Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images

Can Taylan Sari and Cigdem Gunduz-Demir*, *Member, IEEE*

Abstract—Histopathological examination is today's gold standard for cancer diagnosis. However, this task is time consuming and prone to errors as it requires a detailed visual inspection and interpretation of a pathologist. Digital pathology aims at alleviating these problems by providing computerized methods that quantitatively analyze digitized histopathological tissue images. The performance of these methods mainly rely on features that they use, and thus, their success strictly depends on the ability of these features successfully quantifying the histopathology domain. With this motivation, this paper presents a new unsupervised feature extractor for effective representation and classification of histopathological tissue images. This feature extractor has three main contributions: First, it proposes to identify salient subregions in an image, based on domain-specific prior knowledge, and to quantify the image by employing only the characteristics of these subregions instead of considering the characteristics of all image locations. Second, it introduces *a new deep learning based technique* that quantizes the salient subregions by extracting a set of features directly learned on image data and uses the distribution of these quantizations for image representation and classification. To this end, the proposed deep learning based technique constructs *a deep belief network of restricted Boltzmann machines* (RBMs), defines the activation values of the hidden unit nodes in the final RBM as the features, and learns the quantizations by clustering these features in an unsupervised way. Third, this extractor is the first example of successfully using restricted Boltzmann machines in the domain of histopathological image analysis. Our experiments on microscopic colon tissue images reveal that the proposed feature extractor is effective to obtain more accurate classification results compared to its counterparts.

Index Terms—Deep learning, feature learning, histopathological image representation, digital pathology, automated cancer diagnosis, saliency, colon cancer, hematoxylin-eosin staining.

I. Introduction

IN recent years, deep learning has shown great promise as an alternative to employing handcrafted features in computer vision tasks [1]. Since deep learners are end-to-end unsupervised feature extractors, they neither require nor use domain-specific prior knowledge. Nevertheless, in order for humans to accomplish certain tasks, an insight that could only be gained through a specialized training in the related domain

C. T. Sari is with the Department of Computer Engineering, Bilkent University, Ankara TR-06800, Turkey (e-mail: can.sari@bilkent.edu.tr).

*C. Gunduz-Demir is with the Department of Computer Engineering and Neuroscience Graduate Program, Bilkent University, Ankara TR-06800, Turkey (e-mail: gunduz@cs.bilkent.edu.tr).

is typically required. Therefore, incorporating some amount of domain-specific knowledge into the learning method may prove useful in these tasks.

One such task is the diagnosis and grading of cancer via the examination of histopathological tissues [2]. This procedure normally requires a pathologist who has extensive medical knowledge and training to visually inspect a tissue sample. In this inspection, pathologists do not examine just randomly selected subregions but salient ones located around the important sections of a tissue. They first determine the characteristics of these salient subregions and then properly categorize them to decide whether the sample contains normal or abnormal (cancerous) formations. For a trained pathologist, this categorization and decision process relies on human insight and expert knowledge. However, most of these subregions lack a clear and distinct definition that can directly be used in a supervised classifier, and in the framework of learning, the task of annotating these subregions incurs great cost.

In response to these issues, this paper proposes a novel semi-supervised method for the classification of histopathological tissue images. Our method introduces a new feature extractor that uses prior domain knowledge for the identification of salient subregions and devises an unsupervised method for their characterization. A tissue is visually characterized by the traits of its cytological components, which are determined by the appearance of the components themselves and the subregions in their close proximities. Thus, this new feature extractor first proposes to define the salient subregions around the approximated locations of cytological tissue components. It then pretrains a deep belief network, consisting of consecutive *restricted Boltzmann machines* (RBMs) [3], on these salient subregions, allowing the system to extract high-level features directly from image data. To do so, this unsupervised feature extractor proposes to use the activation values of the hidden unit nodes in the final RBM of the pretrained deep belief network and to feed them into a clustering algorithm for quantizing the salient subregions (their corresponding cytological components) in an unsupervised way. Finally, our method trains a supervised learner on the distribution of the quantized subregions/components in a tissue, which is then used to properly classify a tissue image.

Our proposed method differs from the existing studies in the following aspects. The studies that use deep learning for histopathological image analysis either train a learner on entire images for their classification [4], [5] or crop small patches out of these images, train a learner on the patches and

then use the patch labels for entire image classification [6], [7] but more typically for nucleus detection or entire image segmentation [8], [9], [10], [11], [12], [13]. On the other hand, as opposed to our proposed method, these studies either pick random points in an image as the patch centers, or divide the image into a grid, or use the sliding window approach. None of them identify salient subregions/components and use them to determine their patches. Furthermore, most of them use convolutional neural networks (CNNs) trained in a supervised manner, which requires the labels of all training patches. For that, they label a patch with the type of the segmented region covering this patch (e.g., with either nucleus or background label for nucleus detection) if focus is segmentation. Otherwise, if it is classification, they label a patch with the class of the entire image without paying attention to the local characteristics of its subregions since the latter type of labeling is quite difficult and extremely time-consuming. On the other hand, considering the local characteristics of patches/subregions in a classifier may improve the performance since a tissue contains subregions showing different local characteristics and the distribution of these subregions determines the characteristics of the entire tissue. To the best of our knowledge, there exists only one study that labeled its patches in an unsupervised way, using stacked autoencoders. However, this study did pick its patches randomly and did not consider the saliency in tissue images at all [7]. As opposed to all these previous studies, our method uses salient subregions/components in an image, as determined by prior domain knowledge, and learns how to characterize them in an entirely unsupervised manner without the need for expensive and impractical labeling. These two attributes of our proposed method lead to better results as our experiments have demonstrated. Furthermore, to the best of our knowledge, this study is the first example that successfully uses a deep belief network of RBMs for the characterization of histopathological tissue images.

In our earlier study [14], we also proposed to quantize tissue components in an unsupervised manner through clustering and use the distribution of the cluster labels for image classification. However, completely different than this current work, our earlier study used a set of handcrafted features and did not use any deep learning technique at all. Our experiments have revealed that the use of deep learning features directly learned from image data improves the accuracy of using the handcrafted features.

There are three main contributions of this paper: First, it introduces a new deep learning based unsupervised feature extractor to quantize a subregion of a tissue image. This feature extractor feeds the subregion's pixels to a deep belief network of consecutive RBMs and defines the activation values of the hidden units in the last RBM layer as the deep features of this subregion. Then, it clusters these deep features to learn the quantizations in an unsupervised way. Second, it proposes to characterize the tissue image by first identifying its salient subregions and then using only the quantizations of these subregions. Last, it successfully uses RBMs for feature extraction in the domain of histopathological image analysis.

II. RELATED WORK

Digital pathology systems are becoming important tools as they enable fast and objective analysis of histopathology slides. The digital pathology research has focused on two main problems: classification, which is also the focus of this paper, and segmentation. Up to recent studies, the developed methods rely on defining and extracting handcrafted features from a histopathological image and using these features in the design of a classification or a segmentation algorithm. Among those are textural features, which quantify the spatial arrangement of pixel intensities, and structural features, which quantify that of tissue components. Co-occurrence matrices [15], wavelets [16], and local binary patterns [17] are commonly used to define the textural features. For the definition of the structural features, graphs are constructed on nuclei [18] or multi-typed tissue components [19], [20] and global graph features are extracted. Although they yield promising results in numerous applications, defining expressive handcrafted features may require significant insight on the corresponding application. However, this is not always that trivial and improper feature definitions may greatly lower the algorithm's performance.

In order to define more expressive and more robust features, deep learning based studies have proposed to learn the features directly on image data. For that, the majority of these studies train a CNN classifier in a supervised manner and exploit its output for classification or segmentation. Among these studies, only a few feed an entire tissue image to the trained CNN and use the class label it outputs to directly classify the image [4], [5]. Others divide a tissue image into a grid of patches, feed each patch to the CNN, which is also trained on the same-sized patches, and then use either the class labels or the posteriors generated by this CNN. In [6], the labels are voted to classify the image out of which the patches are cropped. In [12], the patch labels are directly used to segment the tissue image into its epithelial and stromal regions. These patch labels are also employed to extract structural features, which are then used for whole slide classification [21] and gland segmentation [22]. Although they are not histopathological images, a similar approach is followed to differentiate nuclear and background regions in fluorescent microscopy images [23] and nuclear, cytoplasmic, and background regions in cervical images [24].

The posteriors generated by the supervised CNN are commonly used to segment a tissue image into its regions of interest (ROI). To this end, for the class corresponding to the ROI (e.g., nucleus or gland class), a probability map is formed using the patch posteriors. Then, the ROI is segmented by either finding local maxima on this posterior map [8], [9], [11], [25] or thresholding it [26]. This type of approach has also been used to detect cell locations in different types of microscopic images such as live cell [27], fluorescent [28], and zebrafish [29] images. As an alternative, nuclei are located by postprocessing the class labels with techniques such as morphological operations [30] and region growing [31]. In [32], after obtaining a nucleus label map, nuclei's bounding boxes are estimated by training another deep neural network.

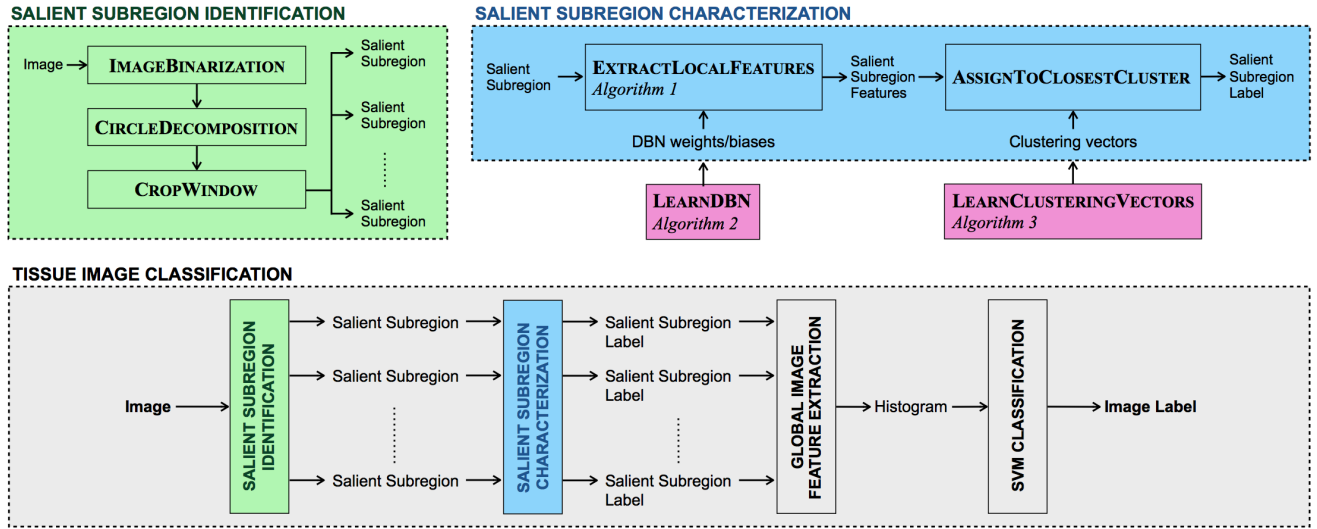There are only a few studies that make use of unsupervised

Fig. 1. A schematic overview of the proposed method.

learning in their systems. In [33], a set of autoencoders are first pretrained on small image patches and the weights of each autoencoder are employed to define a filter for the first convolution layer of a supervised CNN classifier, which is then to be used to classify an entire tissue image. Similarly, in [34], a stacked autoencoder is pretrained on image patches and the outputs of its final layer are fed to a supervised classifier for nucleus detection. As opposed to our proposed method, these previous studies did not cluster the outputs of the autoencoders to label the patches in an unsupervised way and did not use the label distribution for image classification. The study in [7] is similar to our method in the sense that it also clusters the patches based on the outputs of a stacked autoencoder. However, this study did select its patches randomly and did not consider any saliency in a tissue image. On the contrary, our work proposes to determine the salient subregions by prior domain-knowledge, characterize them by an unsupervised deep belief network consisting of consecutive RBMs, and use the characteristics of *only* these salient subregions to classify the entire tissue image. Our experiments have demonstrated that the use of saliency together with this unsupervised characterization improve the accuracy. Additionally, as opposed to all these previous studies, which employ either a CNN or a stacked autoencoder, our study uses a deep belief network of restricted Boltzmann machines.

III. Methodology

Our proposed method relies on representing and classifying a tissue image with a set of features extracted by a newly proposed unsupervised feature extractor. This extractor defines the features by quantifying only the characteristics of the salient subregions in the image instead of considering those of all image locations. To this end, it first proposes to define the salient subregions around cytological tissue components (Sec. III-A). Afterwards, to characterize the subregions/components in an unsupervised way, it learns their local features by a deep belief network consisting of

consecutive RBMs and quantizes them by clustering the local features by the k-means algorithm (Sec. III-B). At the end, it represents and classifies the image with the distribution of its quantized subregions/components (Sec. III-C). A schematic overview of the proposed method is given in Fig. 1 and the details of its steps are explained in the following subsections. The source codes of its implementation are available at http://www.cs.bilkent.edu.tr/~gunduz/downloads/DeepFeature.

The motivation behind this proposed method is the following: A tissue contains different types of cells that serve different functions in the tissue. The visual appearance of a cell and its surrounding may look differently depending on the cell's type and function. Furthermore, some types of cells may form specialized structures in the tissue. The tissue is visually characterized by the traits of all these cytological components. Depending on its type, cancer causes changes in the appearance and distribution of certain cytological tissue components. For example, in colon, epithelial cells line up around a lumen to form a gland structure and different types of connective tissue cells in between the glands support epithelia. In a normal tissue, the epithelial cells are arranged in a single layer and since they are rich in mucin, their cytoplasms appear in light color. With the development of colon adenocarcinoma, this single layer structure is getting disappeared, which causes the epithelial cells' nuclei to be seen as nucleus clutters, and their cytoplasms return to pink as they become poor in mucin. With the further progression of this cancer, the epithelial cells are dispersed in the connective tissue and the regular structure of a gland gets totally lost (see Fig. 2). Some of such visual observations are easy to express, but some others may lack of a clear definition although they are in the eyes of a pathologist. Furthermore, when there exists a clear definition for an observation, its expression and quantification commonly require exact component localization, which emerges a very difficult segmentation problem even for a human eye, and its use in a supervised classifier requires very laborious annotation. Thus, our method approximately represents the tissue components with a set of multi-typed circular objects, defines
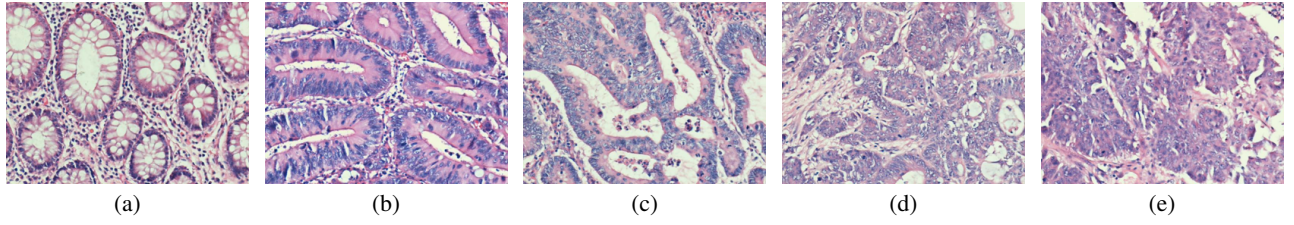
Fig. 2. Example images of tissues labeled with different classes: (a) Normal, (b) low grade cancerous--*grade1*, (c) low grade cancerous--*at the boundary between grade1 and grade2*, (d) low grade cancerous--*grade2*, and (e) high grade cancerous. Note that the normal and high grade cancerous classes are the same for our first and second datasets whereas the low grade cancerous class in the first dataset is further categorized into three in the second one.
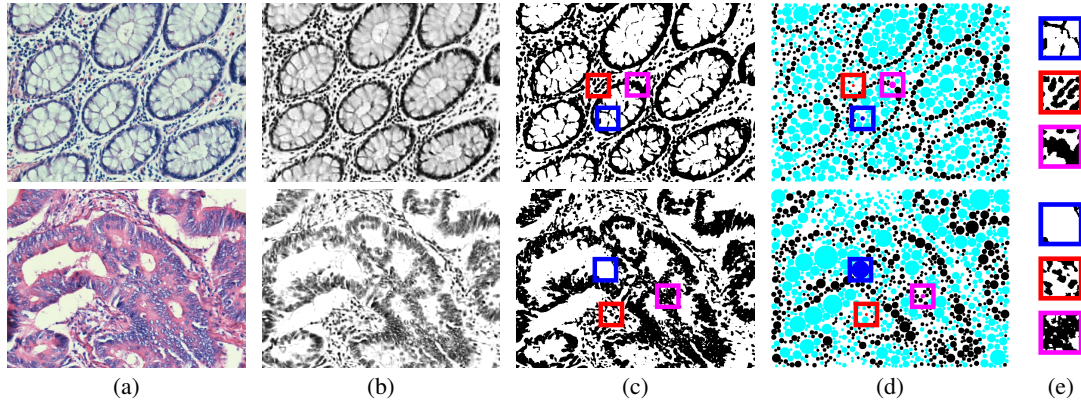

Fig. 3. (a) Original images; top is normal and bottom is cancerous, (b) hematoxylin channels obtained by stain deconvolution, (c) binary images obtained by thresholding, (d) circular objects located by the circle-fit algorithm [36], and (e) examples of salient subregions cropped around the three example located objects. In (d), black and cyan circles represent nuclear and non-nuclear objects, respectively. In (c) and (d), the blue, red, and magenta squares indicate example salient subregions cropped around three example objects, which are also shown in blue, red, and magenta in (c). As seen in the examples given in (e), local properties of small subregions in an image of different types might be similar or different. On the other hand, the distribution of the local properties is different for different types of images.

the local windows cropped around these objects as the salient subregions, and characterizes them in an unsupervised way. Note that this is just an approximate representation where one object can correspond to multiple components or vice versa. It is also worth to noting that the salient subregions cropped around the objects are defined with the aim of approximately representing the components, whose characterizations will further be used in the entire image characterization.

A. Salient Subregion Identification

Salient subregions are defined around tissue components whose locations are approximated by the algorithm that we previously developed in our research group [20]. This approximation and salient subregion identification are illustrated on example images in Fig. 3 and the details are explained below.

The approximation algorithm uses nuclear and non-nuclear types for object representation. For that, it first separates the hematoxylin channel of an image $\mathcal{I}$ by applying color deconvolution [35] and thresholds this channel to obtain the binary image $\mathcal{BW}$. In this thresholding, an average is calculated on all pixel values and a pixel is labeled as nucleus if its value is less than this threshold and non-nucleus otherwise. Then, the circle-fit algorithm [36] is applied on the pixels of each group in $\mathcal{BW}$ separately to locate a set of nuclear and non-nuclear objects. The circle-fit algorithm iteratively locates non-overlapping circles on the given pixels, starting from the largest one as long as the radii of the circles are greater than the threshold r_{min}. At the end, around each object c_i, a salient

region Ω_i is defined by cropping a window out of the binary image $\mathcal{BW}$ where the object centroid determines the window center and the parameter ω_{size} determines its size. Note that although the located objects are labeled with a nuclear or a non-nuclear type by the approximation algorithm, we just use the object centroids to define the salient regions, without using their types. Instead, we will re-type (re-characterize) these objects with the local features that will be learned by a deep belief network (Sec. III-B).

The substeps of this salient subregion identification are herein referred to as IMAGEBINARIZATION, CIRCLEDECOMPOSITION, and CROPWINDOW functions, respectively. We will also use these functions in the implementation of the succeeding steps. To improve the readability of this paper, we provide a list of these functions and their uses in Table I. Note that this table also includes other auxiliary functions, which will be used in the implementation of the succeeding steps.

B. Salient Subregion Characterization via Deep Learning

This step involves two learning systems: The first one, LEARNDBN, acts as an unsupervised feature extractor for the salient subregions, and hence, for the objects that they correspond to. It learns the weights of a deep belief network of RBMs and uses the activation values of the hidden unit nodes in the final RBM to define the local deep features of the salient subregions. The second system, LEARNCLUSTERINGVECTORS, learns the clustering vectors on the local deep features. This clustering will be used to quantize any salient

TABLE I
AUXILARY FUNCTION DEFINITIONS.

Function	Definition
$\mathcal{BW} \leftarrow$ IMAGEBINARIZATION($\mathcal{I}$)	Binarizes the image $\mathcal{I}$ with respect to its hematoxylin channel (see Sec. III-A).
$\mathcal{C} \leftarrow$ CIRCLEDECOMPOSITION($\mathcal{BW}, r_{min}$)	Locates a set $\mathcal{C}$ of circular objects on nuclear and non-nuclear pixels of the binary image $\mathcal{BW}$ with a minimum radius of r_{min} (see Sec. III-A).
$\Omega_i \leftarrow$ CROPWINDOW($\mathcal{BW}, c_i, \omega_{size}$)	Defines a salient subregion Ω_i by cropping a $w_{size} \times w_{size}$ window out of the binary image $\mathcal{BW}$ around an object c_i.
$[W, B] \leftarrow$ CONTRASTIVEDIVERGENCE($\mathcal{D}_{dbn}, P$)	Pretrains a deep belief network on the dataset $\mathcal{D}_{dbn}$ of the salient subregions, by applying the contrastive divergence algorithm to each of its RBMs. The architecture of the deep belief network is denoted by the input parameter P. Returns the weight matrices W and the bias vectors B of the pretrained deep belief network.
$V \leftarrow$ KMEANSCLUSTERING($\mathcal{D}_{kmeans}, K$)	Clusters the dataset $\mathcal{D}_{kmeans}$ of the local deep features of the salient subregions into K using the k-means algorithm and returns the clustering vectors V.
$l_i \leftarrow$ ASSIGNTOCLOSESTCLUSTER(ϕ_i, V)	Labels the salient subregion Ω_i with l_i, which is the id of the closest clustering vector in V, according to the local deep features ϕ_i of this salient subregion.

Algorithm 1 EXTRACTLOCALFEATURES

Input: salient subregion Ω_i, number H of RBMs in the pretrained deep belief network, weight matrices W and bias vectors B of the pretrained deep belief network
Output: local feature set ϕ_i of the salient subregion Ω_i

1: $\Pi_0 = \Omega_i$
2: **for** $j = 1$ to H **do**
3: $\quad \Pi_j = \text{sigmoid}(\Pi_{j-1} \, W_j \, + \, B_j)$
4: **end for**
5: $\phi_i = \Pi_H$

---

subregion, which corresponds to re-typing the object for which this salient subregion is defined. The details of these learning systems are given below.

1) Deep Network Learning: The LEARNDBN algorithm pretrains a deep belief network, which consists of consecutive RBMs. An RBM is an undirected graphical model consisting of a visible and a hidden layer and the symmetric weights in between them. The output of an RBM (the units in its hidden layer) can be considered as a higher representation of its input (the units of its visible layer). To get the representations at different abstraction levels, a set of RBMs are stacked consecutively by linking one RBM's output to the next RBM's input. In this work, the input of the first RBM is fed by the pixels of a salient subregion Ω_i, which is cropped out of the binary image $\mathcal{BW}$, and the output of the last RBM is used as the local feature set ϕ_i of this salient subregion; see Algorithm 1. In this algorithm, W_j and B_j are the weight matrix and the bias vector of the j-th RBM, respectively.

The LEARNDBN function learns the weights and biases of the deep belief network by pretraining it layer by layer using the contrastive divergence algorithm [37]. For this purpose, it constructs a dataset $\mathcal{D}_{dbn}$ from randomly selected salient subregions of randomly selected training images. Algorithm 2 gives its pseudocode; see Table I for explanations of the auxiliary functions. Note that LEARNDBN should also input the parameters that specify the architecture of the network, including the number of hidden layers (the number of RBMs) and the number of hidden units in each hidden layer.

2) Cluster Learning: After learning the weights and biases of the deep belief network, the EXTRACTLOCALFEATURES

Algorithm 2 LEARNDBN

Input: training set $\mathcal{D}$ of original images, size ω_{size} of a salient subregion, minimum circle radius r_{min}, architecture P of the deep belief network
Output: weight matrices W and bias vectors B of the pretrained deep belief network

1: $\mathcal{D}_{dbn} = \emptyset$
2: **for** each randomly selected $\mathcal{I} \in \mathcal{D}$ **do**
3: $\quad \mathcal{BW} \leftarrow$ IMAGEBINARIZATION($\mathcal{I}$)
4: $\quad \mathcal{C} \leftarrow$ CIRCLEDECOMPOSITION($\mathcal{BW}, r_{min}$)
5: $\quad$ **for** each randomly selected $c_i \in \mathcal{C}$ **do**
6: $\quad\quad \Omega_i \leftarrow$ CROPWINDOW($\mathcal{BW}, c_i, \omega_{size}$)
7: $\quad\quad \mathcal{D}_{dbn} = \mathcal{D}_{dbn} \, \cup \, \Omega_i$
8: $\quad$ **end for**
9: **end for**
10: $[W, B] \leftarrow$ CONTRASTIVEDIVERGENCE($\mathcal{D}_{dbn}, P$)

---

function is used to define the local deep features of a given salient subregion. This work proposes to quantify the entire tissue image with the labels (characteristics) of its salient subregions. Thus, these continuous features are quantized into discrete labels. As discussed before, annotating each salient subregion is quite difficult, if not impossible, and hence, it is very hard to learn these labels in a supervised manner. Therefore, this work proposes to follow an unsupervised approach to learn this labeling process. To this end, it uses k-means clustering on the local deep features of the salient subregions. Note that the k-means algorithm learns the clustering vectors V on the training set $\mathcal{D}_{kmeans}$ that is formed up of the local deep features of randomly selected salient subregions of randomly selected training images. The pseudocode of LEARNCLUSTERINGVECTORS is given in Algorithm 3. This algorithm outputs a set V of K clustering vectors. In the next step, an arbitrary salient subregion is labeled with the id of its closest clustering vector.

C. Image Representation and Classification

In the last step, a set of global features are extracted to represent an arbitrary image $\mathcal{I}$. To this end, all salient subregions are identified within this image and their local deep features are

---

Algorithm 3 LEARNCLUSTERINGVECTORS

Input: training set $\mathcal{D}$ of original images, size ω_{size} of a salient subregion, minimum circle radius r_{min}, number H of RBMs, weight matrices W and bias vectors B of the pretrained deep belief network, cluster number K

Output: clustering vectors V

1: $\mathcal{D}_{kmeans} = \emptyset$
2: **for** each randomly selected $\mathcal{I} \in \mathcal{D}$ **do**
3:    $\mathcal{BW} \leftarrow$ IMAGEBINARIZATION$(\mathcal{I})$
4:    $\mathcal{C} \leftarrow$ CIRCLEDECOMPOSITION$(\mathcal{BW}, r_{min})$
5:    **for** each randomly selected $c_i \in \mathcal{C}$ **do**
6:       $\Omega_i \leftarrow$ CROPWINDOW$(\mathcal{BW}, c_i, \omega_{size})$
7:       $\phi_i \leftarrow$ EXTRACTLOCALFEATURES(Ω_i, H, W, B)
8:       $\mathcal{D}_{kmeans} = \mathcal{D}_{kmeans} \cup \phi_i$
9:    **end for**
10: **end for**
11: $V \leftarrow$ KMEANSCLUSTERING$(\mathcal{D}_{kmeans}, K)$

---

extracted. Each salient subregion Ω_i is labeled with the id l_i of its closest clustering vector according to its deep features ϕ_i by the ASSIGNTOCLOSESTCLUSTER auxiliary function (see Table I). Then, to represent the image $\mathcal{I}$, global features are extracted by calculating a histogram on the labels of all salient subregions in $\mathcal{I}$ (i.e., the characteristics of the components that these subregions correspond to). At the end, the image $\mathcal{I}$ is classified by a support vector machine (SVM) with a linear kernel based on its global features. Note that, this study uses the SVM implementation of [38], which employs the one-against-one strategy for multiclass classifications.

IV. EXPERIMENTS

A. Datasets

We test our proposed method on two datasets that contain microscopic images of colon tissues stained with the routinely used hematoxylin-and-eosin technique. The images of these tissues were taken using a Nikon Coolscope Digital Microscope with a $20\times$ objective lens and the image resolution was 480×640. The first dataset is the one that we also used in our previous studies. In this dataset, each image is labeled with one of the three classes: normal, low-grade cancerous, and high-grade cancerous. It comprises 3236 images taken from 258 patients, which were randomly divided into two to form the training and test sets. The training set includes 1644 images (510 normal, 859 low-grade cancerous, and 275 high-grade cancerous) of the 129 patients. The test set includes 1592 images (491 normal, 844 low-grade cancerous, and 257 high-grade cancerous) of the remaining patients. Note that the training and test sets are independent at the patient level; i.e., the images taken from a slide(s) of a particular patient are used either in the training or the test set.

The second dataset includes a subset of the first one with the low-grade cancerous tissue images being further subcategorized. Here only a subset was selected since subcategorization was difficult for some images. Note that we also excluded some images from the normal and high-grade

cancerous classes to obtain more balanced datasets. As a result, in this second dataset, each image is labeled with one of the five classes: normal, low-grade cancerous (grade1), low-grade cancerous (grade2), low-grade cancerous (at the boundary between grade1 and grade2), and high-grade cancerous. The training set includes 182 normal, 188 grade1 cancerous, 121 grade1-2 cancerous, 123 grade2 cancerous, and 177 high-grade cancerous tissue images. The test set includes 178 normal, 179 grade1 cancerous, 117 grade1-2 cancerous, 124 grade2 cancerous, and 185 high-grade cancerous tissue images. Example images from these datasets are given in Fig. 2.

B. Parameter Setting

The proposed method has the following model parameters that should be externally set: minimum circle radius r_{min}, size of a salient subregion ω_{size}, and cluster number K. The parameters r_{min} and ω_{size} are in pixels. Additionally, the support vector machine classifier has the parameter C. In our experiments, the values of these parameters are selected using cross-validation on the training images of the first dataset without using any of its test samples. Moreover, this selection does not consider any performance metric obtained on the second dataset. By considering any combinations of the following values $r_{min} = \{3, 4, 5\}$, $\omega_{size} = \{19, 29, 39\}$, $K = \{500, 1000, 1500\}$, and $C = \{1, 5, 10, 25, 50, 100, 250, 500, 1000, 2500, 5000, 10000\}$, the parameters are set to $r_{min} = 4$, $\omega_{size} = 29$, $K = 1500$, and $C = 500$. In Sec. IV-D, we will discuss the effects of this parameter selection to the method's performance in detail.

In addition to these parameters, one should select the architecture of the deep belief network. In this work, we fix this architecture. In general, the number of hidden layers determines the abstraction levels represented in the network. We set this number to four. We then select the number of hidden units as 2000, 1000, 500, and 100 from bottom to top layers, having the following considerations. For our work, the hidden unit number in the first layer should be selected large enough to effectively represent the pixels in a local subregion. On the other hand, the number in the last layer should be selected small enough to effectively quantize the subregions. The hidden unit numbers in between should be selected consistent to the selected hidden unit numbers in the first and last layers. The investigation of using different network architectures is considered as future work.

C. Results

Tables II and III report the test set accuracies obtained by our proposed *DeepFeature* method for the first and second datasets, respectively. These tables provide the class-based accuracies in their first three/five columns and the average class-based accuracies in the last two. These tables report the average class-based accuracies instead of the overall test set accuracy since especially the first dataset has an unbalanced class distribution. Here we provide the arithmetic mean of the class-based accuracies as well as their harmonic mean since the arithmetic mean can sometimes be misleading when values

TABLE II
TEST SET ACCURACIES OF THE PROPOSED *DeepFeature* METHOD AND THE COMPARISON ALGORITHMS FOR THE FIRST DATASET.

	Norm.	Low	High	Arith. mean	Harm. mean
DeepFeature	98.37	91.59	**98.44**	**96.13**	**96.02**
Handcrafted features					
CooccurrenceMatrix	87.58	84.12	85.60	85.77	85.74
GaborFilter	91.24	82.23	78.60	84.02	83.70
LocalObjectPattern [14]	95.32	92.54	90.27	92.71	92.66
TwoTier [41]	99.18	93.83	93.77	95.59	95.53
Deep learning for supervised classification					
AlexNet	99.39	**97.39**	75.88	90.89	89.53
GoogLeNet	**99.59**	97.04	80.16	92.26	91.40
Inception-v3	**99.59**	93.01	89.11	93.90	93.71
Deep learning for feature extraction (salient points)					
SalientStackedAE	97.35	90.17	93.00	93.50	93.41
SalientConvolutionalAE	96.54	93.96	76.26	88.92	87.94
Deep learning for feature extraction (random points)					
RandomRBM	95.93	87.91	96.89	93.58	93.40
RandomStackedAE [7]	97.96	90.05	90.27	92.76	92.62
RandomConvolutionalAE	95.32	88.63	79.38	87.77	87.28

to be averaged differ greatly. These results show that the proposed method leads to high test set accuracies, especially for the first dataset. The accuracy for the sub-low-grade cancerous classes decreases, as expected, since this subcategorization is a difficult task even for human observers. The receiver operating characteristic (ROC) curves of these classifications together with their area under the curve (AUC) metrics are reported in the supplementary material [39].

We also compare our method with four groups of other tissue classification algorithms; the comparison results are also provided in Tables II and III. The first group includes four methods, namely *CooccurrenceMatrix*, *GaborFilter*, *LocalObjectPattern*, and *TwoTier*, that use handcrafted features for image representation. We use them in our comparisons to investigate the effects of learning features directly on image data instead of manual feature definition. The *Cooccurrence-Matrix* and *GaborFilter* methods employ pixel-level textures. The *CooccurrenceMatrix* method first calculates a gray-level co-occurrence matrix and then extracts Haralick descriptors from this matrix. The *GaborFilter* method first convolves an image with log-Gabor filters in six orientations and four scales. Then, for each scale, it calculates average, standard deviation, minimum-to-maximum ratio, and mode descriptors on the response map averaged over those of all orientations [40]. Both methods use an SVM with a linear kernel for the final image classification. For both datasets, the proposed *DeepFeature* method leads to test set accuracies much better than these two methods, which employ pixel-level handcrafted features.

The *LocalObjectPattern* [14] and *TwoTier* [41] methods, which we previously developed in our research group, use component-level handcrafted features. The first one defines a descriptor with the purpose of encoding spatial arrangements of the components within the specified local neighborhoods. It is similar to this currently proposed method in the sense that it also represents the components with circular objects, labels them in an unsupervised way, and uses the labels'

distribution for image classification. On the other hand, it uses handcrafted features whereas the currently proposed method uses deep learning to learn the features directly from image data. The comparison results show the effectiveness of the latter approach. The *TwoTier* method decomposes an image into irregular-shaped components, uses Schmid filters [42] to quantify their textures and employs the dominant blob scale metric to quantify their shapes and sizes. At the end, it uses the spatial distribution of these components to classify the image. Although this method gives good results for the first dataset, it is not that successful to further subcategorize low-grade cancerous tissue images (Table III). The proposed *DeepFeature* method also gives the best results for this subcategorization. All these comparisons indicate the benefit of using deep learning for feature extraction.

The second group contains the methods that use CNN classifiers for entire image classification [13], [43], [44], [45]. These methods transfer their CNN architectures (except the last softmax layer since the number of classes is different) and their corresponding weights from the *AlexNet* [46], *GoogLeNet* [47], and *Inception-v3* [47] models, respectively, and fine-tune the model weights on our training images. Since these network models are designed for images with 227×227, 224×224, and 299×299 resolutions, respectively, we first resize our images before using the models. The experimental results given in Tables II and III show that the proposed *DeepFeature* method, which relies on characterizing the local salient subregions by deep learning, gives more accurate results than all these CNN classifiers, which are constructed for entire images without considering the saliency.

In the third group of methods (*SalientStackedAE* and *SalientConvolutionalAE*), we extract features from the salient subregions using two other deep learning techniques. Recall that our proposed method trains a deep belief network containing four layers of RBMs and uses the outputs of the RBM in the final layer as the features. We implement these comparison methods to investigate the effectiveness of using an RBM-based feature extractor for this application. The *SalientStackedAE* method trains a four-layer stacked autoencoder, whose architecture is the same with our network, and uses the outputs of the final autoencoder as its features. The *SalientConvolutionalAE* method trains a convolutional autoencoder and uses the encoded representation, which is the output of its encoding network, as the features. This convolutional autoencoder network has an encoder with three convolution-pooling layers (with 128, 64, and 32 feature maps, respectively) and a decoder with three deconvolution-upsampling layers (with 32, 64, and 128 feature maps, respectively). Its convolution/deconvolution layers use 3×3 filters and its pooling/upsampling layers use 2×2 filters. Both methods take the RGB values of a subregion as their inputs. Except using a different feature extractor for the salient subregions, the other steps of the methods remain the same. The test set accuracies obtained by these methods are reported in Tables II and III. When it is compared with *SalientConvolutionalAE*, the proposed *DeepFeature* method leads to more accurate results. The reason might be the following: We use the feature extractor to characterize small local subregions,

TABLE III
TEST SET ACCURACIES OF THE PROPOSED *DeepFeature* METHOD AND THE COMPARISON ALGORITHMS FOR THE SECOND DATASET.

	Norm.	Low (grade1)	Low (grade1-2)	Low (grade2)	High	Arith. mean	Harm. mean
DeepFeature	96.63	88.83	67.52	62.90	80.54	**79.28**	**77.24**
Handcrafted features							
CooccurrenceMatrix	87.64	71.51	50.43	39.52	78.38	65.50	60.03
GaborFilter	85.96	70.95	22.22	58.06	76.22	62.68	49.47
LocalObjectPattern [14]	92.70	89.39	48.72	58.87	77.30	73.40	69.04
TwoTier [41]	**98.88**	80.45	53.85	62.90	79.46	75.11	71.84
Deep learning for supervised classification							
AlexNet	97.19	**96.09**	35.90	52.42	**87.03**	73.73	63.20
GoogLeNet	97.75	81.56	**76.92**	63.71	61.62	76.31	74.17
Inception-v3	**98.88**	89.94	38.46	**66.94**	86.49	76.14	67.81
Deep learning for feature extraction (salient points)							
SalientStackedAE	98.31	87.71	55.56	58.87	83.24	76.74	72.92
SalientConvolutionalAE	**98.88**	80.45	45.30	51.61	70.27	69.30	63.92
Deep learning for feature extraction (random points)							
RandomRBM	87.08	82.12	56.41	58.87	82.16	73.33	70.88
RandomStackedAE [7]	97.19	82.12	47.01	57.26	82.70	73.26	68.22
RandomConvolutionalAE	96.07	72.63	45.30	44.35	59.46	63.56	58.40

whose characterizations will later be used to characterize the entire tissue image. The RBM-based feature extractor, each layer of which provides a fully connected network with a global weight matrix, may be sufficient to quantify a small subregion and learning the weights for such a small-sized input may not be that difficult for this application. On the other hand, a standard convolutional autoencoder network, each convolution/deconvolution layer of which uses local and shared connections, may not be that effective for such small local subregions and it may be necessary to customize its layers. The design of customized architectures for this application is considered as future work. The *SalientStackedAE* method, which also uses a fully connected network in each of its layers, improves the results of *SalientConvolutionalAE*, but it still gives lower accuracies compared to our proposed method.

The last group contains three methods that we implement to understand the effectiveness of considering the saliency in learning the deep features. The *RandomRBM* method is a variant of our algorithm. In this method, subregions are randomly cropped out of each image (instead of using the locations of tissue components) and everything else remains the same. Likewise, the *RandomStackedAE* and *RandomConvolutionalAE* methods are variants of *SalientStackedAE* and *SalientConvolutionalAE*, respectively. They also use randomly selected subregions instead of considering only the salient ones. Note that *RandomStackedAE* uses stacked auto-encoders to define and extract the features, as proposed in [7]. The experimental results are reported in Tables II and III. The results of all these variants reveal that extracting features from the salient subregions, which are determined by prior knowledge, improves the classification accuracies of their counterparts, especially for the second dataset. All these comparisons indicate the effectiveness of using the proposed RBM-based feature extractor together with the salient points.
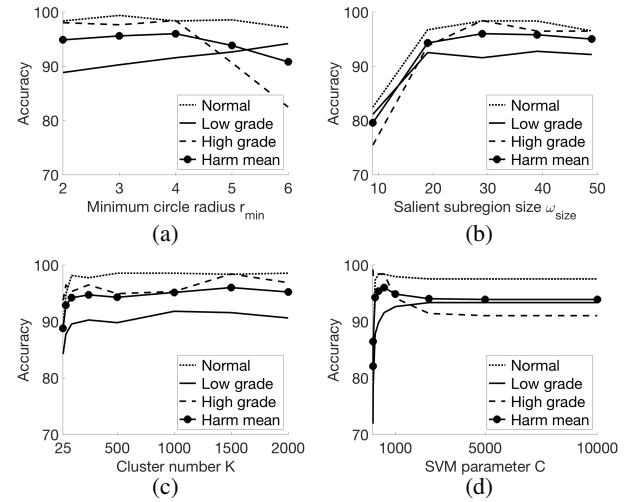

Fig. 4. For the first dataset, test set accuracies as a function of the model parameters: (a) minimum circle radius r_{min}, (b) size of a salient subregion ω_{size}, (c) cluster number K, and (d) SVM parameter C. The parameter analysis for the second dataset are given as supplementary material [39].

D. Parameter Analysis

The *DeepFeature* method has four external parameters: minimum circle radius r_{min}, size of a salient subregion ω_{size}, cluster number K, and SVM parameter C. This section analyzes the effects of the parameter selection on the method's performance. To this end, for each parameter, it fixes the values of the other three parameters and measures the test set accuracies as a function of the parameter of interest. These analyses are depicted in Fig. 4 and discussed below for the first dataset. The analyses for the second dataset are parallel to those of the first one. The reader is referred to the technical report [39] for the latter analyses.

The minimum circle radius r_{min} determines the size of the smallest circular object (tissue component) located by the CIRCLEDECOMPOSITION algorithm. Its larger values cause

not to locate smaller objects, which may correspond to important small tissue components such as nuclei, and not to define salient subregions around them. This may cause an inadequate representation of the tissue, which decreases the accuracy as shown in Fig. 4(a). On the other hand, using smaller values leads to defining noisy objects and the use of the salient subregions around them slightly decreases the accuracy.

The parameter ω_{size} is the size of a salient subregion cropped for each component by the CROPWINDOW algorithm. This parameter determines the locality of the deep features. When ω_{size} is too small, it is not sufficient to accurately characterize the subregion, and thus, the component it corresponds to. This significantly decreases the accuracy. After a certain point, it does not affect the accuracy too much, but of course, increases the complexity of the required deep neural network. This analysis is depicted in Fig. 4(b).

The cluster number K determines the number of labels used for quantizing the salient subregions (components). Its smaller values may result in defining the same label for components of different types. This may lead to an ineffective representation, decreasing the accuracy. Using larger values only slightly affects the performance (Fig. 4(c)).

The SVM parameter C controls the trade-off between the training error and the margin width of the SVM model. Using values smaller and larger than necessary may cause underfitting and overfitting, respectively. Unfortunately, similar to many hyperparameters in machine learning, there is no foolproof method for its selection and its value must be determined empirically. As shown in Fig. 4(d), our application necessitates the use of C in the range between 250 and 1000.

E. Discussion

This work introduces a new feature extractor for histopathological image representation and presents a system that uses this representation for their classification. This system classifies an image with one of the predefined classes, assuming that it is homogeneous. This section discusses how this system can be used in a digital pathology setup, in which typically lower magnifications are used to scan a slide. Thus, the acquired images usually have a larger field of view and may be homogeneous or heterogeneous. To this end, this section presents a simple algorithm that detects the regions belonging to one of the predefined classes in such a large image. Developing more sophisticated algorithms for the same purpose or for different applications could be considered as future research work.

Our detection algorithm first slides a window with a size that the classification system uses (in our case, the size of 480×640) over the entire large image and then extracts the features of each window and classifies it by the proposed *DeepFeature* method. Since these windows may not be homogeneous, it does not directly output the estimated class labels, but instead, it uses the class labels of all windows together with their posteriors in a seed-controlled region growing algorithm. In particular, this detection algorithm has three main steps: posterior estimation, seed identification, and seed growing. All these steps run on circular objects, which we previously define to approximate the tissue components and to represent the

salient subregions, instead of image pixels, since the latter is much more computationally expensive. Thus, before starting these steps, the circular objects are located on the large image and the connectivity between them are defined by constructing a Delaunay triangulation on their centroids.

The first step slides a window over the objects and estimates posteriors for all sliding windows by *DeepFeature*. Then, for each object, it accumulates the posteriors of all sliding windows that cover this object. Since our system classifies a window with a predefined class and since these classes may not cover all tissue formations (e.g., lymphoid or connective tissue), this step defines a reject action and assigns it a probability. It uses a very simple probability assignment; the reject probability is 1 if the maximum accumulated posterior is greater than 0.5, and 0 otherwise. The objects are then relabeled by also considering the reject probabilities. As future work, one may define the reject probability as a function of the class posteriors. As an alternative, one may also consider to define classes for additional tissue formations and retrain the classifier. The second step identifies the seeds using the object labels and posteriors. For that, it finds the connected components of the objects that are assigned to the same class with at least T_{seed} probability. It identifies the components containing more than T_{no} objects as the seeds. In our experiments, we set $T_{seed} = 0.90$ and $T_{no} = 500$. The last step grows the seeds on the objects with respect to their posteriors. At the end, the seeds of objects are mapped to image pixels by assigning each pixel the class of its closest seed object, and the seed boundaries are smoothed by majority filtering.

We test this detection algorithm on a preliminary dataset of 30 large images. These images were taken with a $5\times$ objective lens and the image resolution is 1920×2560. Most of the images are heterogeneous; only five of them are homogeneous to test the algorithm also on large homogeneous images. In our tests, we will directly use the classifier trained for our first dataset without any modification or additional training. Hence, the aim will be to detect low-grade and high-grade colon adenocarcinomatous regions on these large images as well as those containing normal colon glands. Thus, we only annotate those regions on the large images. Example images together with their annotations are given in Fig. 5; more can be found in [39]. The visual results of the algorithm are also given for these examples. For quantitative evaluation, the recall, precision, and F-score metrics are calculated for each class separately. For class C, the standard definitions are as follows: Precision is the percentage of correctly classified C pixels that actually belong to C. Recall is the percentage of actual C pixels that are correctly classified as C by the algorithm. F-score is the harmonic mean of these two metrics. The results for these metrics are reported in Table IV. This table also reports the results obtained by relaxing the precision and recall definitions with respect to our application, in which the aim is colon adenocarcinoma detection. Since this cancer type mainly affects epithelial cells, non-epithelial regions are left as unannotated in our datasets. Indeed, one may include these regions to any class without changing the application's aim. Thus, for class C, we relax the definitions as follows: Precision is the percentage of correctly classified C pixels that
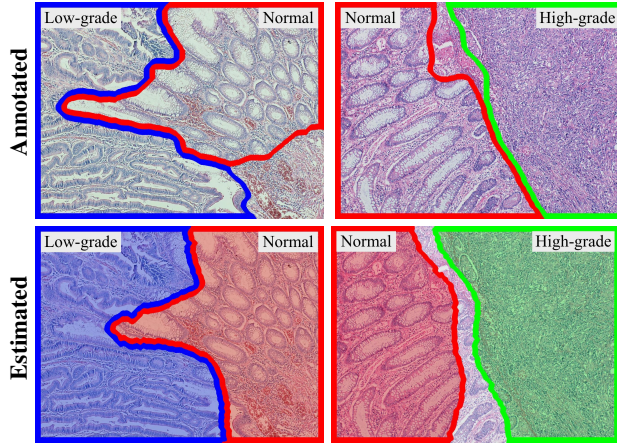
Fig. 5. Examples of large heterogeneous images together with their visual results obtained by the colon adenocarcinoma detection algorithm. The boundaries of the annotated/estimated normal, low-grade cancerous, and high-grade cancerous regions are shown with red, blue, and green, respectively. More examples can be found in [39].

TABLE IV
RESULTS OF THE COLON ADENOCARCINOMA DETECTION ALGORITHM ON A PRELIMINARY DATASET OF LARGE IMAGES.

	Standard Definitions			Relaxed Definitions		
	Precis.	Recall	F-score	Precis.	Recall	F-score
Normal	92.96	79.71	85.83	99.48	88.37	93.60
Low-grade	83.01	91.30	86.96	91.03	93.32	92.16
High-grade	70.82	98.61	82.44	87.00	99.93	93.02

actually belong to C or a non-epithelial region. Recall is the percentage of actual C pixels that are correctly classified as C or with the reject class by the algorithm.

The visual and quantitative evaluations reveal that the detection algorithm, which uses the proposed classification system, leads to promising results. Thus, it has the potential to be used with a whole slide scanner. To do that, a whole slide should be scanned with a low magnification of the scanner, and the acquired image, which has a larger field of view, can be analyzed by this detection algorithm. Although it yields successful results for many large images, it may also give misclassifications for some of them, especially for those containing relatively large non-epithelial regions; an illustrative example is given in [39]. When non-epithelial regions are small, incorrect classifications can be compensated by correct classifications of nearby regions and the reject action. However, when they are large, such compensation may not be possible and the system gives incorrect results since there is no separate class for such regions. Defining an extra class(es) will definitely improve the accuracy on these regions. This is left as future research work.

V. CONCLUSION

This paper presents a semi-supervised classification method for histopathological tissue images. As its first contribution, this method proposes to determine salient subregions in an image and to use only the quantizations (characterizations) of these salient subregions for image representation and classification. As the second contribution, it introduces a new

unsupervised technique to learn the subregion quantizations. For that, it proposes to construct a deep belief network of consecutive RBMs whose first layer takes the pixels of a salient subregion and to define the activation values of the hidden unit nodes in the final RBM as its deep features. It then feeds these deep features to a clustering algorithm for learning the quantizations of the salient subregions in an unsupervised way. As its last contribution, this study is a successful demonstration of using restricted Boltzmann machines in the domain of histopathological image analysis. We tested our method on two datasets of microscopic histopathological images of colon tissues. Our experiments revealed that characterizing the salient subregions by the proposed local deep features and using the distribution of these characterized subregions for tissue image representation lead to more accurate classification results compared to the existing algorithms.

In this work, we use the histogram of quantized salient subregions for defining a global feature set for the entire image. One future research direction is to investigate the other ways of defining this global feature set, such as defining texture measures on the quantized subregions. Another research direction is to explore the use of different network architectures. For example, one may consider combining the activation values in different hidden layers to define a new set of deep features. On an example application, we have discussed how the proposed system can be used in a digital pathology setup. The design of sophisticated algorithms for this purpose is another future research direction of this study.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.
[2] M. N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: a review," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 147-171, Oct 2009.
[3] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504-507, 2006.
[4] H. Rezaeilouyeh, A. Mollahosseini, and M. H. Mahoor, "Microscopic medical image classification framework via deep learning and shearlet transform," *J. Med. Imaging*, vol. 3, no. 4, pp. 044501, Oct. 2016.
[5] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proc. Int. Conf. On Pattern Recogn.*, 2016, pp. 2440-2445.
[6] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol Informatics*, vol. 7, pp. 29, 2016.
[7] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, and F. A. Gonzalez, "An unsupervised feature learning framework for basal cell carcinoma image analysis," *Art. Intel. Medicine*, vol. 64, no. 2, pp. 131-145, Jun. 2015.
[8] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1196-1206, Feb. 2016.

[9] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Med. Image Comput. Comput. Assisted Intervent.*, 2013, pp. 411-418.

[10] 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.

[11] T. Chen and C. Chefd'hotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *Machine Learning in Medical Imaging*, New York: Springer, 2014, pp. 17-24.

[12] 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.

[13] B. E. Bejnordi, *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, 2199-2210, 2017.

[14] G. Olgun, C. Sokmensuer, and C. Gunduz-Demir, "Local object patterns for tissue image representation and cancer classification," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1390-1396, Jul. 2014.

[15] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies", *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1205-1218, May 2012.

[16] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 697-704, Jun. 2003.

[17] O. Sertel, J. Kong, H. Shimada, U. V. Catalyurek, J. H. Saltz, and M.N. Gurcan, "Computer-aided prognosis of neuroblastoma on whole slide images: classification of stromal development," *Pattern Recognit.*, vol. 42, no. 6, pp. 1093-1103, Jun. 2009.

[18] 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.

[19] D. Altunbay, C. Cigir, C. Sokmensuer, and C. Gunduz-Demir, "Color graphs for automated cancer diagnosis and grading", *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 665-674, Mar. 2010.

[20] E. Ozdemir and C. Gunduz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Trans. Med. Imaging*, vol. 32, no. 2, pp. 474-483, Feb. 2013.

[21] B. E. Bejnordi, G. Zuidhof, M. Balkenhol, M. Hermsen, P. Bult, B. van Ginneken, N. Karssemeijer, G. Litjens, and J. van der Laak, "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *J. Med. Imaging*, vol. 4, no. 4, pp. 044504, Dec. 2017.

[22] S. Manivannan, W. Li, J. Zhang, E. Trucco, and S. J. McKenna, "Structure prediction for gland segmentation with hand-crafted and deep convolutional features," *IEEE Trans. Med. Imaging*, vol. 37, no. 1, pp. 210-221, Jan. 2018.

[23] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52-i59, Jun. 2016.

[24] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multi-scale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421-2433, Oct. 2015.

[25] P. Naylor, M. Lae, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *2017 IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 933-936.

[26] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. Med. Imaging*, vol. 1, no. 3, pp. 034003, Oct. 2014.

[27] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Comput. Biol.*, vol. 12, no. 11, pp. e1005177, Nov. 2016.

[28] S. U. Akram, J. Kannala, L. Eklund, and J. Heikkila, "Cell proposal network for microscopy image analysis," in *Proc. Int. Conf. Image Proc.*, 2016, pp. 3199-3203.

[29] B. Dong, L. Shao, M. Da Costa, O. Bandmann, and A. F. Frangi, "Deep learning for automatic cell detection in wide-field microscopy zebrafish images," in *Proc. Int. Symp. Biomed. Imag.*, 2015, pp. 772-776.

[30] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88-99, Mar. 2017.

[31] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1550-1560, Jul. 2017.

[32] C. Li, X. Wang, W. Liu, and L. J. Latecki, "DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks," *Med. Image Anal.*, vol. 45, pp. 121-133, Apr. 2018.

[33] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Proc. Med. Image Comput. Comput. Assisted Intervent.*, 2013, pp. 403-410.

[34] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 119-130, Jan. 2016.

[35] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, pp. 291-299, 2001. Available: http://www.dentistry.bham.ac.uk/landinig/software/cdeconv/cdeconv.html.

[36] A. B. Tosun, M. Kandemir, C. Sokmensuer, and C. Gunduz-Demir, "Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection," *Pattern Recognit.*, vol. 42, no. 6, pp. 1104-1112, 2009.

[37] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, pp. 1771-1800, 2002.

[38] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Tech.*, vol. 2, no. 3, pp. 1-27, 2011. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[39] C. T. Sari and C. Gunduz-Demir, "Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images: Supplementary material," Technical Report BU-CE-1801, Computer Engineering, Bilkent University, 2018. Available http://www.cs.bilkent.edu.tr/tech-reports/2018/BU-CE-1801.pdf.

[40] 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.

[41] T. Gultekin, C. F. Koyuncu, C. Sokmensuer, and C. Gunduz-Demir, "Two-tier tissue decomposition for histopathological image representation and classification," *IEEE Trans. Med. Imaging*, vol. 34, no. 1, pp. 275-283, Jan. 2015.

[42] C. Schmid, "Constructing models for content-based image retrieval," in *Proc. IEEE Int. Conf. Comp. Vis. Pattern Recog.*, Jun. 2001, pp. 39-45.

[43] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *Image Analysis and Recognition, ICIAR 2018*, Lecture Notes in Computer Science, vol. 10882, pp. 914-922.

[44] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, "Classification of breast cancer histology images using transfer learning," in *Image Analysis and Recognition, ICIAR 2018*, Lecture Notes in Computer Science, vol. 10882, pp. 763-770.

[45] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.

[46] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classication with deep convolutional neural networks," in *Proc. Adv. in Neural Inf. Proc. Syst.*, 2012.

[47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, Jun. 2015, pp. 1-9.