

AN AUTOMATIC BREAST CANCER GRADING METHOD IN HISTOPATHOLOGICAL IMAGES BASED ON PIXEL-, OBJECT-, AND SEMANTIC-LEVEL FEATURES

Jiajia Cao[†] Zengchang Qin^{†*} Juan Jing[‡] Jianhui Chen[§] Tao Wan^{†*}

[†] Intelligent Computing & Machine Learning Lab, School of ASEE, Beihang University, China

[‡] School of Biological Science and Medical Engineering, Beihang University, China

[§] No. 91 Central Hospital of PLA, Henan, China

ABSTRACT

We present an automatic breast cancer grading method in histopathological images based on the computer extracted pixel-, object-, and semantic-level features derived from convolutional neural networks (CNN). The multiple level features allow not only characterization of nuclear polymorphism, but also extraction of structural and interpretable information within the images. In this study, a hybrid level set based segmentation method was used to segment nuclei from the images. A quantile normalization approach was utilized to improve image color consistency. The semantic level features are extracted by a CNN approach, which describe the proportions of nuclei belonging to the different grades, in conjunction with pixel-level (texture) and object-level (structure) features, to form an integrated set of attributes. A support vector machine classifier was trained to discriminate the breast cancer between low, intermediate, and high grades. The results demonstrated that our method achieved accuracy of 0.92 (low vs. high), and 0.74 (low vs. intermediate), and 0.76 (intermediate vs. high), suggesting that the present method could play a fundamental role in developing a computer-aided breast cancer grading system.

Index Terms— breast cancer grading, multi-level features, convolutional neural networks, histopathology

1. INTRODUCTION

Breast cancer grading on histopathological images is considered as a standard clinical practice for the diagnosis and prognosis of breast cancer development. The Nottingham grading system is the most widely used criterion for histological diagnosis of invasive breast cancers with combination of nuclear pleomorphism, tubular formation, and mitotic count [1] (see Fig.1). In routine histological analysis, pathologists perform grading by manually examining breast cancer tissue specimen under a microscope, which is a tedious and subjective process and thus suffer inter- and intra-observer variations. In this

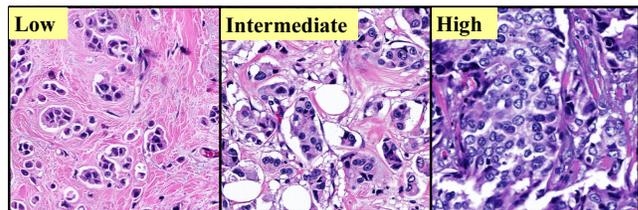


Fig. 1. Examples of breast cancer histopathological image patches for low, intermediate, and high Nottingham grades showing different appearances of nuclear polymorphism in histopathology.

work, we aim to developing an automatic breast cancer grading method in digitized histopathological images, mimicking the Nottingham grading system approach, in order to assist pathologists in enhancing the operational efficiency as well as improving diagnostic confidence.

Recently, many techniques have been presented for automatic breast cancer grading based on the Nottingham scoring system in histopathological images [2]. For instance, Petushi *et al.* modeled the microstructure present in histopathology to characterize the localized tubular formation [3]. Wan *et al.* presented a statistical method to automatically detect mitotic cells [4]. These methods handle only one of three criteria in the Nottingham system. Further, Basavanhally *et al.* introduced a multi-field-of-view framework to integrate image nucleus features including graph-based and textural features to discriminate low, intermediate, and high grades of breast cancers [5]. This motivates our work to differentiate breast cancer grades by combining multiple levels of features.

In this work, we present an automatic method for breast cancer grading based on a combination of pixel-, object-, and semantic-level features. The pixel-level features characterize the micro-textures within the nuclei. The object-level features are computed using three different graphs to quantify nuclear architecture. Although these features are directly computed from images reflecting explicit attributes that pathologists look for when grading breast cancers, there is another category of feature generation inspired by convolutional neural

* Corresponding authors' e-mail: {taowan, zcqin}@buaa.edu.cn. This work was partially supported by the National Natural Science Foundation of China under award Nos. 61305047 and 61401012.

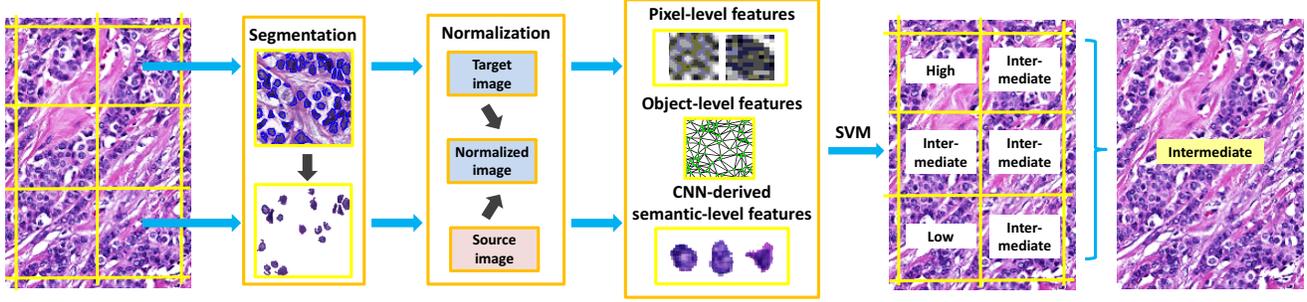


Fig. 2. Workflow of the breast cancer grading method. The nuclei are segmented from the original histopathological images and normalized before extracting pixel-, object-, and semantic-level features. The cancer grade is determined via a SVM classifier.

networks (CNN). In [6], Wang *et al.* showed that a combination of CNN modeled and handcrafted features for mitosis detection yielded improved detection results. We compute the semantic-level features via a data-driven CNN approach which can learn additional feature bases that cannot be represented through any of pixel- and object-level features. Our semantic-level features calculate potential proportions of nuclei belonging to the different cancer grades. The workflow of the presented method is depicted in Fig.2. A hybrid level set based segmentation method was used to automated segment nuclei from the images. A color-map quantile normalization approach was utilized to correct image intensity variations. The features were extracted independently and evaluated through support vector machine (SVM) classification. In summary, key contributions of this work include: (i) A integration of multiple level features for breast cancer grading; (ii) Learning important features for differentiating breast cancers with low, intermediate, and high grades.

The remainder of this paper is organized as follows. The methods are described in Section 2. In Section 3, we demonstrate experimental results with discussion. Section 4 concludes the paper.

2. METHODOLOGY

2.1. Nucleus Segmentation

To segment nuclear regions, we use a hybrid level-set method [7], which combines boundary and region information to perform image segmentation. The method tends to minimize the energy function defined as: $\varepsilon = \alpha \int (I - \mu)H(\phi)d\omega + \beta \int g|\nabla H(\phi)|d\omega$ where I is the image, ϕ is a zero level set function, ω refers to image region, μ is a pre-defined upper bound of the grey value, $g = g(|\nabla I|)$ is the feature gradient of image, $H(\phi)$ is a step function, α and β are weights for the closed curve and geodesic active contour model, respectively.

2.2. Stain Normalization

Our histopathological images are stained with hematoxylin and eosin (H&E). To achieve a consistent color and intensity

appearance, we use a pixel-based color-map quantile normalization method [8]. We define three ranking functions F^c , where $c \in [R, G, B]$ (color space). We adjust the color values on a pixel-by-pixel basis to match the color distribution of a source image s to that of a target image t . For example, $F_s^c(k_s^c) = r_{s,k}^c$, where $k_s^c \in [0, 255]$, and $r_{s,k}^c \in [1, N_s]$, N_s is the number of pixels in s . We map the k^{th} pixel intensity in c color channel to a rank $r_{s,k}^c$. The normalization of k_s^c is computed via $\tilde{k}_s^c = F_t^{-1}(\frac{r_{s,k}^c}{N_s}N_t + \frac{1}{2})$.

2.3. Feature Extraction

We compute three types of features that capture image attributes at the pixel, object, and semantic levels.

Pixel-level features

The extracted pixel-level features consider all image pixels and capture property of texture to quantify image sharpness, contrast, changes in intensity, and discontinuities, etc. We compute Haralick gray-level co-occurrence matrix features, Gabor filter, and speeded up robust features (SURF) [9] to form the pixel-level feature set (F_p). We calculate mean, median, variance, and minimum-to-maximum ratio on the Gabor filtered images. The SURF features are scale and rotation-invariant descriptors to characterize the intensity distribution of the pixels within the neighborhood of the point of interest. These local features are useful in the characterization of small objects, such as nuclei.

Object-level features

The object-level features are used to capture the spatial distribution of nuclear structures within the image figures. We compute three spatial graphs, including Voronoi diagram (VD), Delaunay triangulation (DT), and minimum spanning tree (MST), for extracting topological features. The VD is defined by a set of convex polygons surrounding all nuclear centroids $C = \{c_1, \dots, c_n\}$. The polygons are defined as $P(c_i) = \{C | \mathcal{D}(C, c_j) \leq \mathcal{D}(C, c_i)\}$, where $j \neq i, j = 1, \dots, n$ and $\mathcal{D}(\cdot)$ is Euclidean distance. The DT is simply the dual graph of VD and is constructed by connecting three points c_r, c_s, c_t , when three polygons, $P(c_r), P(c_s)$ and $P(c_t)$ are connected with an intersection point of VD. The MST is a span-

ning tree of a connected, undirected graph which connects all the vertices together with the minimal weighting for its edges. We measure area, perimeter, and chord length for the VD, side length and area for the DT, and branch length for MST. In addition, we compute the number of nuclei within a pre-defined circle region and radius of the circle containing a particular number of nuclei to describe the density of the nuclear distribution. The 4 statistical features (mean, median, variance, and minimum-to-maximum ratio) of these graph and density measures are computed to provide a set of 32 object-level features (F_o).

Semantic-level features

Compared to pixel- and object-level features, semantic-level features are in a higher level of the information hierarchy which can easily capture interpretable concepts [8]. With the large amount of biological variations present in histopathological images due to the heterogeneity of cancer biology, development of CNN-derived semantic-level descriptors which involves a large data training becomes feasible. Since CNN approach is a supervised feature generation method, we train the CNN model with labeled segmented nuclei having low, intermediate, and high grades. The CNN architecture is designed with 2 consecutive convolutional and pooling layers, and a fully-connected layer (see Fig.3) [6]. The extraction of semantic-level features (F_s) is as follows:

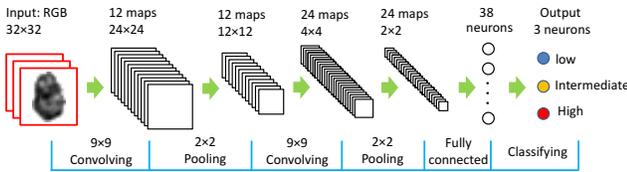


Fig. 3. A 3-layer CNN model with two convolutional and pooling layers, and a fully-connected layer.

1. Input: The nuclear figures with RGB color channels are resized into 32×32 pixels;
2. Convolutional layer: A two-dimensional convolution of the input feature maps with a 9×9 convolution kernel. The convolution function defined as $y_j = \tanh(\sum_i k_{ij}x_i)$, where $\tanh(\cdot)$ is an activation function, k_{ij} is a convolution kernel, x_i and y_j represent the value of the i^{th} input and j^{th} output feature maps, respectively;
3. Pooling layer: We apply a subsampling pooling operation over a 2×2 non-overlapping window on each output feature map, allowing to learn the invariant features;
4. Fully connected layer: 38 neurons in this layer are connected to the output feature maps of pooling layer;
5. Output: Three neurons (low, intermediate, and high grades) in this layer are activated by a logistic regression model;
6. The proportions of nuclei belonging to different grades are computed.

2.4. SVM based Classification

The computerized features are evaluated via a SVM classifier to distinguish images with different grades. The SVM classifier projects the feature set onto a higher dimensional space using a linear kernel and the hyperplane that most accurately separates the two classes is determined. The grading algorithm presented below is finally performed to determine the cancer grade of image I .

The Grading Algorithm

Input: Image I , SVM classifier $\mathcal{C}(F, p)$

Output: Cancer grade $G(F, P)$

begin

1. Divide I into M image patches $P = \{p_1, \dots, p_M\}$;
2. Segment K nuclei $n^i = \{n_1^i, \dots, n_K^i\}$ from p_i ;
3. Stain normalization of n^1, \dots, n^M ;
4. Train CNN model for extracting F_s ;
5. Feature set $F_i = \{F_p, F_o, F_s\}$, $i \in \{1, \dots, M\}$;
6. $\mathcal{C}(F_i, p_i)$, $G(F, P) = \text{sign}(\frac{1}{M} \sum_{i=1}^M \mathcal{C}(F_i, p_i))$.

end

3. EXPERIMENTAL RESULTS

3.1. Experimental Design

A total of 89 H&E stained slides were collected from 89 patient studies diagnosed using the Nottingham grading system (low: $N = 22$; intermediate: $N = 49$; high: $N = 18$). All slides were digitized via a whole slide scanner (Motic[®], Xiamen, China) at $40\times$ magnification. An expert pathologist manually delineated the regions of interest containing cancerous tissues, which were divided into 6740 non-overlapping image patches (1000×1000 pixels). We used 865 patches from 18 patient studies containing about 4000 nuclei for each grade category to compute CNN-based semantic features. The rest of 71 cases (5875 patches) were randomly sampled to generate a training set Z_{tra} and a testing set Z_{tes} with $Z_{tra} \cap Z_{tes} = \emptyset$. The slide grade was determined by the majority of grade categories associated with the patches.

The classification was performed via an iterative 2-fold cross-validation process, and the resulting mean and standard deviation of the classification accuracy were computed. In addition, the receiver operating characteristic (ROC) analysis and an area under the curve (AUC) were used to quantitatively measure the extracted features' ability in distinguishing low, intermediate, and high grades of breast cancers.

3.2. Results

The classification results using pixel-level features (F_p), object-level features (F_o), semantic-level features (F_s), and their combinations are listed in Table 1. We noted that the combination of CNN-learned semantic-level and object-level

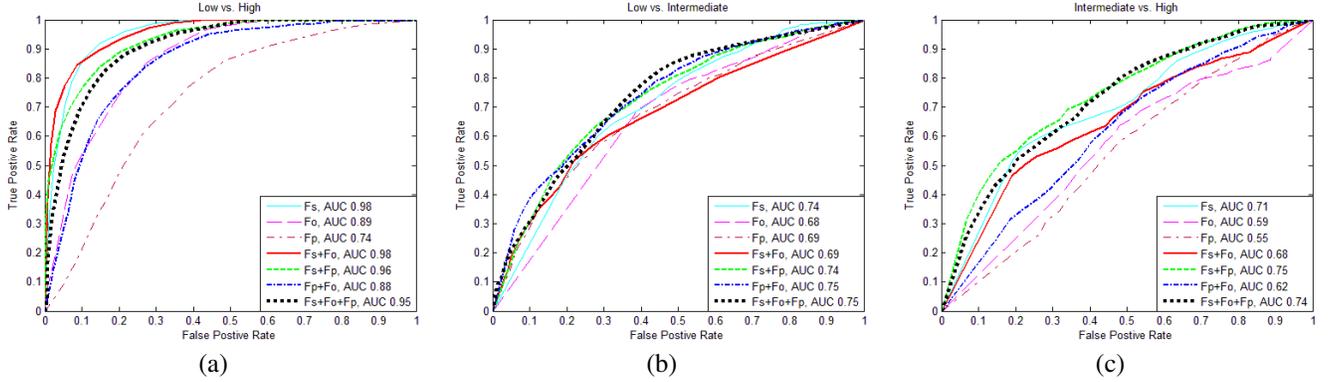


Fig. 4. ROC curves of different feature sets in (a) low versus high grades; (b) low versus intermediate grades; (c) intermediate versus high grades. The semantic-level features (F_s) or its combinations yielded the best classification performance.

features yielded the best performance in distinguishing low from high grade breast cancers with the accuracy of 0.92. The combinations of three feature types achieved the highest accuracy in the other two classification tasks. The ROC curves (see Fig.4) also provide consistent results in terms of AUC values. In all the three classification tasks, the semantic-level features outperformed the other two types of features and their combinations, due to the fact that the CNN-derived features were fully data-drive, therefore they are more accurate in representing training samples and are able to find feature patterns that the other two feature types fail to describe. Further, the extracted semantic-level features are low dimensional which result in a significant reduction in storage of the image features, as well as in the computational complexity of the classification task.

Table 1. Classification accuracy using pixel-level (F_p), object-level (F_o), semantic-level (F_s) features, and their combinations.

Feature set	Classification task		
	Low vs. High	Low vs. Inter	Inter vs. High
F_s	0.91±0.03	0.73±0.02	0.75±0.02
F_o	0.79±0.06	0.72±0.07	0.71±0.02
F_p	0.68±0.10	0.71±0.03	0.72±0.02
$F_s + F_o$	0.92±0.03	0.70±0.02	0.74±0.02
$F_s + F_p$	0.89±0.04	0.72±0.03	0.75±0.03
$F_o + F_p$	0.80±0.07	0.71±0.03	0.71±0.03
$F_s + F_o + F_p$	0.90±0.03	0.74±0.03	0.76±0.04

4. CONCLUSION

In this paper, we presented a multi-level feature based method for automatic breast cancer grading of histopathological images. The pixel-, object-, and semantic-level features and their combinations were evaluated via a SVM classifier for distinguishing different grades of breast cancers. The quantitative results suggested that our new CNN-derived semantic-level features are promising image markers to stratify more or

less aggressive breast cancers.

5. REFERENCES

- [1] C.W. Elston and I.O. Ellis, "Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.
- [2] M. Veta et al., "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, 2014.
- [3] S. Petushi et al., "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Med. Imaging*, vol. 6, no. 1, 2006.
- [4] T. Wan et al., "Wavelet-based statistical features for distinguishing mitotic and non-mitotic cells in breast cancer histopathology," in *ICIP*, 2014, pp. 2290–2294.
- [5] A. Basavanahally et al., "Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2089–2099, 2013.
- [6] H. Wang et al., "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. of Med. Imaging*, vol. 1, no. 3, 2014.
- [7] Y. Zhang et al., "Medical image segmentation using new hybrid level-set method," in *MEDIVIS*, 2008, pp. 71–76.
- [8] S. Kothari et al., "Pathology imaging informatics for quantitative analysis of whole-slide images," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 6, pp. 1099–1108, 2013.
- [9] H. Bay et al., "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, pp. 346–359, 2008.