Computer Methods and Programs in Biomedicine 198 (2021) 105807

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Cervical cell classification with graph convolutional network

Jun Shi[a,*], Ruoyu Wang[a], Yushan Zheng[b,c,d], Zhiguo Jiang[b,c,d], Haopeng Zhang[b,c,d], Lanlan Yu[e]

[a] School of Software, Hefei University of Technology, Hefei 230601, China
[b] Image Processing Center, School of Astronautics, Beihang University, Beijing, 100191, China
[c] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, 100191, China
[d] Beijing Key Laboratory of Digital Media, Beihang University, Beijing, 100191, China
[e] Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen 361101, China

ARTICLE INFO

Article history:
Received 26 June 2020
Accepted 12 October 2020

Keywords:
Cervical cancer screening
Cervical cytology
Cervical cell classification
Graph convolutional network

ABSTRACT

Background and objective: Cervical cell classification has important clinical significance in cervical cancer screening at early stages. In contrast with the conventional classification methods which depend on hand-crafted or engineered features, Convolutional Neural Network (CNN) generally classifies cervical cells via learned deep features. However, the latent correlations of images may be ignored during CNN feature learning and thus influence the representation ability of CNN features.

Methods: We propose a novel cervical cell classification method based on Graph Convolutional Network (GCN). It aims to explore the potential relationship of cervical cell images for improving the classification performance. The CNN features of all the cervical cell images are firstly clustered and the intrinsic relationships of images can be preliminarily revealed through the clustering. To further capture the underlying correlations existed among clusters, a graph structure is constructed. GCN is then applied to propagate the node dependencies and thus yield the relation-aware feature representation. The GCN features are finally incorporated to enhance the discriminative ability of CNN features.

Results: Experiments on the public cervical cell image dataset SIPaKMeD from International Conference on Image Processing in 2018 demonstrate the feasibility and effectiveness of the proposed method. In addition, we introduce a large-scale Motic liquid-based cytology image dataset which provides the large amount of data, some novel cell types with important clinical significance and staining difference and thus presents a great challenge for cervical cell classification. We evaluate the proposed method under two conditions of the consistent staining and different staining. Experimental results show our method outperforms the existing state-of-arts methods according to the quantitative metrics (i.e. accuracy, sensitivity, specificity, F-measure and confusion matrices).

Conclusions: The intrinsic relationship exploration of cervical cells contributes significant improvements to the cervical cell classification. The relation-aware features generated by GCN effectively strengthens the representational power of CNN features. The proposed method can achieve the better classification performance and also can be potentially used in automatic screening system of cervical cytology.

1. Introduction

Cervical cancer is one of the primary causes of cancer death in women [1]. Screening at early stages is crucial to the prevention and early detection of cervical cancer. As the most common screening test, cervical cytology has been extensively used and effectively reduced the incidence and mortality. At present, manual screening of abnormal cells from a cervical cytology slide is still the common practice. However, it is usually tedious, inefficient and high-cost. Consequently, the automatic screening method has attracted increasing attention [2–4]. It generally applies image processing and machine learning techniques to automatically detect abnormal cells from a given cytology slide, determine the categories and finally present the analysis results of the whole slide according to The Bethesda System (TBS).

Accurate cervical cell classification is crucial to the automatic screening method. Over the last decades, many cervical cell image

* Corresponding author.
E-mail addresses: juns@hfut.edu.cn (J. Shi), aywry@mail.hfut.edu.cn (R. Wang), yszheng@buaa.edu.cn (Y. Zheng), jiangzg@buaa.edu.cn (Z. Jiang), zhanghaopeng@buaa.edu.cn (H. Zhang), yull@motic.com (L. Yu).

https://doi.org/10.1016/j.cmpb.2020.105807
0169-2607/

classification methods have been proposed. Chankong et al. [5] use morphological features of segmented nuclei and cytoplasm to achieve the multi-label classification. Bora et al. [6] combine the shape, texture and color features of segmented nuclei to classify the cervical dysplasia. Arya et al. [7] fuse multiple texture features including first-order histogram, grey-level co-occurrence matrix (GLCM), local binary pattern (LBP), Laws energy texture features and discrete wavelet transform (DWT) for the cervical cell classification. Wang et al. [8] perform feature selection on the shape, textural and Gabor features of nuclei, and then apply generated optimal feature set and support vector machine (SVM) for the classification of normal, uninvolved and abnormal cells. These methods essentially use hand-crafted or engineered features and one or multiple classifiers for classification. As a result, they are inevitably influenced by feature or classifier selection.

In the last few years, Convolutional Neural Network (CNN) [9–13] has been proposed to automatically learn multi-level features through hierarchical deep architecture. A variety of CNN models have been successfully used in computational pathology [14–19] and cervical cytology [20–25]. Zhang et al. [20] take the cell image patch coarsely centered on each nucleus as the network input and fine-tune the pre-trained CNN model for classification. Taha et al. [21] use the pre-trained AlexNet [9] as feature extractor and couple with SVM. Wieslander et al. [22] employ VGG [10] and ResNet [11] for binary classification of benign and malignant cervical cells. Plissiti et al. [23] propose the annotated cervical cell image dataset SIPaKMeD and apply VGG-19 [10] for classifying five types of cervical cells. Gautam et al. [24] perform transfer learning on AlexNet for both single and multi-cell images. Sompawong et al. [25] use Mask Regional Convolutional Neural Network (Mask R-CNN) [13] to simultaneously achieve the detection and classification. Obviously, these CNN-based methods obtain superior classification performance compared to the traditional methods which rely on hand-crafted or engineered features. However, they ignore the potential relationships among cervical cell images during feature learning and thus may influence the representation ability of CNN features. Meanwhile, some research on cervical cell analysis show that each independent cervical cell has the intrinsic similarity [26–28]. For example, superficial and intermediate cells generally have relatively small nuclei, and have a clear cytoplasm and nucleus margin while Dyskeratotic and Metaplastic cells have an overlapping cytoplasm and nucleus margin [26,29]. Besides Koilocytotic cells have the presence of perinuclear cavity while other cells have a relatively thick cytoplasm [26,29]. These observations indicate there exists the potential relationship of cervical cell images.

Graph offers an intuitive and reasonable way to represent the relationships of images and how to apply CNN for graph has increasingly become the focus. Recently, graph convolutional network [30–34] has attracted more attention, which intends to generalize CNN on non-Euclidean graph data. It can be roughly divided into spatial convolution and spectral convolution methods. Spatial methods generally directly perform graph convolution operation on the node neighbors of the graph. Duvenaud et al. [30] model a molecule as a graph where nodes represent atoms and edges describe chemical bonds and directly operate convolution on the graphs for molecular feature extraction. Atwood and Towsley [31] propose diffusion-convolutional neural networks to obtain the latent representations of nodes incorporating the contextual information of nodes into the graph diffusion. In contrast with spatial methods, spectral methods perform the convolution on the Laplacian spectrum of the graph. Bruna et al. [32] first employ the eigen-decomposition of graph Laplacian matrix for defining the graph convolution. Defferrard et al. [33] use the Chebyshev polynomial filters in the convolutional layer and efficiently decrease the computation complexity of eigen-decomposition. Kipf and Welling [34] present the Graph Convolutional Network (GCN) using a localized first-order approximation of spectral convolutions on graph. It encodes the relationships of nodes from the graph structure into the node features through a simple layer-wise propagation rule and thus generates more informative feature representations. Thanks to its simplicity and scalability, GCN has been successfully applied in computer vision applications, such as scene understanding [35], person re-identification [36] and action localization [37]. In the community of medical image analysis, Zhang et al. [38] take the supervoxels from the brain MRI volume as the nodes of graph and use GCN to classify supervoxel into different types of tissues. Song et al. [39] construct the brain structural connectivity graphs to design a multi-class GCN classifier for Alzheimer's disease. Zhou et al. [40] model the tissue microenvironment through a cell-graph where each node is described by a nucleus and each edge reflects cellular interactions, and then apply GCN for grading of colorectal cancer histology images. Wang et al. [41] build a graph to capture the proliferation and community structure of tumor cells and use the node features of GCN to implement the prostate tissue micro-arrays (TMA) classification. However, few work has been done to use GCN to exploit the intrinsic relationship of cervical cell images with the aim of improving the classification performance of cervical cytology.

In this paper, we propose a novel cervical cell classification method with graph convolutional network. Concretely all the CNN features of cervical cell images are firstly clustered. Consequently, the relationships of images can be revealed according to the clustering structure. However, these relationships obtained by clustering can be regarded as a preliminary exploration of relationship among images since there exists the potential correlations among clusters. Therefore, we build a graph to capture the dependencies of cluster centroids for further exploiting the latent relationship of cell images. Each node of the graph is characterized by one clustering centroid and the edges characterize the similarities of centroids. The node features and corresponding adjacency relationship of nodes are then fed into GCN for learning relation-aware representation of nodes through layer-wise message passing. The final GCN feature representations are incorporated to enhance the representational power of CNN features. The proposed method is evaluated on two cervical cell image datasets (i.e. SIPaKMeD and Motic) and experimental results on cervical cell image classification demonstrate the effectiveness of our method.

The contribution and novelty of this paper is three-folds:

1. We propose a novel cervical cell classification method based on graph convolutional network. To the best of our knowledge, this is the first to apply GCN for cervical cytology classification. The relationships of images are explored through clustering. More importantly, a graph structure is constructed to further discover the potential correlations of clusters and the relationships of nodes are encoded into node features through the layer-wise propagation rule of GCN. The relation-aware representations generated by GCN greatly improve the discriminant ability of CNN features.
2. We introduce a large-scale Motic liquid-based cytology image dataset with seven categories. It consists of two subsets where each cell image is manually annotated by experts. It is worth noting that these two subsets are respectively stained by ThinPrep and Motic medical laboratory, different from the current available cervical cytology image datasets which only have a single staining. We fully consider the influence of staining difference to cervical cell classification in practical application. Moreover, the large amount of data and some novel cell types with important clinical significance also provide a new challenge for cervical cell image analysis field.

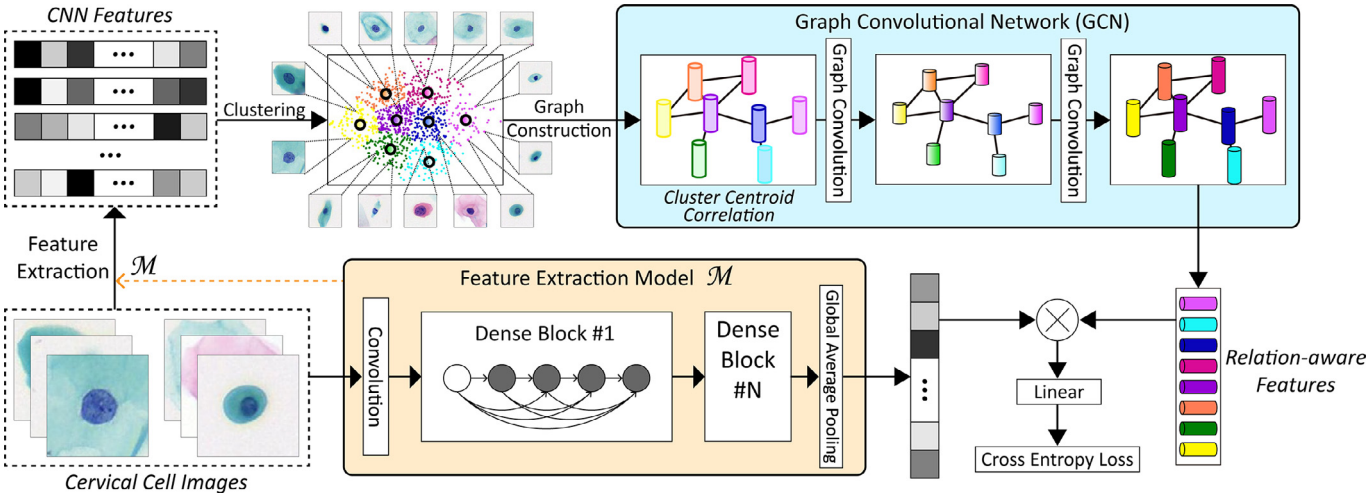

Fig. 1. Classification framework of our method for cervical cell images.

3. We conduct experiments to evaluate the proposed method and compare it with the existing state-of-the-art cervical cell classification methods on the public SIPaKMeD dataset and the proposed Motic dataset. Experimental results have demonstrated the effectiveness of our method. It can be potentially applied in automatic screening system of cervical cytology.

A preliminary version of this work has been reported [42]. In this paper, we introduce more comprehensive reviews related to cervical cytology classification and GCN, show the detailed methodology, make extensive experiments to validate the performances of the proposed method w/o GCN and the proposed method, hyperparameters determination and provide discussion. In addition, considering staining difference existed in the Motic dataset which is also common in practical application of cervical cytology screening, we also compare our method with the state-of-the-art methods under this condition and discuss the influence of staining variation to all the methods.

The remainder of this paper is organized as follows: Section 2 introduce the methodology of our method. The experimental results and discussion are presented in Sections 3 and 4. Section 5 summarizes the conclusions and future work.

2. Methodology

2.1. Overview

The cervical cell image classification framework of the proposed method is illustrated in Fig. 1. The features of cell images are firstly extracted by a CNN model pre-trained in cervical cell classification task. *k*-means clustering is performed on these CNN features and thus the cluster centroids can be gained. It can be regarded that the relationships of images have been preliminarily revealed through clustering. Considering the clusters are likely to be highly correlated, the graph of clustering centroid correlation is then built in terms of their intrinsic similarities. Actually, the potential relationships among images can be further explored through the constructed graph. Subsequently, two-layer GCN is applied to learn over the graph structure and node features, aiming to produce relation-aware features of nodes. All of these representations encoded by GCN are incorporated into the CNN features in the form of dot product. Cross-entropy loss is finally used to train the entire network after linear projection.

2.2. Graph convolutional network

Graph Convolutional Network (GCN) [34] aims to learn the relation-aware feature representations of nodes via propagating the intrinsic structure information of graph. Compared with the CNN-based methods which operate convolution on local Euclidean structure, GCN generalizes the operation of convolution to non-Euclidean data (e.g. graph). Specifically, it performs spectral graph convolution on the features of neighbor nodes, and updates the feature representation of each node incorporating the intrinsic information of graph structure during the learning process.

Given a graph $G = (V, E)$ with N nodes $v_i \in V (i = 1, \cdots, N)$, edges $(v_i, v_j) \in E$ and an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ which describes the correlations of N nodes, GCN aims to encode the graph G via a neural network model $f = (\mathbf{X}, \mathbf{A})$ where $\mathbf{X} \in \mathbb{R}^{N\times D}$ is the original features of N nodes. A multi-layer GCN updates the node features according to the following layer-wise propagation rule [34]:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{1}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ is the adjacency matrix $\mathbf{A}$ with self-connections and $\mathbf{I}_N$ is the identity matrix, and $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$ is the degree matrix. $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ is a layer-specific trainable weight matrix and $\sigma(\bullet)$ represents an activation function (e.g. ReLU). $\mathbf{H}^{(i)} \in \mathbb{R}^{d_l \times d_l}$ denotes the feature representations of nodes in the lth layer, d_l is the feature dimension and $\mathbf{H}^{(l+1)} \in \mathbb{R}^{N\times d_{l+1}}$ indicates the updated node features. Note that $\mathbf{H}^{(0)} = \mathbf{X}$. Consequently, the relation-aware feature representations of nodes can be obtained through the message-passing mechanism of the multi-layer GCN.

2.3. Relation-aware feature representation with GCN

To address the cervical cell image classification, pre-trained CNN model is firstly used to represent each cell image in our method. Considering the potential correlations among cell images may be ignored in CNN feature learning, *k*-means clustering is used to cluster all the CNN features and thus the intrinsic relationships of images can be gained. However, these relationships guided by the clustering structure only preliminarily reflect the connections of images due to the latent correlations among clusters. Therefore, we construct a graph to discover the dependencies of clusters, and then GCN is applied to encode the dependencies of clusters into the feature representations of nodes. Different from the conventional GCN [34] which achieves the node classification

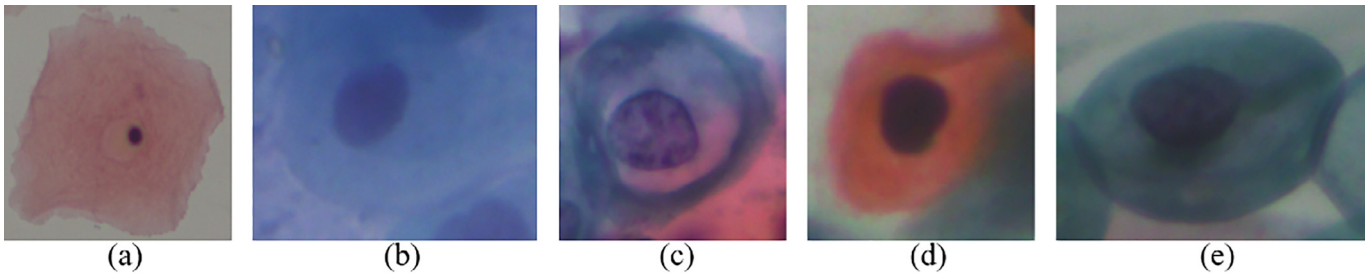

Fig. 2. Cervical cell images of 5 categories in SIPaKMeD dataset: (a) Superficial-Intermediate, (b) Parabasal, (c) Koilocytotic, (d) Dyskeratotic, (e) Metaplastic.

with the node features, the proposed method can obtain the more informative features through incorporating the GCN representation into the CNN features. It can be understood that the potential correlations among images are well preserved in the final features. More importantly, the relation-aware characteristic of GCN significantly enhances the discriminant ability of CNN features and thus improves the performance of cervical cell image classification.

Specifically, DenseNet-121 [12] pre-trained by cervical cell classification task is taken as the CNN base model in our method. Global average pooling is used to produce the image-level feature representation $\mathbf{I} \in \mathbb{R}^D$where $D = 1024$ is the feature dimension. We perform k-means clustering on the DenseNet-121 features and thus obtain the cluster centroids $\mathbf{X} \in \mathbb{R}^{N\times D}$where N is the number of centroids. Subsequently, the graph structure G is built and the corresponding adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$can be defined:

$$\mathbf{A}_{ij} = \begin{cases} 1, & if\ \mathbf{X}_i \in KNN\left(\mathbf{X}_j\right)\ \text{or}\ \ \mathbf{X}_j \in KNN(\mathbf{X}_i) \\ 0, & otherwise \end{cases} \tag{2}$$

where $KNN(\mathbf{X}_i)$ denotes the k nearest neighbors of $\mathbf{X}_i$ based on Cosine similarity. It characterizes the correlations of nodes (i.e. cluster centroids) and further explores the potential relationships among images. The node features $\mathbf{X}$ and the adjacency matrix $\mathbf{A}$ are fed into stacked two-layer GCN. Note that $\mathbf{H}^{(0)} = \mathbf{X}$ is the features of cluster centroids in our method. The node features $\mathbf{H}^{(2)} \in \mathbb{R}^{N\times D}$in the last layer of GCN can be produced through the propagation rule shown in (1). The output $\mathbf{H}^{(2)}$ can be viewed as the relation-aware feature representation generated by GCN, which is encoded by the node dependencies and the layer-wise message-passing. In the following, we will apply the relation-aware features for cervical cell classification.

2.4. Classification using relation-aware features

Given the DenseNet features of an image $\mathbf{I} \in \mathbb{R}^D$, the final feature representation of cell image $\mathbf{y}$can be obtained through dot product between the relation-aware features $\mathbf{H}^{(2)} \in \mathbb{R}^{N\times D}$and $\mathbf{I}$:

$$\mathbf{y} = \mathbf{H}^{(2)}\mathbf{I} \tag{3}$$

Afterwards, we can get the predicted scores of all the categories for the given image via linear projection and softmax function:

$$\mathbf{s} = (s_1, s_2, \cdots, s_C)^{\mathrm{T}} = \mathrm{softmax}(\mathbf{z}), \mathbf{z} = \mathbf{y}\mathbf{W} + \mathbf{b}\left(\mathbf{s}, \mathbf{z} \in \mathbb{R}^C\right) \tag{4}$$

where $s_i (i = 1, \cdots, C)$ denotes the predicted score of the ith category, C is the number of categories of cervical cells, softmax(•) is the softmax function which is used to guarantee the output scores meet the probability condition (i.e. $\sum_{i=1}^{C} s_i = 1$). $\mathbf{W} \in \mathbb{R}^{N\times C}$is the weight and $\mathbf{b}$ is the bias. The cross-entropy loss is finally used to train the entire network:

$$\mathcal{L} = -\sum_{i=1}^{C} s_i^{gt} \log\left(s_i\right) \tag{5}$$

where $\mathbf{s}^{gt} = (s_1^{gt}, s_2^{gt}, \cdots, s_C^{gt})^{\mathrm{T}}$denotes the ground-truth of the given image, especially $s_i^{gt} = 1$means the image belongs to the ith category and $s_i^{gt} = 0$otherwise.

During online inference, the DenseNet features of each cell image are firstly obtained and its corresponding GCN features are then produced by the pre-constructed graph and the trained two-layer GCN. The final feature representation of each image is generated by the incorporation of the DenseNet and the GCN features, and then the predicted category is returned through softmax.

3. Experiments

3.1. Datasets

To evaluate the classification performance of our method, the cervical cell image dataset SIPaKMeD [23] is used in this paper. It consists of 4049 images of isolated cells with 5 different categories from 966 cluster cell images of Pap smear slides. Example images of each category are displayed in Fig. 2 and the data distribution of SIPaKMeD is listed in Table 1.

In addition, the large-scale liquid-based cytology image dataset with 7 categories and 20 × objective provided by Motic is proposed for cervical cytology research in this paper and also used for the evaluation of our method. In contrast with the existing cervical cell image datasets, such as SIPaKMeD, Motic cytology image dataset contains 2 subsets, i.e., Subset-1 from 35 slides stained by ThinPrep and Subset-2 from 111 slides stained by Motic medical laboratory. Example images of each category are shown in Fig. 3. Obviously there exists staining difference in these two subsets which is a common phenomenon in practical application of cervical cytology. It inevitably poses a great challenge for cervical cell image classification. On the other hand, some novel cell types with important clinical significance (e.g. granulocyte, glandular cells, koilocytotic cells and cells with high nuclear-cytoplasmic ratio) are included. It extends the categories of cervical cell and also increases the difficulty of classification. The data distribution is given in Table 2. Clearly, the Motic dataset has larger data size than the SIPaKMeD dataset. Note that each image is 128 × 128 pixels, which is collected in this way that the centroid of segmented nucleus in a slide is taken as the center and then the area 128 × 128 around the center is captured. The ground-truth of each cell is carefully annotated by experts.

Table 1
Data distribution of SIPaKMeD dataset.

Category	Num. of Cells	Property
Superficial/Intermediate	813	Normal
Parabasal	787	
Koilocytotic	825	Abnormal
Dyskeratotic	813	
Metaplastic	793	Benign
Total	4049	

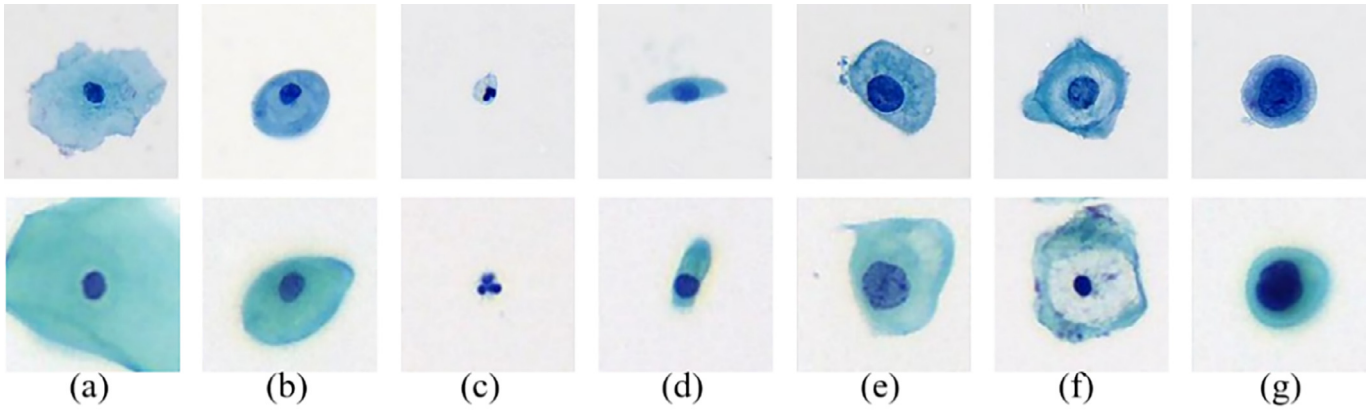

Fig. 3. Cervical cell images of 7 categories in Motic dataset: (a) Superficial squamous cells, (b) Intermediate squamous cells, (c) Granulocyte, (d) Glandular cells, (e) Atypical squamous cells (Atypical), (f) Koilocytotic cells, (g) Cells with high nuclear-cytoplasmic ratio (High-N/C-Ratio). The first row is from Subset-1 and the last one is Subset-2.

3.2. Experimental settings

Our method (denoted as Ours) is compared with 6 state-of-arts methods for cervical cell image classification, i.e., 27-dimensional cell morphological features (Mor-27), DeepPap [20], CNN features in SIPaKMeD dataset proposed by Plissiti et al. [23], ResNet-101 [11], DenseNet-121 [12] and our method w/o GCN (denoted as Ours-w/o-GCN) which only incorporates the cluster centroids and DenseNet features (i.e. $\mathbf{y} = \mathbf{H}^{(0)}\mathbf{I}$). It should be noted that Ours-w/o-GCN is used for ablation study to verify the effectiveness of GCN. Besides, Mor-27 method uses 26-dimensional cell features [23] and the perimeter of the nucleus as feature extractor, and SVM with Radial Basis Function (RBF) kernel as the classifier. It is more discriminant than 26-dimensional cell features [23] by our experiment. Each cervical cell image is represented by a 1024-dimensonal DenseNet-121 feature vector. The neighbor parameter is tuned from 10 to 100 at intervals of 10 and finally set as 40. The final GCN dimension is 1024 due to the incorporation of DenseNet-121 and GCN features. The classification accuracy, sensitivity, specificity, F-measure and confusion matrices are used for the evaluation metrics. Our model is implemented on PyTorch. The batch size is set to 32 and the network is trained for 80 epochs by stochastic gradient descent (SGD). Note that the k-means clustering is not applied to each minibatch. The cluster centroids X and the adjacency matrix A of the graph structure are obtained before GCN training. The initial learning rate is set to 0.01 and divided by 10 for every 20 epochs. All the experiments are conducted on a computer with an Intel Core i7-7820X CPU of 3.60 GHz and a GPU of NVIDIA GTX 1080Ti.

Following the preprocessing [23], each image of SIPaKMeD dataset is resized to 80 × 80 pixels, and the training set is augmented through flipping each image horizontally, vertically, and in both directions. For the fairness of the comparison in SIPaKMeD dataset, 5 training-test folds cross-validation is used for the evaluation of classification performance [20,23]. Concretely 4 of 5 folds are used as the training set and the other one as the validation set for 5 rounds. The classification evaluation metrics are obtained by averaging results from the 5 validation sets [20,23].

For the Motic dataset, we respectively conduct three experiments on Subset-1, Subset-2 and Subset-1&Subset-2 (Mixed) for verifying the classification performance of all the methods under different staining conditions. Note that 70% of cell images from each category is used as the training set and the remaining as the test set for each experiment. All the methods are evaluated on the test set.

Table 2
Data distribution of MOTIC dataset.

Category	Num. of Images of Subset-1	Num. of Images of Subset-2	Property
Superficial	2713	6872	Normal
Intermediate	364	514	
Granulocyte	5202	5527	
Glandular	36	596	
Atypical	764	1263	Abnormal
Koilocytotic	382	151	
High-N/C-Ratio	342	652	
Total	9803	15575	25378

3.3. Hyperparameters determination

Both Ours-w/o-GCN and Ours use k-means clustering to mine the relationships of image-level CNN features, so the number of cluster centroids N is likely to influence the performance of these two methods. Besides, the dimension of the first GCN layer d_1 has effect on learning the relation-aware feature representation. In our experiments, the parameter N is selected from {128, 256, 512, 1024, 2048}, and the dimension d_1 varies from {256, 512, 1024, 2048}. 5 folds cross-validation is applied for the determination of these two optimal parameters. Specifically, when one parameter is fixed, we calculate the classification accuracy (mean± std) under different values of another parameter. The parameters corresponding to the best accuracy are taken as the optimal parameters.

The 5-folds cross-validation results of Ours-w/o-GCN and Ours on the four datasets (i.e. SIPaKMeD, Motic Subset-1, Subset-2 and Mixed) are shown in Figs. 4 and 5. As can been seen from Fig. 4(a) and (b), Ours yields better classification accuracies on SIPaKMeD dataset as N and d_1 vary, but has the relatively large standard deviations. To some extent, the results of these two methods show the sensitivity to the parameters and the instability of classification performance on SIPaKMeD dataset. It may be due to the fact that the limited data size of SIPaKMeD dataset influences these two methods. Fig. 5(a)–(f) exhibit the validation results of Ours-w/o-GCN and Ours on Motic Subset-1, Subset-2 and Mixed datasets. Obviously, Ours outperforms Ours-w/o-GCN and also has the smaller standard deviations. It is likely to benefit from the large amount of Motic dataset. More importantly, Ours updates the node features through propagating the graph structure information, and thus the final node features have the relation-aware ability, while Ours w/o GCN only encodes the clustering relationships into the CNN features and fails to consider the topological relation of cluster centroids. According to the results of Figs. 4 and 5, the parameter N of

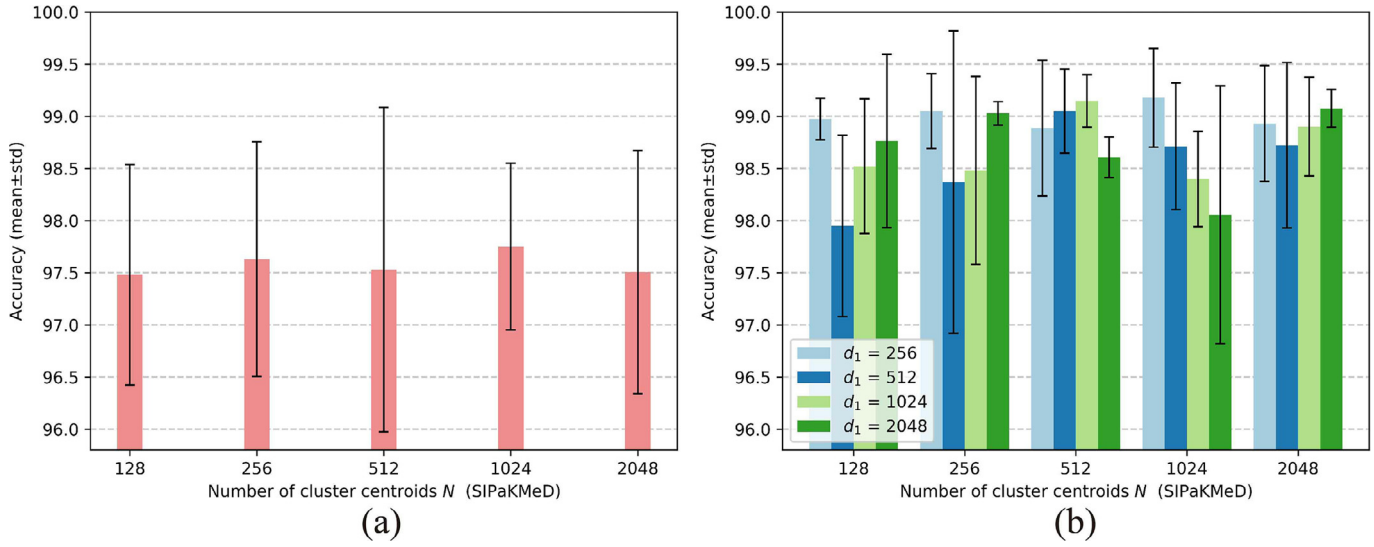

Fig. 4. Classification accuracy (mean±std) of Ours-w/o-GCN and Ours on SIPaKMeD dataset under different parameters N and d_1 where the mean of 5 folds is drawn by the column and the standard deviation is presented by the black vertical line. (a) Ours-w/o-GCN. (b) Ours.

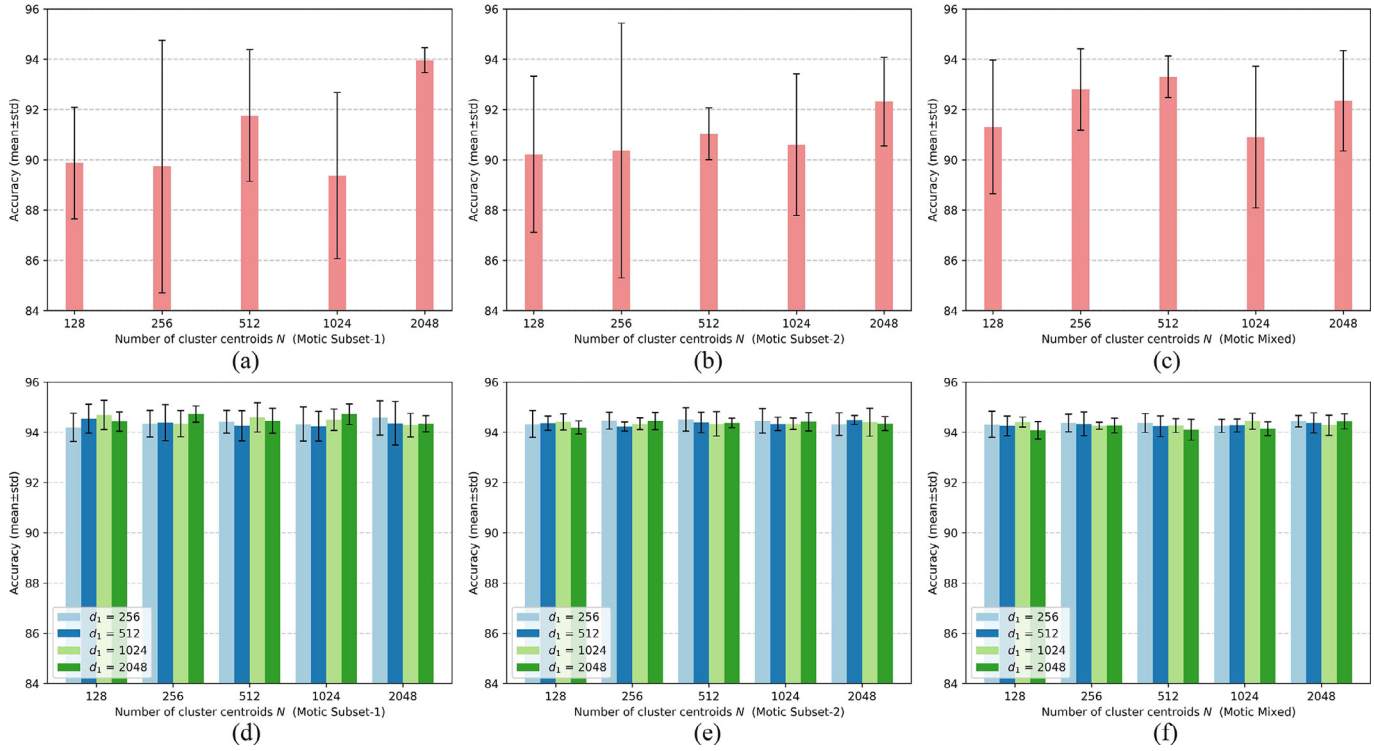

Fig. 5. Classification accuracy (mean±std) of Ours-w/o-GCN and Ours on Motic dataset under different parameters N and d_1 where the mean of 5 folds is drawn by the column and the standard deviation is presented by the black vertical line. (a) Ours-w/o-GCN on Motic Subset-1. (b) Ours-w/o-GCN on Motic Subset-2. (c) Ours-w/o-GCN on Motic Mixed. (d) Ours on Motic Subset-1. (e) Ours on Motic Subset-2. (f) Ours on Motic Mixed.

Ours-w/o-GCN on SIPaKMeD, Motic Subset-1, Subset-2 and Mixed datasets, is respectively set as {1024, 2048, 2048, 512}. The parameter N of Ours is respectively determined as {1024, 256, 512, 2048} and d_1 is set as {256, 2048, 256, 256}.

3.4. Comparison of classification performance

3.4.1. Comparison on SIPaKMeD dataset

As mentioned in Section 3.3, following the practice [20,23], after the optimal parameters N and d_1 are determined, the classification performance of all the methods are compared by averaging results from the 5 validation sets. Table 3 shows the classification results (accuracy, sensitivity, specificity, F-measure) of all the methods. Compared with Mor-27 which applies cell morphological features and SVM for classification, the CNN based methods (e.g. DeepPap [20], the method proposed by Plissiti et al. [23], ResNet-101 [11] and DenseNet-121 [12]) yield better classification results. Particularly DenseNet-121 basically outperforms DeepPap, the method proposed by Plissiti et al. and ResNet-101 due to its intrinsic network architecture. Both Ours-w/o-GCN and Ours which consider the potential relationship of images have superior classification performance compared to DenseNet-121. It can be regarded that the relationship exploration of cervical cell images contributes to the improvement of classification perfor-

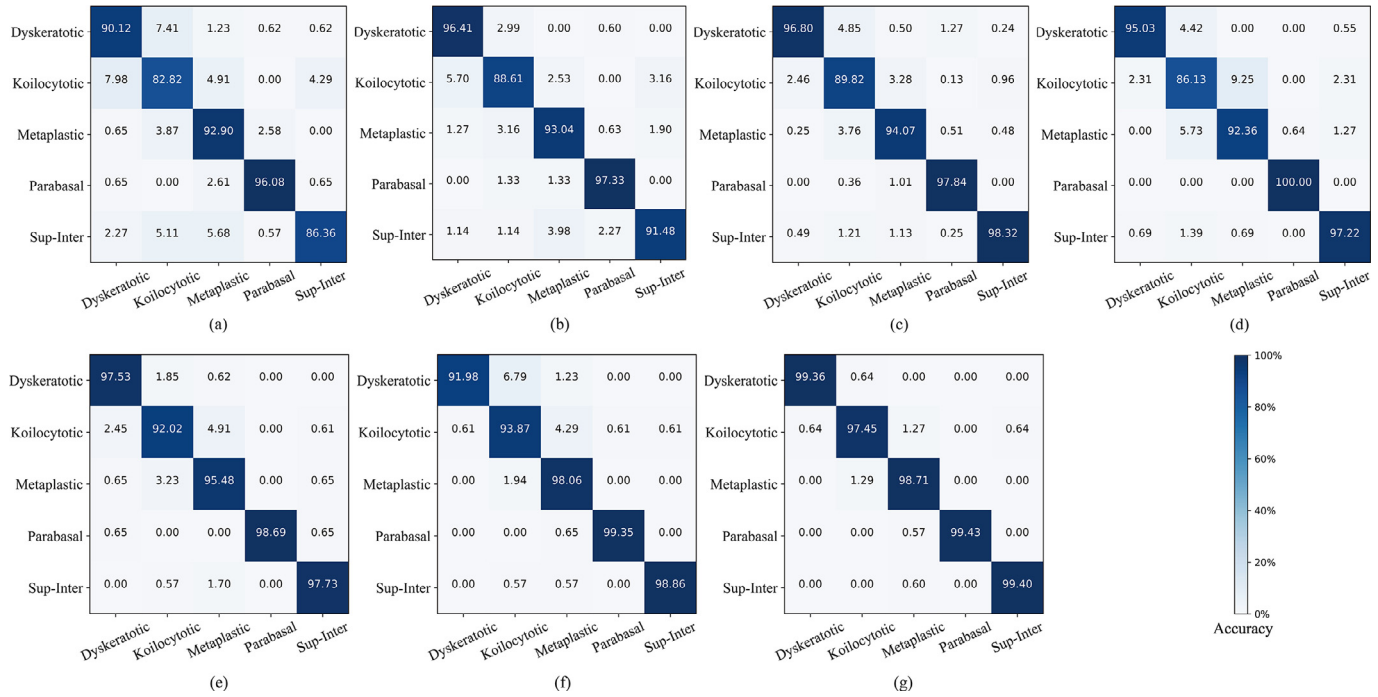

Fig. 6. Confusion matrices for classification on SIPaKMeD dataset. (a) Mor-27. (b) DeepPap. (c) Plissiti et al. (d) ResNet-101. (e) DenseNet-121. (f) Ours-w/o-GCN. (g) Ours.

mance. It should be noted that Ours is more effective than Ours-w/o-GCN, and the accuracy, sensitivity, specificity, F-measure reach 98.37%±0.57, 99.80%±0.10, 99.60%±0.20 and 99.80%±0.10. It can be explained that the intrinsic correlations of clusters can better describe the potential relationships of cell images and are encoded into the node features through the layer-wise message-passing of GCN. Therefore, the relation-aware representations generated by GCN greatly enhance the discriminative ability of CNN features.

Confusion matrices of all the methods are exhibited in Fig. 6 and they display a detailed visualization of the classification results for all the cell categories. As shown in Fig. 6(a), Mor-27 has poor classification accuracies since the variation of cellular morphology inevitably influences the representation of its morphological features. As the CNN based classification methods (e.g. DeepPap [20], the method proposed by Plissiti et al. [23], ResNet-101 [11] and DenseNet-121 [12]), the accuracies of all these methods have obviously increased. Among them, the accuracies of DenseNet-121 exceed 90% and particularly the accuracies of the abnormal cells (i.e. Koilocytotic and Dyskeratotic) reach 92.02% and 97.53% respectively, which can be seen from Fig. 6(e). However, there exists the misclassification problem for DenseNet-121. As the most challenging cell, the misclassification rate of Koilocytotic cell is 7.98% and it is most easily misclassified into Metaplastic cell (4.91%). As for the analysis of Ours-w/o-GCN shown in Fig. 6(f), the normal cells (i.e. Superficial/Intermediate and Parabasal) and the benign cell (Metaplastic) have better classification results than DenseNet-121. The accuracies of Superficial/Intermediate and Parabasal are 98.86% and 99.35%, and the accuracy of Metaplastic obtains 98.06%. Koilocytotic reaches the accuracy of 93.87% and slightly precedes the result of DenseNet-121. Its misclassification rate is 6.13% and it is still most easily incorrectly classified into Metaplastic cell (4.29%). Besides, the accuracy of Dyskeratotic is only 91.98%, which is 5.55% below DenseNet-121. It indicates that the relationship exploration of cell images indeed enhances the representational ability of CNN features. But Ours-w/o-GCN only uses the initial clustering relation for describing the relationships among images. The potential correlations existed among clusters are not fully considered, which may lead to the misclassification problem and instability of classification performance. For the results of Ours shown in Fig. 6(g), the accuracies of normal cells are more than 99% and the benign cell obtains the highest accuracy of 98.71%. It is worth noting that the results of the abnormal cells are greatly improved and the accuracies of Koilocytotic and Dyskeratotic achieve 97.45% and 99.36%. Especially the misclassification rate of Koilocytotic cell is reduced to 2.55% and its misclassification problem is effectively alleviated. It can be explained that the correlations among cluster centroids are more beneficial to characterizing the relationships of cell images and the relation-aware features generated by GCN improve the discriminative ability of CNN features.

Table 3
Comparison of classification performance on SIPaKMeD dataset (%).

Methods	Accuracy	Sensitivity	Specificity	F-measure
Mor-27	88.47± 0.92	95.90± 1.20	90.90± 1.10	95.00± 0.60
DeepPap	93.58± 0.16	97.40± 0.80	96.90± 1.40	97.60± 0.50
Plissiti et al.	95.35± 0.42	–	–	–
ResNet-101	94.86± 0.74	99.10± 0.70	97.70± 0.80	98.70± 0.20
DenseNet-121	96.79± 0.42	99.00± 0.50	98.90± 0.50	99.20± 0.10
Ours-w/o-GCN	97.75± 0.80	99.60± 0.10	99.50± 0.40	99.60± 0.20
Ours-GCN	**98.37± 0.57**	**99.80± 0.10**	**99.60± 0.20**	**99.80± 0.10**

3.4.2. Comparison on Motic dataset

In this experiment, we evaluate the classification performance of all the methods in Motic Subset-1, Subset-2 and Mixed datasets under two conditions of the consistent staining and different staining. The classification results on the test sets corresponding to these three datasets are listed in Table 4. It can be easily found that Ours has achieved the best accuracy, sensitivity, specificity and F-measure on the Subset-1 and Subset-2 datasets, and when the training data grow, our method has obvious classification advantage. Specifically, the accuracy, sensitivity, specificity, F-measure of our method in Subset-1 stained by ThinPrep reach 94.23%, 92.41%, 98.76% and 92.72%. For the Subset-2 stained by Motic medical laboratory, the accuracy, sensitivity, specificity, F-measure of our method are respectively 95.17%, 93.24%, 99.24% and 94.07%.

Table 4
Comparison of classification performance on Motic dataset (%).

Motic dataset	Methods	Accuracy	Sensitivity	Specificity	F-measure
Subset-1	Mor-27	86.66	82.38	98.69	86.85
	DeepPap	90.83	85.71	98.24	87.67
	Plissiti et al.	90.90	87.50	98.32	88.89
	ResNet-101	91.10	88.84	98.48	90.05
	DenseNet-121	92.87	91.74	98.64	92.05
	Ours-w/o-GCN	93.78	90.62	98.92	92.17
	Ours	**94.23**	**92.41**	**98.76**	**92.72**
Subset-2	Mor-27	82.28	85.52	98.37	88.12
	DeepPap	92.62	90.18	98.64	90.61
	Plissiti et al.	92.13	87.92	98.72	89.58
	ResNet-101	92.71	87.28	99.04	90.18
	DenseNet-121	93.82	91.14	99.19	92.79
	Ours-w/o-GCN	94.18	94.20	98.74	93.08
	Ours	**95.17**	**93.24**	**99.24**	**94.07**
Mixed	Mor-27	82.56	75.27	98.72	82.64
	DeepPap	91.50	84.00	98.63	87.31
	Plissiti et al.	91.48	88.03	98.50	89.28
	ResNet-101	91.69	86.16	98.82	89.11
	DenseNet-121	93.61	91.39	**99.07**	92.74
	Ours-w/o-GCN	93.33	90.74	98.89	91.86
	Ours	**94.93**	**92.98**	98.84	**92.94**

The superior classification performance of Ours is likely to benefit from two reasons. On the one hand, the potential relationships among images are better captured through the intrinsic correlations of clusters. More importantly, the underlying dependencies of nodes are propagated through the message-passing mechanism of GCN and thus the relation-aware features produced by GCN effectively improve the classification ability of CNN features. Moreover, Ours-w/o-GCN is overall better than DeepPap, the method proposed by Plissiti et al., ResNet-101 and DenseNet-121. It may be explained that Ours-w/o-GCN results from exploiting the relationships among images guided by the preliminary clustering. When the Motic Mixed dataset (i.e. Subset-1&Subset-2) is used for evaluation, it can be seen that all the methods are not improved as expected when the amount of training data increases. It indicates that the staining difference has an important influence on the classification performance. Under this condition, the accuracy, sensitivity, specificity, F-measure of our method obtain 94.93%, 92.98%, 98.84% and 92.94%. The overall classification results are superior to other methods and the specificity is comparable to DenseNet-121 (99.07%). Compared with DeepPap, the method proposed by Plissiti et al. and ResNet-101, Ours-w/o-GCN has better classification results but lower than DenseNet-121 and Ours. It shows that the staining difference affects the clustering structure and may cause the cells with similarity are viewed as they are from different clusters. Therefore, the relationships among images in Ours-w/o-GCN may be not well exploited and the generated feature representations are likely to be not discriminative. Nevertheless, our method captures the correlations of clusters, which can be regarded as a further exploration of the intrinsic relationships among images. More critically, our method encodes these correlations into the features of images through the layer-wise propagation of GCN, which improves the classification performance and also enhances the robustness to staining difference.

Confusion matrices of all the methods on Motic Subset-1, Subset-2 and Mixed datasets are illustrated in Figs. 7–9. As the results of Subset-1 shown in Fig. 7, our method outperforms other methods. To be specific, the accuracies of normal cells (i.e. Superficial, Intermediate, Granulocyte and Glandular) for our method shown in Fig. 7(g) are respectively 93.00%, 93.64%, 98.01% and 63.64%. The result of Glandular cell is unsatisfactory due to its limited data size. The abnormal cells (i.e. Atypical, Koilocytotic, High-N/C-Ratio) have the accuracies of 87.39%, 79.13% and 82.52%, which are higher than other methods. Obviously, the results of abnormal cells are inferior to Superficial, Intermediate and Granulocyte cells. It may be explained that the abnormal cells have the relatively high interclass similarity and thus increase the difficulty for classification. Especially, similar to the results on SIPaKMeD dataset, Koilocytotic cell is still the most challenging category. The misclassification rate is 20.87% and it is easily classified into Superficial (10.43%), Atypical (8.70%) and Granulocyte (1.74%). The same situation goes for other methods, and the misclassification rate of Koilocytotic for DenseNet-121 and Ours-w/o-GCN are respectively 26.96% and 23.48%, as presented in Fig. 7(e) and (f).

As displayed in Fig. 8, we can find that our method has more desirable classification results for all the categories in Subset-2 compared to other methods. Superficial, Intermediate, Granulocyte and Glandular cells have the accuracies of 97.48%, 76.13%, 98.13% and 88.83%, and Atypical, Koilocytotic and High-N/C-Ratio cells have the accuracies of 91.03%, 52.17% and 84.69%, which can be seen from Fig. 8(g). Clearly as the training data increase, the results for most of categories have been improved in Subset-2 except for Intermediate and Koilocytotic cells. Note that Intermediate cell in Subset-2 has the lower accuracy than the result in Subset-1, and it is most easily classified into Superficial (14.19%). It may because Intermediate cells of Subset-2 stained by Motic medical laboratory have strong similarity with Superficial cells and thus the small interclass dissimilarity leads to the misclassification problem. Moreover, Koilocytotic cells reach the lowest accuracy due to the small training data in Subset-2.

When the Motic Mixed dataset (i.e. Subset-1&Subset-2) is applied, we can see that the results of all the methods are not greatly improved as expected as shown in Fig. 9. It verifies that the staining difference influences the classification performance. Under this situation, our method has better classification results for all the categories than other methods. Concretely Superficial, Intermediate, Granulocyte and Glandular cells reach the accuracies of 96.24%, 86.04%, 97.86% and 86.94%, and the abnormal cells (i.e. Atypical, Koilocytotic, High-N/C-Ratio) respectively obtain the accuracies of 88.51%, 78.26% and 85.95%, as presented in Fig. 8(g). It is worth noting that compared with the results of Ours in Subset-2, the accuracies of Intermediate and Koilocytotic cells significantly raise. However, it is lower than the results of Ours in Subset-1. It may be due to that the increase of training data leads to the improvement of classification accuracy to a great extent, but the staining variation limits the promotion of performance as expected. Besides, the misclassification of Koilocytotic is not completely solved and it is

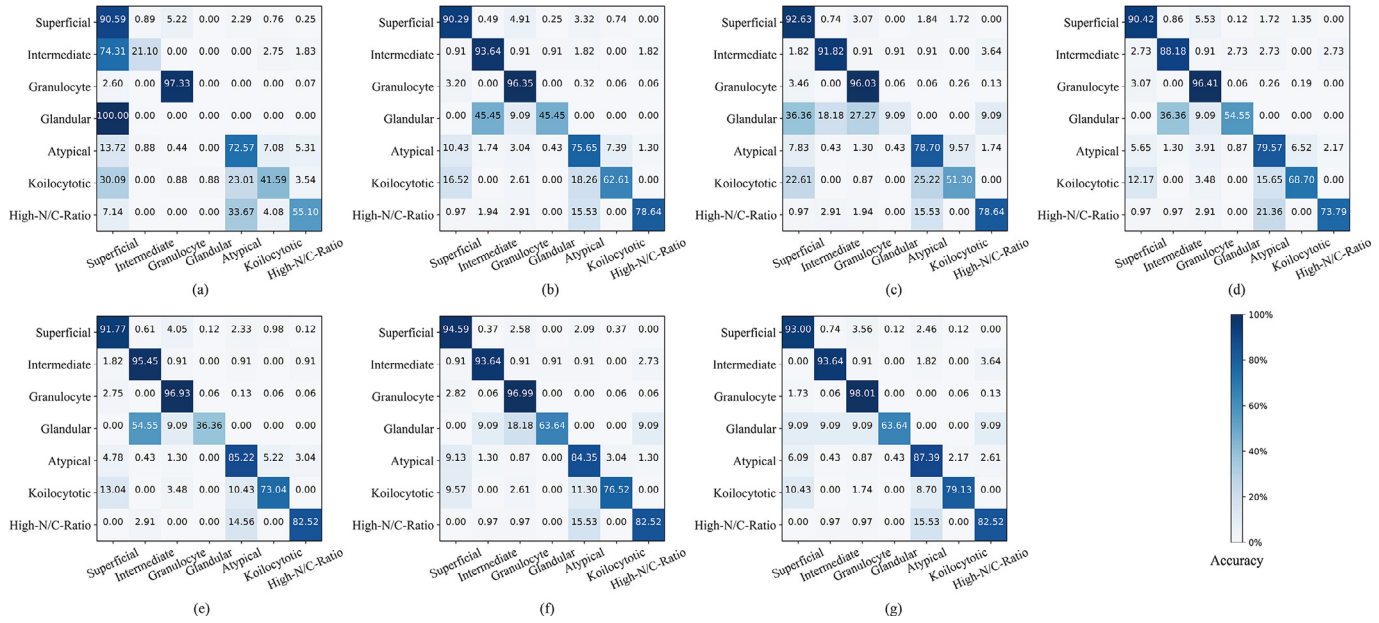

Fig. 7. Confusion matrices for classification on Motic Subset-1 dataset. (a) Mor-27. (b) DeepPap. (c) Plissiti et al. (d) ResNet-101. (e) DenseNet-121. (f) Ours-w/o-GCN. (g) Ours.

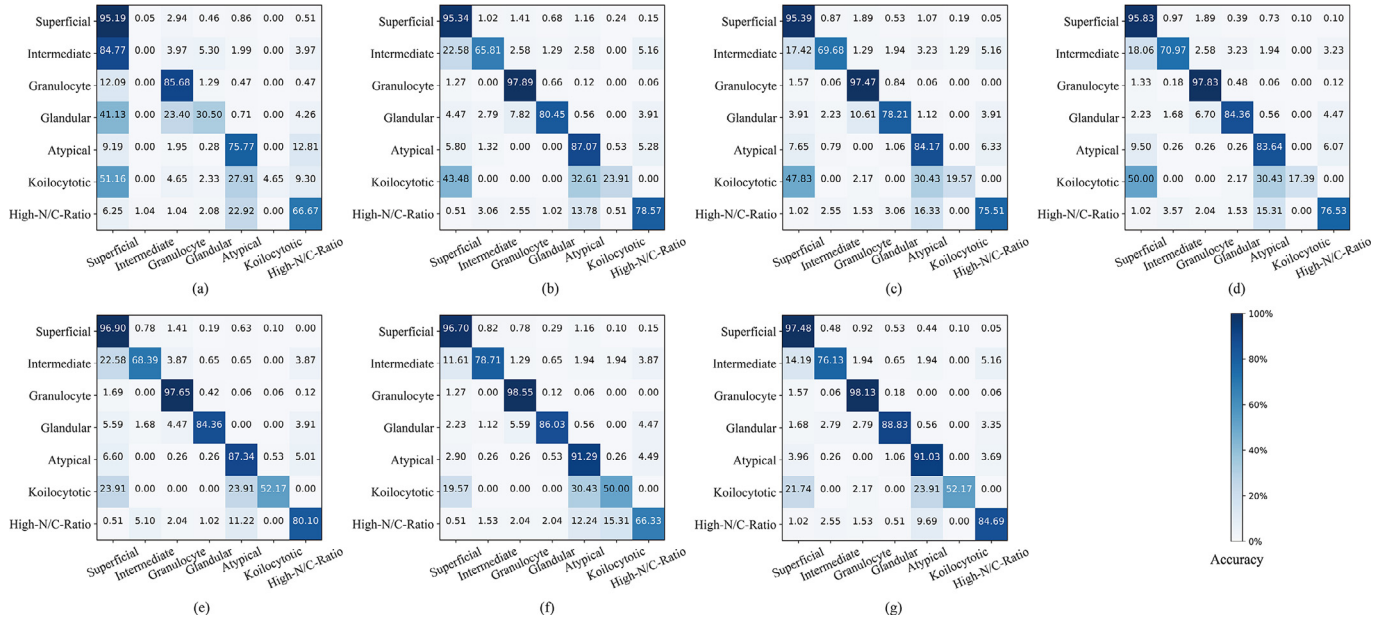

Fig. 8. Confusion matrices for classification on Motic Subset-2 dataset. (a) Mor-27. (b) DeepPap. (c) Plissiti et al. (d) ResNet-101. (e) DenseNet-121. (f) Ours-w/o-GCN. (g) Ours.

still easily classified into Atypical (11.80%), Superficial (9.32%) and Granulocyte (0.62%). Compared with the results of DenseNet-121 (61.49%) and Ours-w/o-GCN (66.46%) shown in Figs. 9(e) and 9(f), the accuracy of Koilocytotic in our method has been effectively improved. It indicates that the feature representation generated by our method is more discriminative.

4. Discussion

Cervical cell classification is a challenging issue for automatic screening of cervical cytology. The performance of the classification method determines whether it can bring convenience to cytoscreener and woman. As the current mainstream classification framework, Convolutional Neural Network (CNN) uses the hierarchical deep architecture to automatically learn high-level features. It has been gradually applied in cervical cytology and is proved to be more effective than traditional cell morphological features. As shown in Tables 3 and 4, the CNN-based methods (e.g. DeepPap [20], the method proposed by Plissiti et al. [23], ResNet-101 [11] and DenseNet-121 [12]) obviously outperform Mor-27. Moreover, with the continuous optimization of network architecture, the optimized CNN model exhibits the better classification performance, such as DenseNet. However, the accuracies of these CNN-based methods for all the cell categories are not desirable, as presented in Figs. 6(b)-(e), 7(b)-(e), 8(b)-(e) and 9(b)-(e). It indicates that the CNN features may be not optimal and we need to make some improvements on CNN features.

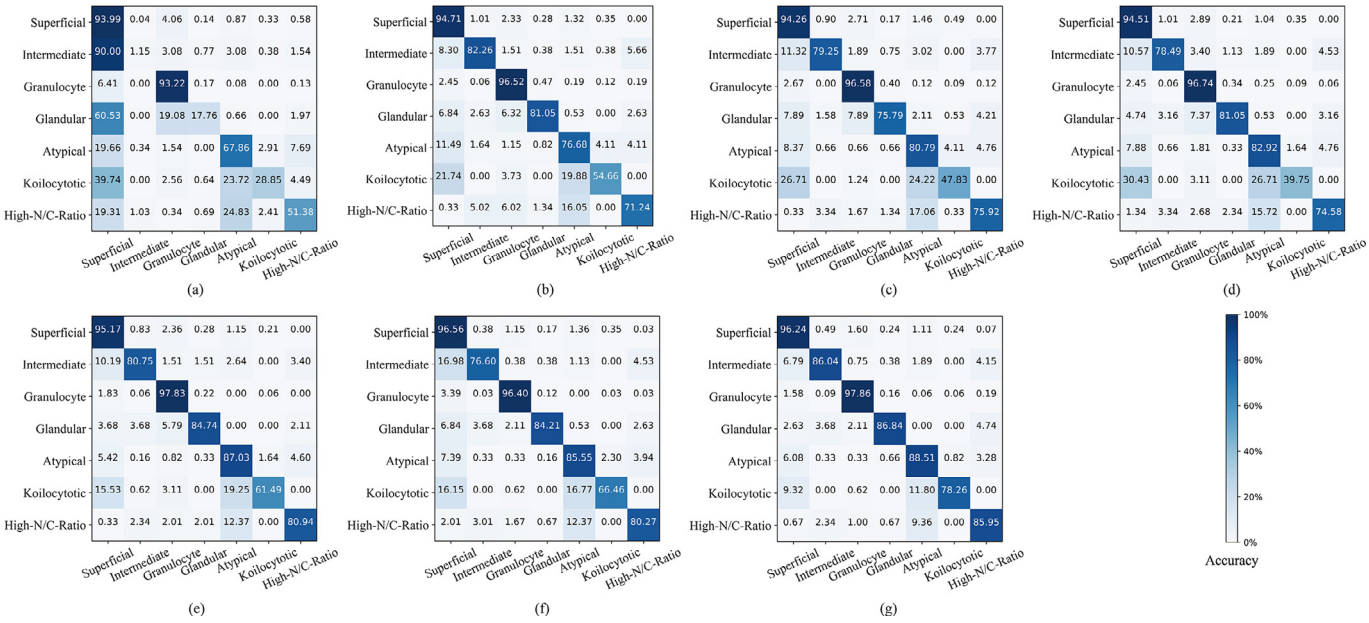

Fig. 9. Confusion matrices for classification on Motic Mixed dataset. (a) Mor-27. (b) DeepPap. (c) Plissiti et al. (d) ResNet-101. (e) DenseNet-121. (f) Ours-w/o-GCN. (g) Ours.

Considering the potential relationships among cervical cell images are ignored during CNN feature learning, we aim to exploit the relationships of cell images for enhancing the discriminative ability of CNN features. The clustering is the first way to describe the relations of images. However, there exists the intrinsic correlations among clusters and thus causes the relations guided by the clustering may be not accurate. Therefore, we use the graph structure to further explore the dependencies of clusters. More importantly, Graph Convolutional Network (GCN) [34] is used to encode these dependencies into node features with the aim of learning relation-aware feature representation. As illustrated in Tables 3 and 4, our method has the best classification results and Ours-w/o-GCN is overall superior to DenseNet. It demonstrates that the relationship exploration of cell images is conductive to classification and the relation-aware ability of GCN can better improve the representational power of CNN features.

For the proposed large-scale Motic liquid-based cytology image dataset, we can find that it extends the categories of the normal and abnormal cells and fully considers the staining difference which is common in practical reading procedure for cytoscreeners. As shown in Table 4 and Fig. 8, it can be seen that all the methods are affected by staining difference and they are not greatly improved as expected as the training data increase. Therefore, future work will focus on eliminating the influence of staining difference to cervical cell classification methods. Meanwhile we notice that the accuracies of abnormal cells (i.e. Atypical, Koilocytotic, High-N/C-Ratio) are not satisfactory. The limited data size may restrict the classification performance for these categories compared with the amount of Superficial and Granulocyte cells. However, it may be unrealistic to spend a lot of manpower collect and annotate the abnormal cells. Therefore, another work will concentrate on using self-supervised learning to define a suitable annotation-free pretext task for alleviating the strong demand of annotated samples.

Recently, some semi-supervised learning methods are applied for histopathology image [43,44]. Inspired by these studies, we notice that our method can also be used in semi-supervised learning. To evaluate the semi-supervised performance of our method, we collected another 2600 unlabeled cervical cell images from Motic medical laboratory and added these cell images into Motic Mixed dataset. We used the CNN model pre-trained in cervical cell classification task to extract the cell features, and then performed the k-means clustering on all the training samples including the extended unlabeled samples. Then, the cluster centroids $\mathbf{X}$ and the adjacency matrix $\mathbf{A}$ of the graph structure are fed into the two-layer GCN for learning the relation-aware features of nodes. Finally, the features encoded by GCN were incorporated into the CNN features and the whole network was trained using the labeled samples. The experimental results are shown in Table 5. It can be seen that the classification performance of our method is improved when the additional unlabeled samples are considered. It indicates that the k-means clustering module has found a more reasonable underlying relationship of the cell features with the help of the unlabeled samples. It has certified the advantage of using the k-means clustering to generate the graph structure.

Moreover, we have tried to apply the proposed method for practical cervical cytology screening. The candidate cells in region of interest (ROI) or whole slide imaging (WSI) were firstly detected, then represented by the bounding boxes, and finally classified by our model. As shown in Fig. 10, most of detected cells in ROIs are correctly classified by our method. It demonstrates our method is potential to be applied in automatic screening system of cervical cytology. On the other hand, we note that there exist cells undetected and thereby misclassified during the screening. It indicates that both the detection and classification are important in practical screening system. The future work will focus on combining self-supervised learning or other learning methods (e.g. attention) to improve the capacity and generalization of our method. We have also reviewed the development history of automatic cervical cancer screening system, which can help us better understand how to apply the artificial intelligence technology for clinical implementation. As the first commercially available automatic screening system, PAPNET [45,46] uses neural network to select 128 digitized images of potentially abnormal cells or clusters. If any of the 128 selected images contains significant abnormalities to the cytoscreener, the corresponding slide needs further microscopic review. It is demonstrated that PAPNET increases the sensitivity and effectively reduces the time and cost of false-negative slide screening [45,46] compared to the manual screening. More importantly, it indicates that the computer technology contributes to assisting

Table 5
Semi-supervised classification performance of our method on Motic dataset (%).

Num. of images	Accuracy	Sensitivity	Specificity	F-measure
25378	94.93	**92.98**	98.84	92.94
25378 + 2600 (unlabeled)	**95.08**	92.14	**99.25**	**93.68**

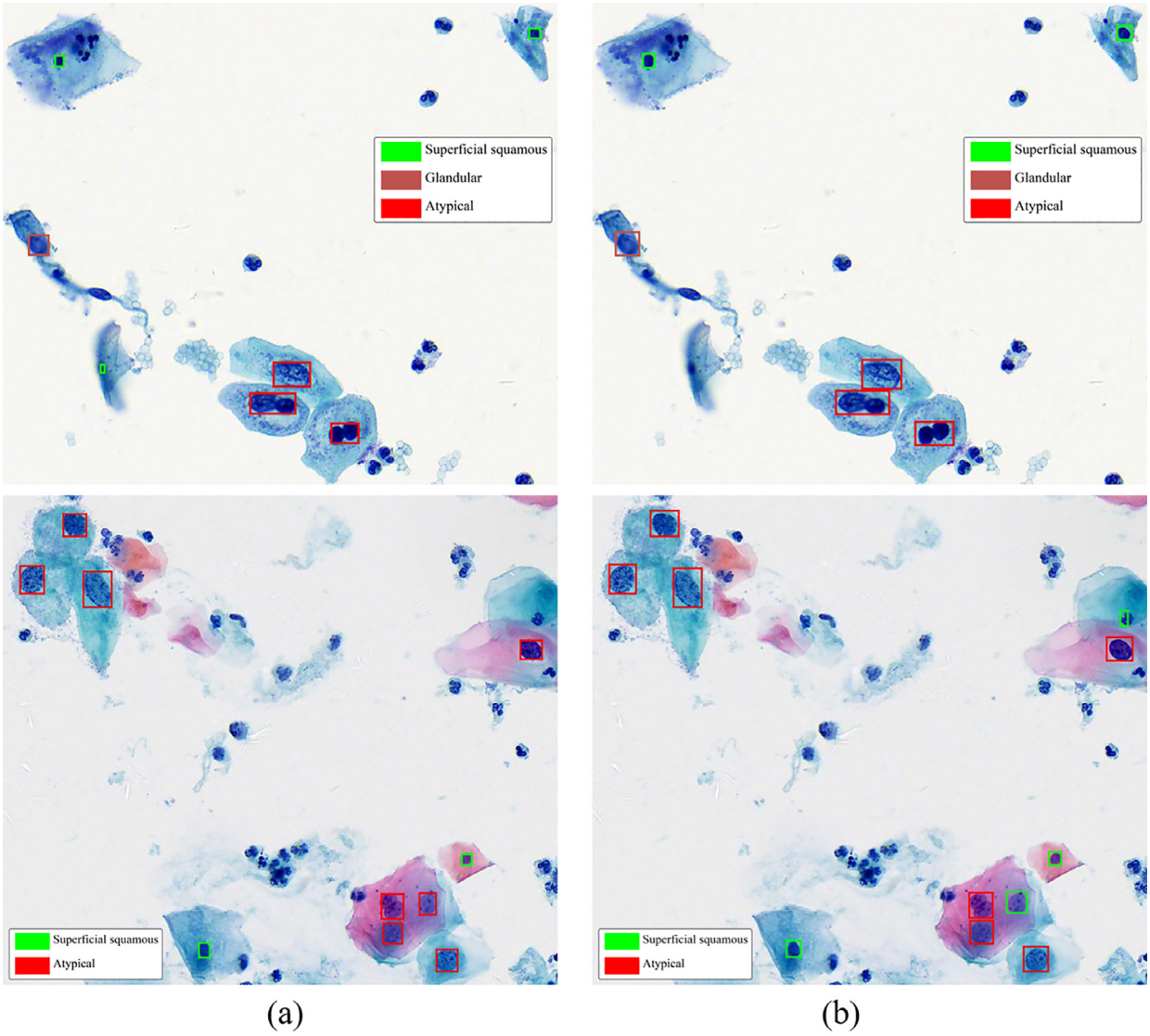

Fig. 10. Classification results of our method for ROIs. (a) Ground truth. (b) Our method.

manual screening. However, the recognition performance of PAPNET is limited by the level of technology at that time. Following the PAPNET, the ThinPrep Imaging System (Imager) designed by Hologic has been widely commercially applied in ThinPrep liquid-based cytology slides. The Imager uses the algorithm based on cellular morphological features and nuclear darkness to recognize the diagnostically relevant cells or cell groups [47,48]. Unlike the PAPNET, the Imager scans the whole slide, identifies 22 fields of view (FOV) which contain cells of concern recognized by the Imager algorithm, and stores their corresponding coordinates [47,48]. If all the 22 fields are judged normal by the cytoscreener, the corresponding slide can be viewed as negative without further screening. If any of the fields has suspicious cells, the entire slide needs to be manually reviewed. Compared with the PAPNET, the Imager further enhances the screening efficiency and raises the sensitivity. Particularly it has achieved the better identification performance of low-grade Squamous Intraepithelial Lesion (LSIL) and high-grade squamous intraepithelial lesion (HSIL) [48]. With the recent development of CNN based image analysis algorithms, it offers an opportunity to further enhance the performance of cervical cell classification method. Our attempts on applying CNN for cervical cell classification demonstrate the rapid improvement in artificial intelligence will make the automatic cervical cancer screening system achieve the full potential.

5. Conclusion

In this paper, a novel cervical cell classification method with Graph Convolutional Network (GCN) is developed. Our method firstly exploits the relationships of cell images through clustering and then constructs a graph where the cluster centroids are modeled as the nodes for further capturing the potential correlations

among clusters. More critically, GCN is applied to encode the correlations into node features and thus generate the relation-aware representations. They are finally incorporated into CNN features for improving the representational power of features. Therefore, our method not only effectively considers the potential relationships of images, but also yields more informative feature representations. Furthermore, the large-scale Motic liquid-based cytology image dataset is also proposed. The large data size, some novel cell types with important clinical significance and staining variation pose a new challenge for cervical cytology field. Experiments on the public SIPaKMeD dataset and the proposed Motic dataset demonstrate the proposed method has better classification results and more robustness to staining difference compared with existing state-of-arts cervical cell classification methods.

In future work, we will focus on stain standardization of cervical cytology for erasing the effect of staining difference to classification methods. Additionally, considering the large amount of unlabeled cell image in the practical application, we will try to combine GCN and self-supervised learning for improving the cell classification performance and reducing the annotation cost.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (grant no. 61906058, 61901018, 61771031 and 61471016), partly supported by the Anhui Provincial Natural Science Foundation (grant no. 1908085MF210), partly supported by the China Postdoctoral Science Foundation (grant no. 2019M650446), and partly supported by the Fundamental Research Funds for the Central Universities of China (grant no. JZ2020YYPY0093).

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, CA Cancer J. Clin. 69 (2019) 7–34.
[2] H.C. Kitchener, R. Blanks, G. Dunn, L. Gunn, M. Desai, R. Albrow, J. Mather, D.N. Rana, H. Cubie, C. Moore, R. Legood, A. Gray, S. Moss, Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial, Lancet Oncol. 12 (2011) 56–64.
[3] L. Zhang, H. Kong, C.T. Chin, S. Liu, X. Fan, T. Wang, S. Chen, Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining, Cytometry Part A 85 (2014) 214–230.
[4] W. William, A. Ware, A.H. Basaza-Ejiri, J. Obungoloch, A review of image analysis and machine learning techniques for automated cervical cancer screening from Pap-smear images, Comput. Methods Programs Biomed. 164 (2018) 15–22.
[5] T. Chankong, N. Theera-Umpon, S. Auephanwiriyakul, Automatic cervical cell segmentation and classification in Pap smears, Comput. Methods Programs Biomed. 113 (2014) 539–556.
[6] K. Bora, M. Chowdhury, L.B. Mahanta, M.K. Kundu, A.K. Das, Automated classification of Pap smear images to detect cervical dysplasia, Comput. Methods Programs Biomed. 138 (2017) 31–47.
[7] M. Arya, N. Mittal, G. Singh, Texture-based feature extraction of smear images for the detection of cervical cancer, IET Comput. Vision 12 (2018) 1049–1059.
[8] P. Wang, L. Wang, Y. Li, Q. Song, S. Lv, X. Hu, Automatic cell nuclei segmentation and classification of cervical Pap smear images, Biomed. Signal Process. Control 48 (2019) 93–103.
[9] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097-1105.
[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015.
[11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[12] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.
[13] K. He, G. Gkioxari, P. Dollár, R. Girshick, R-CNN Mask, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
[14] Y. Zheng, Z. Jiang, F. Xie, H. Zhang, Y. Ma, H. Shi, Y. Zhao, Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification, Pattern Recognit. 71 (2017) 14–25.
[15] F. Ciompi, O. Geessink, B.E. Bejnordi, G.S. de Souza, A. Baidoshvili, G. Litjens, B. van Ginneken, I. Nagtegaal, J. van der Laak, The importance of stain normalization in colorectal tissue classification with convolutional networks, in: Proceedings of the IEEE 14th International Symposium on Biomedical Imaging(ISBI), 2017, pp. 160–163.
[16] C. Li, X. Wang, W. Liu, L.J. Latecki, DeepMitosis: mitosis detection via deep detection, verification and segmentation networks, Med. Image Anal. 45 (2018) 121–133.
[17] T. Song, V. Sanchez, H.E. Daly, N.M. Rajpoot, Simultaneous cell detection and classification in bone marrow histology images, IEEE J. Biomed. Health Inform. 23 (2019) 1469–1476.
[18] T. Qaiser, N.M. Rajpoot, Learning where to see: a novel attention model for automated immunohistochemical scoring, IEEE Trans. Med. Imaging 38 (2019) 2620–2631.
[19] B. Hu, Y. Tang, E.I. Chang, Y. Fan, M. Lai, X. Yan, Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks, IEEE J. Biomed. Health Inform. 23 (2019) 1316–1328.
[20] L. Zhang, L. Lu, I. Nogues, R.M. Summers, S. Liu, J. Yao, DeepPap: deep convolutional networks for cervical cell classification, IEEE J. Biomed. Health Inform. 21 (2017) 1633–1643.
[21] B. Taha, J. Dias, N. Werghi, Classification of cervical-cancer using Pap-smear images: a convolutional neural network approach, in: Proceedings of the Medical Image Understanding and Analysis, 2017, pp. 261–272.
[22] H. Wieslander, G. Forslid, E. Bengtsson, C. Wählby, S.K. Sadanandan, Deep convolutional neural networks for detecting cellular changes due to malignancy, in: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 82–89.
[23] M.E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, A. Charchanti, SIPaKMeD: a new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images, in: Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3144–3148.
[24] S. Gautam, K.K. Harinarayan, N. Jith, A.K. Sao, A. Bhavsar, A. Natarajan, Considerations for a Pap smear image analysis system with CNN features, arXiv:1806.09025 (2018).
[25] N. Sompawong, J. Mopan, P. Pooprasert, W. Himakhun, C. Tantibundhit, Automated pap smear cervical cancer screening using deep learning, in: Proceedings of 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 7044–7048.
[26] K. Kiran GV, G.M. Reddy, Automatic classification of whole slide Pap smear images using CNN with PCA based feature interpretation, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1074–1079.
[27] Y. Zhou, H. Chen, J. Xu, Q. Dou, P.-A. Heng, IRNet: instance relation network for overlapping cervical cell segmentation, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), 2019, pp. 640–648.
[28] T. Conceição, C. Braga, L. Rosado, M.J.M. Vasconcelos, A review of computational methods for cervical cells segmentation and abnormality classification, Int. J. Mol. Sci. 20 (2019) 5114.
[29] K. Kim, B. Naylor, Practical Guide to Surgical Pathology with Cytologic Correlation, Springer, New York, 1992.
[30] D.K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, R.P. Adams, Convolutional networks on graphs for learning molecular fingerprints, in: Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS), 2015, pp. 2224–2232.
[31] J. Atwood, D. Towsley, Diffusion-convolutional neural networks, in: Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS), 2016, pp. 1993–2001.
[32] J. Bruna, W. Zaremba, A. Szlam, Y. Lecun, Spectral networks and locally connected networks on graphs, in: Proceedings of the International Conference on Learning Representations (ICLR), 2014.
[33] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS), 2016, pp. 3844–3852.
[34] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the International Conference on Learning Representations (ICLR), 2017.
[35] J. Yang, J. Lu, S. Lee, D. Batra, D. Parikh, Graph R-CNN for scene graph generation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 690–706.
[36] Y. Shen, H. Li, S. Yi, D. Chen, X. Wang, Person re-identification with deep similarity-guided graph neural network, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 508–526.
[37] R. Zeng, W. Huang, M. Tan, Y. Rong, P. Zhao, J. Huang, Graph convolutional networks for temporal action localization, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, pp. 7093–7102.
[38] Y. Zhang, Y. Kong, J. Wu, G. Coatrieux, H. Shu, Brain tissue segmentation based on graph convolutional networks, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2019, pp. 1470–1474.

[39] T. Song, S.R. Chowdhury, F. Yang, H.I.L. Jacobs, G.E. Fakhri, Q. Li, K.A. Johnson, J. Dutta, Graph convolutional neural networks for Alzheimer's disease classification, in: Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI), 2019, pp. 414–417.
[40] Y. Zhou, S. Graham, N.A. Koohbanani, M. Shaban, P.-A. Heng, N. Rajpoot, CGC-net: cell graph convolutional network for grading of colorectal cancer histology images, in: Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 388–398.
[41] J. Wang, R.J. Chen, M.Y. Lu, A.S. Baras, F. Mahmood, Weakly supervised prostate TMA classification via graph convolutional networks, arXiv:1910.13328 (2019).
[42] J. Shi, R. Wang, Y. Zheng, Z. Jiang, L. Yu, Graph convolutional networks for cervical cell classification, MICCAI Workshop on Computational Pathology, 2019 [Online]. Available https://openreview.net/forum?id=S1gX_tlc-S .
[43] H. Su, X. Shi, J. Cai, L. Yang, Local and global consistency regularized mean teacher for semi-supervised nuclei classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2019, pp. 559–567.
[44] X. Shi, H. Su, F. Xing, Y. Liang, G. Qu, L. Yang, Graph temporal ensembling based semi-supervised convolutional neural network with noisy labels for histopathology image analysis, Med. Image Anal. 60 (2020) 101624.
[45] L.G. Koss, E. Lin, K. Schreiber, P. Elgert, L. Mango, Evaluation of the PAPNET cytologic screening system for quality control of cervical smears, Am. J. Clin. Pathol. 101 (1994) 220–229.
[46] N. Bosanquet, D.V. Coleman, C.J. Dore, G. Douglas, L.J. Magee, Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial, Lancet North Am. Ed. 353 (1999) 1381–1385.
[47] ThinPrep® review scope manual+. 2015. URL: https://www.hologic.com/sites/default/files/package-insert/MAN-03695-001_002_02.pdf.
[48] F.S. Miller, L.E. Nagel, M.B. Kenny-Moynihan, Implementation of the ThinPrep imaging system in a high-volume metropolitan laboratory, Diagn. Cytopathol. 35 (2007) 213–217.