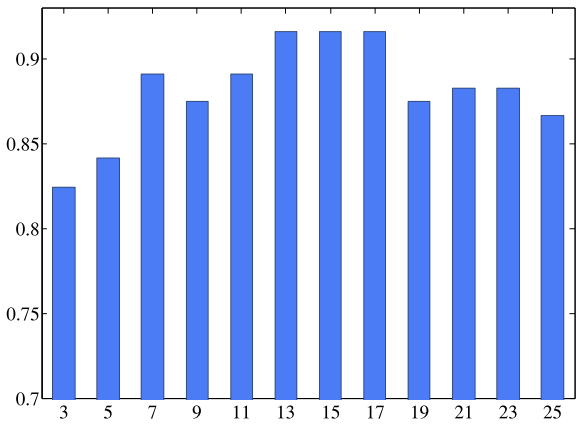

Fig. 5. Evaluation of parameter k when constructing the graphs, ranging from 3 to 25.

nodes that are loosely related with the query. This also adversely affects the accuracy. Therefore, it is desired to choose a proper value of k. The motivation is to have sufficient and related candidates (i.e., nodes in graphs) incorporated into each graph, so our graph fusion algorithm can combine their strengths. Fig. 5 shows that our fusion results are consistently better than each of the baseline, in fact, it can achieve promising results (i.e., more than 90.0%) in a certain range of values, indicating that our method is not sensitive to small variations of k.

In our experiment, graph-based rank-level fusion significantly outperforms feature-level fusion and rank aggregation. However, this is not guaranteed and depends on the properties of features. When two types of features are heterogeneous, their histograms may have dramatically different properties, e.g., sparsity and dimensions. Our method becomes particularly useful, as it adaptively decides the quality of retrieval results on the fly. On the other hand, if these features have similar characteristics, e.g., features from multiple color spaces, they tend to generate similar ranks such that majority voting can be effective. In this case, rank aggregation or concatenation of histograms are able to achieve accurate results for fusion. We have conducted an experiment on fusing subtypes of architecture features. In fact, this 48-dimensional architecture feature is a concatenation of four holistic features, i.e., Voronoi features, Delaunay features, minimum spanning tree features, and Nuclear features, whose accuracy are 69.2%, 70.8%, 77.5%, and 78.3%, respectively. Both feature-level fusion and our graph fusion achieves 80.0% accuracy, indicating that these four features are not heterogeneous.

V. Conclusion

In this paper, we investigate the fusion of heterogeneous features for histopathological image analysis. Specifically, we employ a graph-based framework to fuse the holistic architecture feature and the local appearance feature that are generated from the cell detection results. These features are complementary but have dramatically different characteristics and representations, causing difficulties for traditional fusion methods. Our framework is able to measure online the retrieval quality by the consistency of the neighborhoods of candidate images. Therefore, the fused results significantly improve the baseline using single feature. In the future, we will test our method on larger dataset (e.g., thousands of images) and employ more features for fusion. We are also interested in the relevance feedback that takes domain experts' feedback to improve the retrieval and fusion methods.

REFERENCES

[1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 2, pp. 147–171, Dec. 2009.

[2] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 642–653, Mar. 2010.

[3] B. Weyn, G. van de Wouwer, A. van Daele, P. Scheunders, D. van Dyck, E. van Marck, and W. Jacob, "Automated breast tumor diagnosis and grading based on wavelet chromatin texture description," *Cytometry*, vol. 33, no. 1, pp. 32–40, 1998.

[4] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 496–506, Feb. 2015.

[5] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, Oct. 2007.

[6] O. Sertel, J. Kong, U. V. Catalyurek, G. Lozanski, J. H. Saltz, and M. N. Gurcan, "Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading," *J. Signal Process. Syst.*, vol. 55, no. 1–3, pp. 169–183, 2009.

[7] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *Proc. Int. Symp. Biomed. Imag.*, 2008, pp. 496–499.

[8] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applicationsł clinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1, pp. 1–23, 2004.

[9] H. Müller, A. Geissbühler, and P. Ruch, "ImageCLEF 2004: Combining image and multi-lingual search for medical image retrieval," in *Multilingual Information Access for Text, Speech and Images*. New York, NY, USA: Springer, 2005, pp. 718–727.

[10] D. J. Foran, L. Yang *et al.*, "Imageminer: A software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology," *J. Amer. Med. Informat. Assoc.*, vol. 18, no. 4, pp. 403–415, 2011.

[11] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data," *J. Dig. Imag.*, vol. 26, no. 6, pp. 1025–1039, 2013.

[12] G. Langs, H. Müller, B. H. Menze, and A. Hanbury, "VISCERAL: Towards large data in medical imaging—Challenges and directions," in *MCBR-CDS MICCAI Workshop* (Lecture Notes in Computer Science). New York, NY, USA: Springer, vol. 7723, 2013.

[13] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Artificial Intelligence in Medicine*. New York, NY, USA: Springer, 2009, pp. 126–135.

[14] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, and M. N. Gurcan, "Computerized classification of intraductal breast lesions using histopathological images," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1977–1984, Jul. 2011.

[15] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 1470–1477.

[16] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang, "Supervised hashing with kernels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2074–2081.

[17] X. Zhang, L. Yang, W. Liu, H. Su, and S. Zhang, "Mining histopathological images via composite hashing and online learning," in *Proc. Medical Image Computing and Computer-Assisted Intervention*. New York, NY, USA: Springer, 2014, pp. 479–486.

[18] S. Zhang, J. Huang, H. Li, and D. N. Metaxas, "Automatic image annotation and retrieval using group sparsity," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 3, pp. 838–849, Jun. 2012.

[19] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 221–228.

[20] R. Fagin, R. Kumar, and D. Sivakumar, "Efficient similarity search and classification via rank aggregation," in *Proc. ACM SIGMOD*, 2003, pp. 301–312.

[21] S. Zhang, M. Yang, T. Cour, K. Yu, and D. Metaxas, "Query specific rank fusion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 803–815, Apr. 2015.

[22] X. Zhang, H. Dou, T. Ju, and S. Zhang, "Fusing heterogeneous features for the image-guided diagnosis of intraductal breast lesions," in *Proc. Int. Symp. Biomed. Imag.*, 2015, pp. 1288–1291.

[23] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Med. Imag.*, vol. 6, no. 1, pp. 1–11, 2006.

[24] C. Bilgin, C. Demir, C. Nagi, and B. Yener, "Cell-graph mining for breast tissue modeling and classification," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2007, pp. 5311–5314.

[25] C. Demir, S. H. Gultekin, and B. Yener, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, vol. 21, pp. ii7–ii12, 2005.

[26] L. Yang, W. Chen, P. Meer, G. Salaru, M. D. Feldman, and D. J. Foran, "High throughput analysis of breast cancer specimens on the grid," in *Proc. Med. Image Comput. Comput.-Assisted Intervention Conf.*, 2007, pp. 617–625.

[27] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 1470–1477.

[28] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Visual Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, 1999.

[29] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu, "Visual-textual joint relevance learning for tag-based social image search," *IEEE Trans, Image Process.*, vol. 22, no. 1, pp. 363–376, Jan. 2013.

[30] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4290–4303, Sep. 2012.

[31] D. Comaniciu, P. Meer, and D. J. Foran, "Image-guided decision support system for pathology," *Mach. Vis. Appl.*, vol. 11, no. 4, pp. 213–224, 1999.

[32] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," in *Proc. ACM Int. Workshop Med. Multimedia Anal. Retrieval*, 2011, pp. 1–6.

[33] F. Schnorrenberg, C. Pattichis, C. Schizas, and K. Kyriacou, "Content-based retrieval of breast cancer biopsy slides," *Technol. Health Care*, vol. 8, no. 5, pp. 291–297, 2000.

[34] L. Zheng, A. W. Wetzel, J. Gilbertson, and M. J. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 249–255, Dec. 2003.

[35] H. C. Akakin and M. N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 758–769, Jul. 2012.

[36] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[37] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," 1999.

[38] J. Xu, L. Xiang, R. Hang, and J. Wu, "Stacked sparse autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology," in *Proc. Int. Symp. Biomed. Imag.*, 2014, pp. 999–1002.

[39] A. Ng, "Sparse autoencoder," *CS294A Lecture notes*, p. 72, 2011.

[40] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool, "Hello neighbor: Accurate object retrieval with *k*-reciprocal nearest neighbors," in *Proc. Comput. Vision Pattern Recog.*, 2011, pp. 777–784.

[41] M. Richardson and P. Domingos, "The intelligent surfer: Probabilistic combination of link and content information in pagerank," in *Proc. Neural Inf. Process. Syst.*, 2001, pp. 1441–1448.

[42] A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for image annotation," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 88–105, 2010.

[43] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 2–11, Jan. 2010.

Authors' photographs and biographies not available at the time of publication.