IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 22, 2018, accepted September 11, 2018, date of publication September 19, 2018, date of current version October 17, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871153

Automatic Segmentation of Cervical Nuclei Based on Deep Learning and a Conditional Random Field

YIMING LIU, PENGCHENG ZHANG, QINGCHE SONG, ANDI LI, PENG ZHANG, AND ZHIGUO GUI

Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China

Corresponding author: Zhiguo Gui (gzgtg@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11605160, in part by the Research Project through the Shanxi Scholarship Council of China under Grant 2016-089, in part by the Fund for Shanxi 1331 Project Key Innovative Research Team, in part by the National Key Scientific Instrument and Equipment Development Project of China under Grant 2014YQ24044508, in part by the National Natural Science Foundation of China under Grant 61671413, and in part by the National Key Research and Development Program of China under Grant 2016YFC0101605.

ABSTRACT Automatic and accurate cervical nucleus segmentation is important because nuclei carry substantial diagnostic information for automatic computer-assisted cervical cancer screening and diagnosis systems. In this paper, we propose a cervical nucleus segmentation method in which pixel-level prior information is utilized to provide the supervisory information for the training of a mask regional convolutional neural network (Mask-RCNN), which is then employed to extract the multi-scale features of the nuclei, and the coarse segmentation and bounding box of the nuclei are obtained by forward propagation of the Mask-RCNN. To refine the segmentation, a local fully connected conditional random field (LFCCRF) that contains unary and pairwise energy terms is employed. The nuclear region of interest is determined by extending the bounding box, the coarse segmentation in the nuclear region is used to construct the unary energy, and the pairwise energy is contributed by the position and intensity information of all of the pixels in the nuclear region. By minimizing the energy of the LFCCRF, the final segmentation is realized. We evaluated our method by using cervical nuclei from the Herlev Pap smear data set in this paper, and the precision, recall, and Zijdenbos similarity index were all found to be greater than 0.95 with low standard deviations, demonstrating that our method enables more accurate and stable cervical nucleus segmentation than the current state-of-the-art methods.

INDEX TERMS Conditional random field, deep learning, Mask-RCNN, Pap smear screening.

I. INTRODUCTION

Cervical cancer is one of the most common causes of cancer death among women worldwide, with nearly 0.27 million deaths every year. Most of these deaths occur in low- to middle-income countries due to the financial investments required to establish and maintain the necessary health infrastructure, including laboratory and skilled human resources, which are unavailable or insufficient in many settings. In contrast, in high-income countries, where Pap smears have been used for population-based cytology screening for more than three decades, cervical cancer morbidity and mortality are much lower [1]. Pap smear testing was proposed by George Papanicolaou in 1942 [2] and is one of the simplest and most important methods of clinical cytology screening for the prevention and early detection of cervical cancer [3].

Traditionally, cytology screening is performed manually by a cytologist or pathologist; hence, it is highly repetitive and time-consuming and may be subject to human mistakes even when the cytologist or pathologist is experienced. Therefore, the development of automatic computer-assisted cytology screening and diagnosis systems for cervical cancer is of great significance. The main objectives of such systems are to identify abnormal cells in samples containing thousands of cells that contain mostly normal cells and to show these abnormal cells to doctors. In reality, doctors determine anomalies mainly based on the 2001 Bethesda System [1],

and these anomalies are mostly found in the characteristics of the nuclei (e.g., in their shapes, colors, and sizes), which play a prominent role in the decision [4]. Therefore, such systems generally require segmentation, feature extraction, and classification of nuclei [5]. The accuracy and efficiency of the segmentation are important, since segmentation is the first step in this process, but it is quite challenging to achieve sufficient accuracy and efficiency due to complicated cases such as those involving irregular shapes, pool contrast, blurriness, and inhomogeneous chromatin distributions.

Many cervical nucleus segmentation methods have been proposed thus far, using various techniques such as Otsu-based thresholding [6], k-means[7], superpixel [8]–[11], level set [12]–[15], watershed [16]–[19], and active contour model (ACM) [20] methods. In some approaches, mainly spatial information is used to segment the nuclei [6], [21]. For instance, Tareef *et al*. [6] segmented nuclei by applying anisotropic diffusion filtering to reduce the noise and smooth the inconsistent regions, performing contrast-limited adaptive histogram equalization to improve the local image contrast, adopting H-maxima filtering to remove small and low-contrast regions, applying Otsu thresholding to the filtered images to identify the candidate nuclei, and finally using the regularized level set algorithm to refine the nuclear boundaries. Li *et al*. [21] converted color images into grayscale images; denoised the grayscale images by non-local means filtering; applied spatial k-means clustering to divide the image pixels into three classes, where the darkest class corresponded to nuclei; and extracted the nucleus regions by performing filtering using simple geometric information. In these methods, however, almost no prior information from the nucleus is used, which is significant for nucleus segmentation.

A number of methods have been proposed for nucleus segmentation from cervical smear images using shape prior information such as a regular shape [5], [22]–[24] or trained shape [25]. Specifically, Garcia-Gonzalez *et al*. [5] introduced a multiscale algorithm for nucleus extraction in which a multiscale edge detector and mean-shift filter were used to determine the spatial location of the edge and select the candidate nucleus by ellipse parameter calculation, a variety of techniques such as level set and watershed methods were employed to segment the nucleus region, and finally a support vector machine and an artificial neural network were utilized to extract the true nucleus. Plissiti *et al*. [25] obtained a compact representation of the shape of the nuclear boundary by performing training on a set of images containing single nuclei and then used it to guide the deformation of the initial nuclear boundaries to achieve the final segmentation. In recent years, it has become a trend to use trained shapes to perform segmentation of cytology images of cervical cells [8], [26] due to the ability to represent complex shapes, which are widespread in abnormal cervical cells. Recently, methods involving using pixel level priors to train nucleus segmentation models [27], [28], which contain not only shape features, but also more complex features, have been investigated. Song *et al*. [27] utilized a multiscale convolutional network (MSCN) to extract scale-invariant features, acquired a coarse segmentation to assign each pixel by maximizing the prediction of the MSCN at its location, and finally employed a graph partitioning method to refine the coarse segmentation. Zhang *et al*. [28] trained a fully convolutional network (FCN) [29] to learn a model that could perform pixel-to-pixel prediction for nuclei to achieve coarse segmentation, cropped the sub-regions of both the prediction and the nucleus image according to the coarse segmentation, and finally used the prediction and nucleus image along with a graph-based approach to achieve the refined segmentation. The experimental results demonstrated state-of-the-art performance. However, an FCN only utilizes single-scale feature maps, so the coarse segmentation may be inconsistent when the receptive field of the last pool layer of the backbone of the FCN is smaller than the nuclei, and the coarse segmentation may interpret the nucleus region as background when the receptive field is larger than the nuclei.

To overcome these issues, we combined a mask regional convolutional neural network (Mask-RCNN) [30] and a local fully connected conditional random field (LFCCRF) to segment cervical cell nuclei in this study. The Mask-RCNN uses multi-scale feature maps to acquire stronger semantic features to localize the cervical nuclear boundary generally, that is, to perform coarse segmentation of the cervical nuclei. Then, we incorporated the coarse segmentation and abundant spatial information into the LFCCRF to refine the nuclear boundary. We also conducted quantitative comparisons between our method and the current state-of-the-art methods based on the Herlev dataset [31]. The remainder of the paper is organized as follows. The proposed cervical nucleus segmentation method is introduced in Section 2. In Section 3, the description and preprocessing of the Herlev dataset as well as the evaluation metrics are presented, and the experimental results are shown. Discussions are given in Section 4. Finally, conclusions and topics requiring future work are presented in Section 5.

II. PROPOSED FRAMEWORK

To make full use of spatial information and prior knowledge, a Mask-RCNN and an LFCCRF are employed in the proposed method. A feature pyramid network (FPN) [32] based on ResNet [33], and modified according to the images of the cervical nucleus, is utilized as the backbone of the Mask-RCNN. Then, the Mask-RCNN can be employed to extract pyramid feature maps suitable for the images through the pixel-level prior information of the nucleus. Next, recognition and coarse segmentation of the nuclei are performed, the nuclear region of interest (RoI) is obtained by extending the bounding box of the nuclei provided by the recognition of the Mask-RCNN, and the coarse segmentation in the region is fed into the LFCCRF, which contains position and intensity information about all of the pixels in the region, to achieve refinement. The flowchart of the proposed method is shown in Fig. 1.

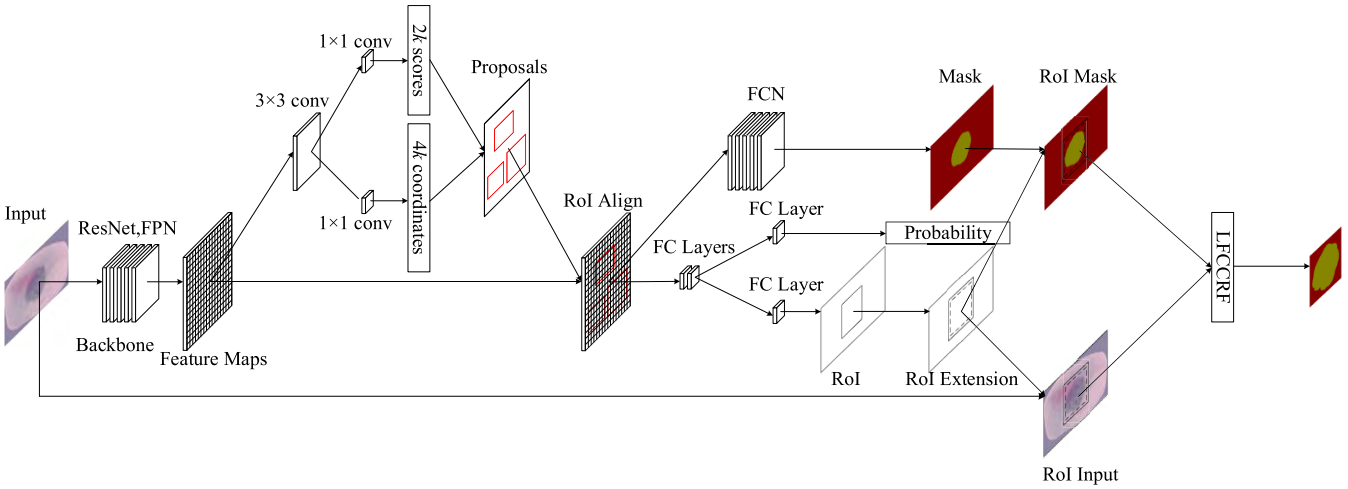


FIGURE 1. Flowchart of the proposed method. For simplicity, only feature maps after the backbone are shown for RPN's architecture, but all feature maps exist behind every layer.

A. MASK-RCNN-BASED COARSE SEGMENTATION

The architecture of a Mask-RCNN is based on that of a convolutional neural network (CNN), which is a deep neural network mainly consisting of various types of layers. The input of a CNN is usually an original image without complex preprocessing, and the outputs of the layers of a CNN are feature maps of various scales with various levels. Generally, the extracted features go from low level to high level as the receptive field expands owing to the deepening of the layers. A CNN uses a back-propagation algorithm [34] to optimize the weights of all of the weighted layers by minimizing the loss, which can be defined differently as different image tasks. Generally, a CNN has three main types of layers: convolutional, pooling, and fully connected layers.

A traditional neural network establishes a connection between the input and output by multiplying the input matrix by the weight matrix, which is a fully connected operation; consequently, it is time consuming. Thus, only a few features are allowed to be trained. To extract features more efficiently, the convolutional layer uses a local connection. It moves a small fixed-size convolutional filter horizontally and vertically along the input matrix to produce linear activation responses in the form of a feature map. Each neuron in the feature map is connected to the local receptive field of the input matrix and obtained by two steps. In the first step, it multiplies the values of the corresponding positions in the receptive field and the filter, and then adds the multiplied values to get the final value. The weights of the convolutional filter are shared by all neurons. To extract features in a more varied manner, multiple convolutional filters in the convolutional layer are operated in parallel to produce multiple feature maps, and various local features are then extracted. To prevent vanishing/exploding gradients and to increase the efficiency of the CNN back-propagation, the feature maps produced by the convolutional filter are fed into a rectified linear unit (ReLU) [35] in the proposed method. These feature maps of the ReLU are then input into the next type of layer, which is known as pooling, and are subsampled by selecting an overall statistical feature of a small neighborhood to replace the neighborhood. The overall statistical feature employed in this case uses the maximum value (called max-pooling). The pooling layer makes the CNN have translational, rotation, and scaling invariance, and the dimensional reduction of the feature vectors prevents network overfitting and further reduces the computational complexity, while also causing the details to be lost. Another important layer is the fully connected layer with a filter size of 1×1, which encodes the high-level features with strong semantic information suitable for different image tasks. The weights of the convolutional and fully connected layers are learnt and adjusted to represent the input data better during the training process.

A Mask-RCNN constructs three stages for coarse segmentation: feature extraction, region proposal, and prediction.

In the first stage, an FPN based on ResNet, with the refined basic feature extraction layers and the reduced subsequent feature extraction layers according to the images of the cervical nucleus, is applied as the Mask-RCNN backbone. Recent evidence [36]–[38] indicates that network depth is of crucial importance but has some side effects on network learning, such as vanishing/exploding gradients and training accuracy degradation. The former problem has been addressed by normalized initialization and intermediate normalization layers, and ResNet can solve the latter problem by applying a residual learning block to a CNN, as in Fig. 2.

The residual block is defined as

$$y = F(x) + x. \tag{1}$$

Here, the input feature maps are represented as x and the output of the merge layer is represented as y, which can be obtained by adding the weights of the corresponding positions of x and $F(x)$. Therefore, $F(x)$, the output of the three convolutional layers with the ReLU, becomes a residual

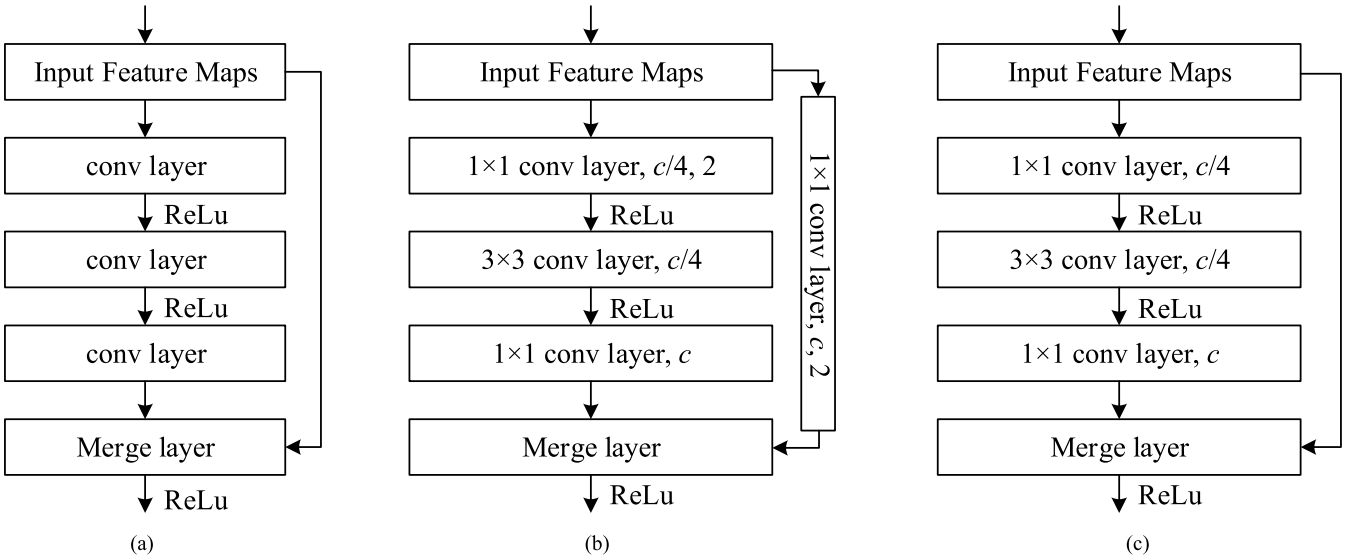


FIGURE 2. Residual learning block. (a) Flowchart of a residual learning block. (b) Convolutional block applied in our study. (c) Identity block applied in our study. In (b) and (c), the elements of the convolutional layer are arranged as "Convolutional layer with filter size, number of channels of filters, stride of filter (if not specified, it has a default value of 1)."

mapping from x to y. The identity mapping is a shortcut connection between the input feature maps and the merge layer (as the side line of Fig. 2(a)), simply performing identity mapping from x to y. Thus, because of the reference of the input, the weights of the convolutional layers are easier to learn than they would be if there were no shortcut connections. In the ResNet, almost all of the weighted layers consist of residual learning blocks, except for the first convolutional layer. Thus, because of the region field size requirement, the block must expand the region field. Hence, there are two types of blocks, i.e., convolutional and identity blocks, whose forms in this study are as shown in Figs. 2(b) and 2(c), respectively. The convolutional block can increase the region field size by increasing the stride of the first convolutional layer from one to two. To match the dimension, the shortcut connection should also increase the stride to two. The identity block extracts sufficient features of the scale determined by the convolutional block in front of it. The first, second, and third convolutional layers in Figs. 2(b) and 2(c) have filter sizes of 1×1, 3×3, and 1×1, respectively, and the first and third convolutional layers are responsible for reducing and restoring the number of filters (shown in Figs. 2(b) and 2(c) as c) to leave the second convolutional layer with fewer weights to learn. This design is reliable and has been demonstrated to provide high accuracy efficiently [30]. The details of the ResNet architecture used in this study are provided in Table 1.

A nucleus image is commonly simpler than an image of a scene in nature, so the number of blocks of conv4_x is reduced to prevent overfitting [33]. In addition, considering the low resolution that is widespread in images containing normal nuclei, the convolutional layer with a filter size of 7 in conv1_x is replaced with three convolutional layers with a filter size of 3 to extract more detailed basic features.

TABLE 1. Details of the ResNet architecture used in this study.

Stage	Component
conv1_x	$3\times3,\ 64,\ 2$ $3\times3,\ 64$ $3\times3,\ 64$ 3×3 max pool, stride 2
conv2_x	$\begin{bmatrix}1\times1,\ 64\\3\times3,\ 64\\1\times1,\ 256\end{bmatrix}\times3$
conv3_x	$\begin{bmatrix}1\times1,\ 128\\3\times3,\ 128\\1\times1,\ 512\end{bmatrix}\times4$
conv4_x	$\begin{bmatrix}1\times1,\ 256\\3\times3,\ 256\\1\times1,\ 1024\end{bmatrix}\times2$
conv5_x	$\begin{bmatrix}1\times1,\ 256\\3\times3,\ 256\\1\times1,\ 1024\end{bmatrix}\times3$

Although ResNet has a powerful representation, it has been proven that better performance can be obtained by using pyramid representations to address multi-scale image tasks [32]. Therefore, an FPN based on the ResNet is employed to extract multi-scale features in the proposed method.

An FPN contains three parts: a bottom-up pathway, a top-down pathway, and lateral connections, as illustrated in Fig. 3.

The bottom-up pathway is the feed-forward computation of the ResNet, because every stage has the same scale and the

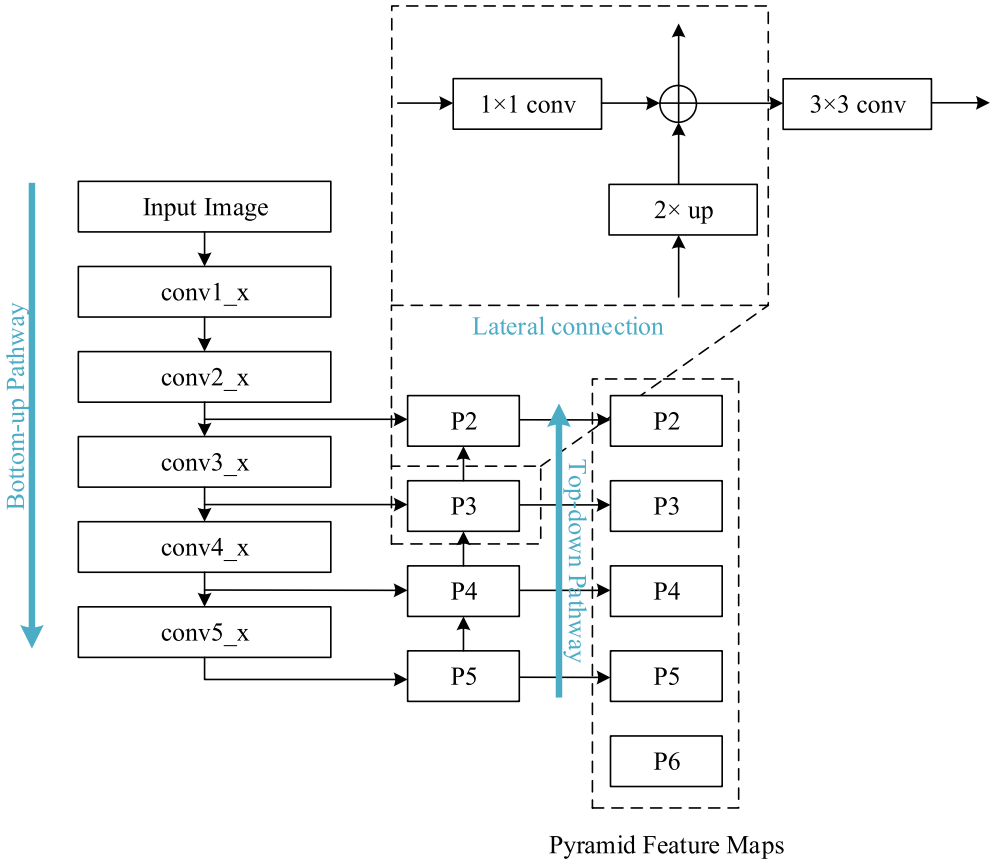


FIGURE 3. Flowchart of an FPN based on the ResNet. An FPN contains three parts: a bottom-up pathway, a top-down pathway, and lateral connections.

deeper the layer, the better the feature representation, so the activation of the output of the last residual block of each stage is employed to form the corresponding reference set of feature maps of the pyramid level.

The top-down pathway first upsamples the feature maps with a coarser spatial resolution but stronger semantic information from the higher pyramid levels than those of the lower pyramid levels by a factor of two. The upsampled features are then merged with the corresponding bottom-up features, which go through a 1×1 convolutional layer (to reduce the channel dimensions of all of the levels of the pyramid and to make them have a uniform value of 256) by element-wise addition. This process is called lateral connection and is performed by each stage in the top-down pathway, as depicted. Then, a 3×3 convolutional layer is appended to each merged feature map to reduce the aliasing effect due to upsampling. The outputs of all of the 3×3 convolutional layers form the final set of feature maps for our ResNet based FPN backbone; here, the set is denoted as $\{P_2, P_3, P_4, P_5\}$. Note that in Fig. 3, P_6 is simply a stride two subsampling of P_5 and only for covering a larger nuclear region.

In the second stage, a region proposal network is slid across the multi-scale feature maps provided by the FPN for region proposal.

An RPN [39] is a small network, as shown in Fig. 4, where the red box is a window that slides over the feature pyramid of the abovementioned backbone. Note that the level of the feature pyramid in which sliding is performed is determined by the size of the RoI [32]. Then, the sliding window is mapped to a 512-dimensional vector by a convolutional layer with a filter size of 3×3, and the vector is fed into two sibling 1×1 convolutional layers, one of which is used for box regression and the other for box classification. At each sliding window location, k region proposals are obtained simultaneously, so the layer for box regression outputs $4k$ coordinates for k boxes, and the layer for box classification outputs $2k$ scores for probability estimation of whether the box represents an object in each proposal. The k proposals are parameterized relative to k reference boxes, which are called anchors and are centered at the sliding window.

The anchors are selected based on the intersection-over-union ratio (IoU) of the anchor and ground-truth (GT) boxes. The anchor with the highest IoU overlap with a single GT box and anchors with IoU higher than 0.7 for all of the GT boxes are assigned positive labels, whereas anchors with IoU less than 0.3 for all of the GT boxes are assigned negative labels.

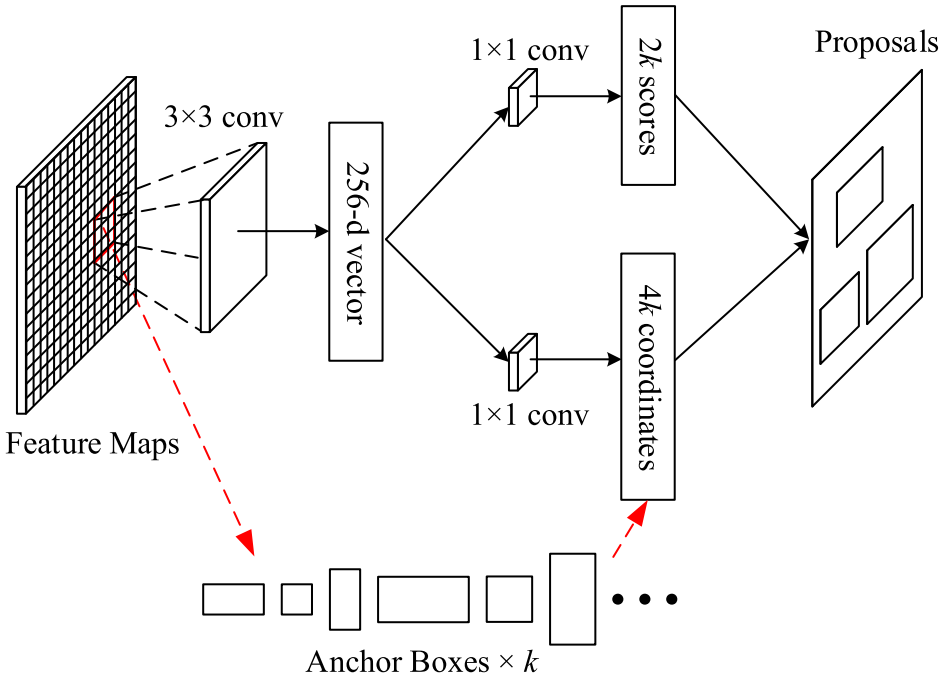


FIGURE 4. Architecture of the RPN. The red box is a window that slides over the feature pyramid of the abovementioned backbone.

The loss function of the RPN is defined as

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*). \tag{2}$$

Here, p_i is the probability that the predicted i-th anchor is an object. If the anchor is positive, the GT label p_i^* is one, and vice versa. t_i is the parameterized coordinate of the bounding box predicted by the RPN, and t_i^* is the coordinate of the GT corresponding to the positive anchor. L_{cls} is the log loss of the binary classification (object/non-object), and L_{reg} is the smooth L1 loss [40] of the predicted bounding box and GT box. The total loss of the RPN is normalized by N_{cls}, N_{reg}, and a balancing weight λ. Finally, the RPN acquires a series of region proposals by adjusting the network weights via training.

In the third stage, the features inside any region proposal are converted into a small feature map with fixed spatial dimensions of H×W (7×7 for nuclear recognition and 14×14 for nuclear segmentation because of the required segmentation precision) by a technique called RoIAlign, which can fix the misalignment of a traditional technique called RoIPool, caused by quantization when calculating the RoI boundaries and bins. Two branches follow to finish the bounding-box recognition and mask segmentation of the nuclei, as shown in Fig. 1.

In the first branch, each feature map is fed into two concatenated fully connected layers followed by two sibling fully connected layers. One of the sibling layers is used for box regression and outputs four values encoding refined bounding-box positions for each region proposal of N object classes, and the other is used for box classification and outputs the probability estimates over $N + 1$ classes (one is background). The classification loss L_{cls} is the log loss of the $N + 1$ classes, and the regression loss L_{reg} is the smooth L1 loss [40] of the bounding box generated by the RPN and the GT box. In our case, the object class is only nuclei, so N is one.

The second branch is parallel to the first branch and composed of RoIAlign and a tiny FCN, which consists of simply a few stacked convolutional layers followed by a transposed convolutional layer. The FCN enables the training for semantic segmentation due to the pixel-to-pixel correspondence. It should be noted that because the input of the tiny FCN consists of the feature maps of RoIAlign, which have low resolutions, and the stride of the transposed convolutional layer of the tiny FCN is set to two to control the number of weights, the pixel-to-pixel correspondence is actually one "pixel" of feature maps to four "pixels" of the fixed-size GT mask (resized to 28×28) of the RoI.

The mask branch predicts the nuclear binary segmentation mask for each RoI from the RPN. For each mask, a sigmoid is performed on each pixel of the last feature map of the tiny FCN:

$$S(\omega_{i,j}) = \frac{1}{1+e^{-\omega_{i,j}}}. \tag{3}$$

where $\omega_{i,j}$ is the value at position (i, j) of the feature map, and L_{mask} is the average binary log loss of all $S(\omega_{i,j})$.

Hence, the network is trained by using the definition of the multi-task loss for each sampled RoI:

$$L = L_{cls} + L_{box} + L_{mask}. \tag{4}$$

The Mask-RCNN method is a region-based technique for semantic segmentation and can be used to predict object bounding boxes and corresponding masks. The mask prediction is based on the region proposals provided by the RPN, and adding the mask loss to the multi-task loss promotes more accurate region proposal. The mutual promotion between the mask prediction and region proposal leads to precise localization of the nuclear boundary in general.

B. LFCCRF-BASED REFINEMENT

Some problems still exist in coarse nucleus segmentation produced by the Mask-RCNN. Because the second branch of the prediction stage is performed using feature maps with strong semantic information but low resolution, the details of the nuclear boundary are lost. Thus, if the RoI has a resolution similar to that of the fixed-size GT mask, the loss of segmentation details is mainly due to the low-resolution feature map, whereas more details are lost when the GT mask of the RoI is reduced to the fixed-size GT, which is coarse itself. To address this issue, we chose to apply an LFCCRF, which is performed on a local region including detected nuclei. The local region is obtained by extending the nucleus bounding box provided by the Mask-RCNN. For a bounding box with width W_b and height H_b, the local region can be obtained simply by extending W_b and H_b to $W_b + \Delta W_b$ and $H_b + \Delta H_b$, respectively. In this study, ΔW_b and ΔH_b were both set to 20 pixels to ensure the entire nucleus area was involved, and the center of the local region was the same as the bounding box.

The LFCCRF takes into account the relationships among all of the pixels in the nuclear RoI. To refine the coarse segmentation of the Mask-RCNN, the cost function of the LFCCRF, consisting of all of the pixel labels obtained from the coarse segmentation, as well as the position and intensity information of three channels in the local region, is minimized.

The pixel label information is derived from the segmentation of the Mask-RCNN, in which the value of each position corresponds to a label variable in the label set $L = \{l_1, \ldots, l_k\}$, where $k = 2$ represents the background and nuclei. All of the variables constitute a random field $Y = \{Y_1, \ldots, Y_{NP}\}$, where Y_j is the label corresponding to position j. Random field $X = \{X_1, \ldots, X_{NP}\}$ is defined on another set of variables, NP is the number of pixels in the nuclear RoI, and X_j represents the color vector of pixel j in the corresponding region from the cell image. Then, the conditional random field (X, Y) can be represented by a Gibbs distribution:

$$P(Y|X) = \frac{1}{Z(X)} \exp(-E(Y|X)), \tag{5}$$

where $Z(X)$ is the normalization term

$$Z(X) = \sum_{Y} \exp(-E(y|X)). \tag{6}$$

$E(Y|X)$ represents the energy when the pixel distribution of the input cell image is X and the label distribution is Y. The goal of the LFCCRF is to find the random field $y \in L^N$ corresponding to the maximum a posteriori (MAP). Because the pixel distribution of cell image X is used as the condition, the representation of condition X is omitted from the subsequent derivation for convenience. Then, for the distribution of $y \in L^N$, the corresponding Gibbs energy is

$$E(Y) = \sum_{i} \psi_u(y_i) + \sum_{i<j} \psi_p(y_i, y_j). \tag{7}$$

The first item is the unary energy term, which is only related to the labels from the coarse segmentation. In this study, the coarse segmentation of the Mask-RCNN was used directly, enabling the formula to be expressed as

$$\psi_u(y_i) = -\log(y_i)\, \psi_u(y_i) = -\log(y_i). \tag{8}$$

The second term is a pairwise energy term that is related to each pixel label from the coarse segmentation and the positions and intensity vectors of all of the pixels. The equation is

$$\psi_p(y_i, y_j) = \mu(y_i, y_j)\left\{\exp\left(-\frac{|pp_i - pp_j|^2}{2\theta_\alpha^2} - \frac{|pp_i - pp_j|^2}{2\theta_\beta^2}\right) + \exp\left(-\frac{|pp_i - pp_j|^2}{2\theta_\gamma^2}\right)\right\}, \tag{9}$$

where $\mu(y_i, y_j)$ is the label compatibility function, and the Potts model was used in this study:

$$\mu(y_i, y_j) = \begin{cases} 0, & y_i = y_j \\ 1, & y_i \neq y_j \end{cases} \tag{10}$$

The first exponential term in (9) is called the appearance kernel, the second exponential term is called the smooth kernel, and pp represents the pixel position. For RGB images, the appearance kernel measures the similarity of pixels in a five-dimensional space. That is, encouraging pixels with similar positions and similar colors have consistent classification. θ_α and θ_β control the sensitivity of the proximity and similarity between the pixels, the smooth kernel is used to remove isolated small areas, and θ_γ controls the sensitivity of the proximity between the pixels.

The quantity of pairwise energy terms for the LFCCRF is tremendous, which typically entails high computational complexity. Hence, we utilized an efficient inference method based on the mean field approximation proposed by Krähenbühl and Koltum in this study [41]. In this method, a simple distribution $Q(y)$ is computed that can replace the original distribution $P(y)$ of the LFCCRF, and each y is independent within this distribution:

$$Q(y) = \prod_{i} Q_i(y_i). \tag{11}$$

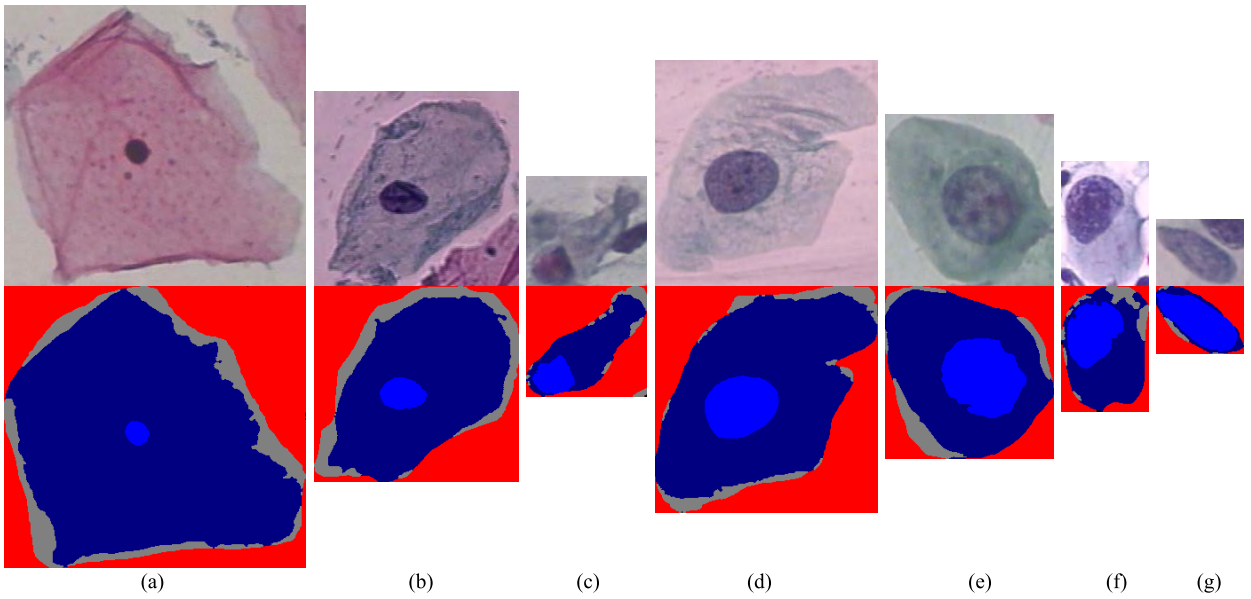


FIGURE 5. Samples of seven classes and the corresponding GT images in the Herlev dataset. First row: samples; second row: corresponding GT images.

Then, the Kullback–Leibler (KL) divergence of Q and P is minimized:

$$D(Q\|P) = \sum_{y} Q(y) \log\left(\frac{Q(y)}{P(y)}\right). \tag{12}$$

Finally, the equation of the iterative update process is

$$Q_i(y_i = l) = \frac{1}{Z_i} \exp\{-\psi_u(y_i) - \sum_{l' \in L} \sum_{i \neq j} \psi_p(y_i, y_j) Q_j(l')\}. \tag{13}$$

The nuclear RoI of the coarse segmentation of the Mask-RCNN and the corresponding nuclear RoI of the cell image is taken as the input of the inference, and the iterated y is the output of the LFCCRF, that is, the final segmentation of the nucleus.

III. EXPERIMENTAL RESULTS

A. DATASET

The method proposed in this paper was evaluated by using a common dataset, the new version of the Herlev Pap smear dataset. The dataset was collected by Herlev University Hospital and the Technical University of Denmark and contains 917 images of single Pap smear cells. The images in the Herlev dataset were obtained at a magnification of 0.201 μm/pixel with an average image size of 156×140. Among all of the images, the longest length of a side is 768 pixels, and the shortest is 32 pixels, so the range of side lengths in the dataset is wide. The cell images in the Herlev Pap smear dataset are divided into seven classes: (a) superficial squamous, (b) intermediate squamous, (c) columnar, (d) mild dysplasia, (e) moderate dysplasia, (f) severe dysplasia, and (g) carcinoma in situ, where the first three classes of cells correspond to normal cells, and the last four classes correspond to abnormal cells. Each cell image has an associated GT image depicting the cell nuclei, cytoplasm, and background area manually labeled by cytologists and physicians. Samples of the seven classes and the corresponding GT images are shown in Fig. 5.

TABLE 2. Details of the distribution of 917 cell images in the Herlev dataset.

Class	Category	Cell type	Cell count	Subtotal
1	Normal	Superficial squamous	74	242
2		Intermediate squamous	70	
3		Columnar	98	
4	Abnormal	Mild dysplasia	182	675
5		Moderate dysplasia	146	
6		Severe dysplasia	197	
7		Carcinoma in situ	150	

The details of the distribution of the 917 cell images in the Herlev dataset are summarized in Table 2.

In the GT, the nuclei and cytoplasm of the major cell are represented in light and dark blue, respectively; the gray area is the surrounding background of the major cell; and the red area is unrelated to the major cell and may contain the nuclei and cytoplasm of other cells, which are not labeled. Thus, if the GT is utilized for training directly, some interference will be introduced, and it is difficult to separate the region of the major cell completely by limiting the length and width. Considering the subtraction of the mean values of the

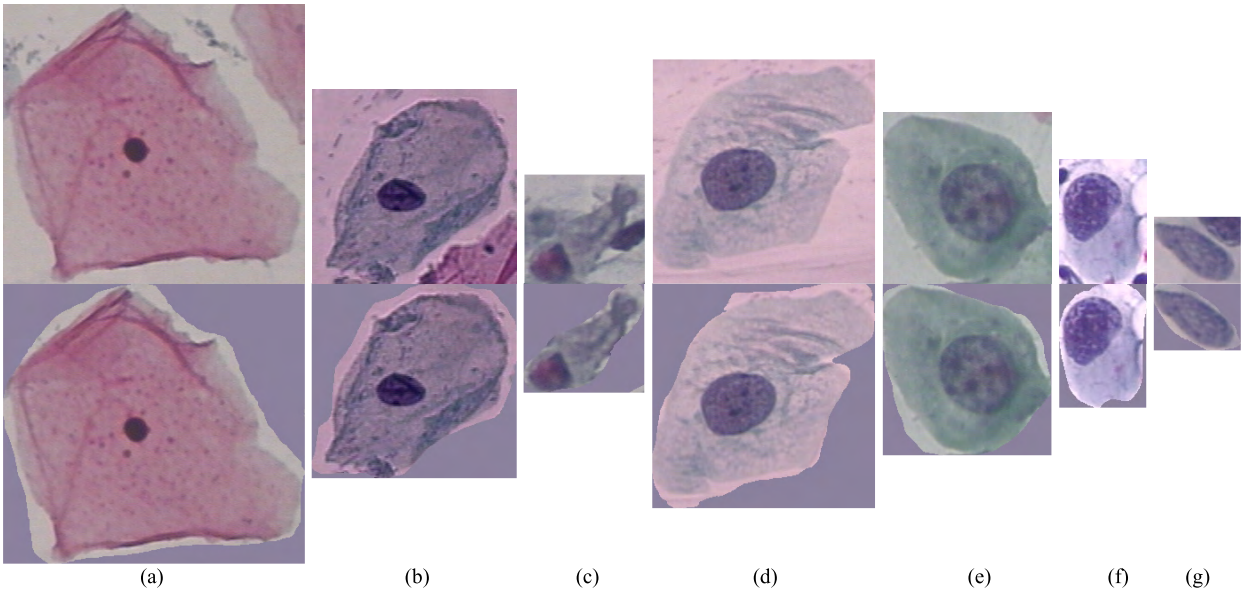


FIGURE 6. Extraction of major cell regions. First row: samples of seven classes; second row: corresponding extraction of major cell regions of seven classes.

three channels of the dataset during the Mask-RCNN training, the areas in the original images corresponding to the red areas in the GT images were replaced by the mean values of the three channels for the dataset in this study. Fig. 6 shows the results of region extraction from the cell images in Fig. 5.

For each pixel label in the cell images, if the pixel corresponds to the region of the cell nuclei in the GT, it is assigned a value of one, and the pixels in the remaining regions are assigned a value of zero.

B. EVALUATION MEASURES

We employed three pixel-level measurements to compare the segmentation quality of our proposed approach with those of other methods: the precision, recall, and Zijdenbos similarity index (*ZSI*) [42]:

$$Precision = \frac{TP}{TP + FP}, \tag{14}$$

$$Recall = \frac{TP}{TP + FN}, \tag{15}$$

and

$$ZSI = \frac{2TP}{2TP + FP + FN}, \tag{16}$$

where *TP* is the number of correctly detected pixels, *FP* is the number of detected pixels that are not in the GT, and *FN* is the number of pixels in the GT that are not detected. That is, the precision is the ratio between the number of correctly detected pixels and the total number of detected pixels, and the recall is the ratio between the number of correctly detected pixels and the total number of pixels in the GT. Thus, low precision indicates high false detection and low recall indicates high missed detection. Therefore, reliable segmentation requires both precision and recall to perform well. Another measurement, *ZSI* considers *TP*, *FP*, and *FN* comprehensively, and according to [42], the detected segmentation boundary and GT are well matched if *ZSI* is higher than 0.7.

C. PERFORMANCE COMPARISON

In this section, all of the experimental details and comparisons are provided. We performed all of the experiments using NVidia GeForce GTX 1080 Ti with 12 GB memory and deep learning architecture [43], [44] and Windows 10 installed on an Intel Core i7 CPU with 16 GB RAM. For objective and robust demonstration of our method, we conducted 10-fold cross validation, i.e., we trained 10 models for 10-fold validation sets.

We trained the Mask-RCNN for 400 epochs using stochastic gradient descent with a momentum of 0.9, a learning rate of 0.001, a batch size of four, weights decayed by 0.0001, and gradients clipped to 5.0 in each epoch. The weights were initialized by a model pre-trained on ImageNet [45], which is composed of natural images. We experimented with longer training times and other settings of the training hyper-parameters, but little improvement was achieved and the loss curve after 300 training epochs became smooth. Thus, we adopted models trained with 400 epochs for the coarse segmentation. To adapt the feature extraction network and minimize the image compression, the shorter sides of all of the cell images were rescaled to a length of 512.

In the refinement, θ_α, θ_β, θ_γ, and the number of iterations *ni* should be set. By combining many experimental results and the analysis of [41] [CRF], we set $\theta_\alpha = 1$, $\theta_\beta = 20$, and $\theta_\gamma = 10$. For the number of iterations, the KL divergence

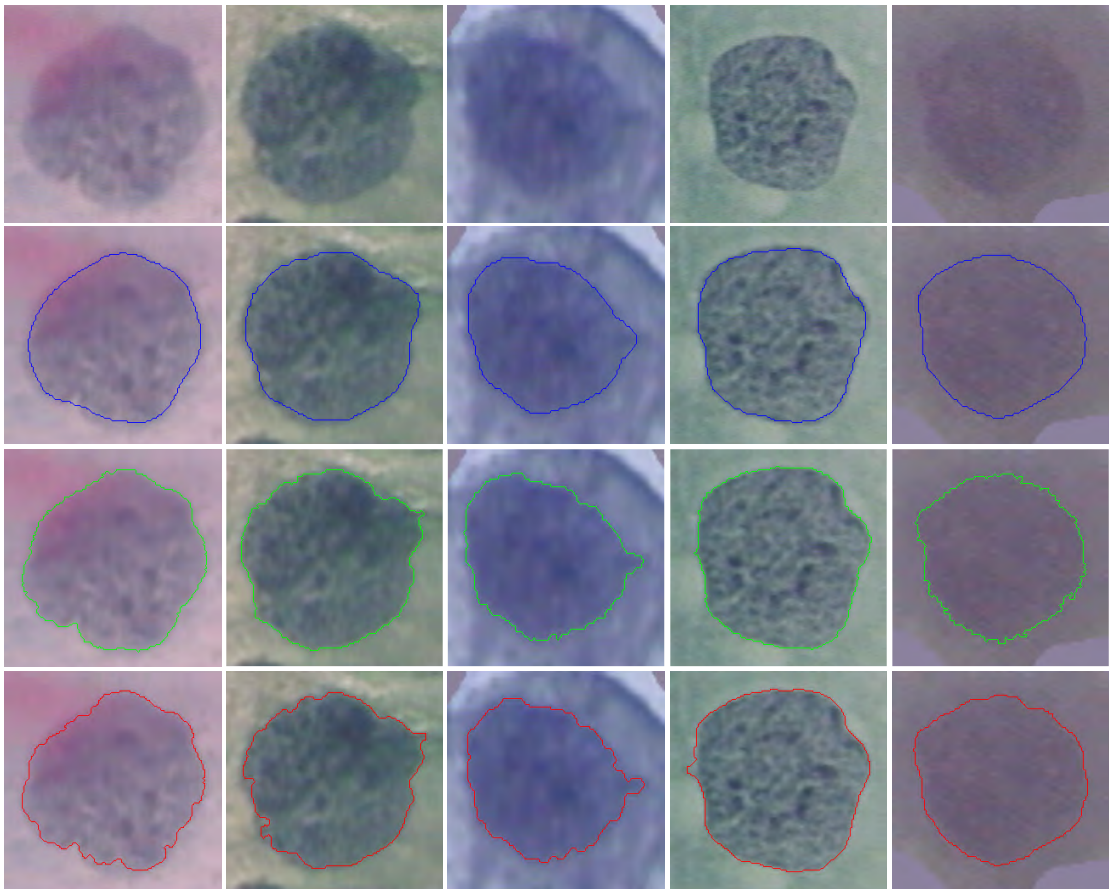


FIGURE 7. Results obtained by applying our segmentation method to the Herlev dataset. First row: original images; second row: results of coarse segmentation by the Mask-RCNN (blue boundaries); third row: results of refined segmentation (green boundaries); fourth row: GT results (red boundaries).

of Q and P almost stopped descending when $ni > 10$; hence, we set $ni = 10$ considering the computational complexity.

Several Herlev dataset samples segmented by using our method are shown in Fig. 7. Despite many sources of interference, such as irregular shapes, pool contrast, blurriness, and inhomogeneous chromatin distributions, the boundaries obtained by using our method are highly matched with the GT. In addition, the details that were not obvious in the coarse segmentation were refined well by the LFCCRF, indicating that our method overcomes these types of inference effectively.

The results were compared with those obtained using other methods [17], [21], [28], [46] that utilize popular techniques such as ACM, multi-scale watershed, and FCN. It should be noted that the results of [17] and [21] were obtained from [17]. The quantities used for comparison are shown in Table 3. For the average of three measurements, our method achieved 0.96, 0.96, and 0.95 for the precision, recall, and *ZSI* respectively, representing significant improvement over the other methods. For abnormal types of nuclei, it becomes more difficult to overcome the inhomogeneous chromatin distribution and the highly irregular shape of the nucleus with increasing abnormality. Thus, it is difficult to perform all three measurements well. Nonetheless, our method yielded parameter measurements all greater than 0.95 with low standard deviations for abnormal types of nuclei.

IV. DISCUSSION

In this study, we utilized both pixel-level prior information and information about the spatial domain of the nuclear RoI to overcome the difficulties that are widespread in cervical nucleus images, such as irregular shapes, pool contrast, blurriness, and inhomogeneous chromatin distributions. The Mask-RCNN method is a powerful technique that uses pixel-level prior information for coarse segmentation; the second row of Fig. 7 reveals accurate general localization. However, the segmentation is based on an FCN, which inevitably causes loss of spatial details. Hence, we chose to employ an LFCCRF to refine the coarse segmentation by minimizing the cost function, which consists of all of the pixel labels from the coarse segmentation, as well as the position and intensity information of three channels of the nuclear RoI. As a result of the full use of pixel-level prior information and the spatial

TABLE 3. Comparison of the performance of our method with those of different nucleus segmentation methods using the Herlev dataset.

Method	Cell type	Precision	Recall	ZSI
Mask-RCNN + LFC-CRF	Superficial squamous	0.96 ± 0.06	0.97 ± 0.12	0.95 ± 0.16
	Intermediate squamous	0.95 ± 0.05	0.99 ± 0.02	0.97 ± 0.02
	Columnar	0.93 ± 0.10	0.94 ± 0.17	0.92 ± 0.16
	Mild dysplasia	0.96 ± 0.04	0.98 ± 0.07	0.97 ± 0.07
	Moderate dysplasia	0.96 ± 0.04	0.97 ± 0.08	0.96 ± 0.08
	Severe dysplasia	0.97 ± 0.04	0.95 ± 0.12	0.95 ± 0.12
	Carcinoma in situ	0.96 ± 0.05	0.96 ± 0.10	0.95 ± 0.09
	Average	**0.96 ± 0.05**	**0.96 ± 0.11**	**0.95 ± 0.10**
Multi-scale Watershed + Binary classifier [17]	Superficial squamous	0.69 ± 0.37	0.63 ± 0.37	0.98 ± 0.12
	Intermediate squamous	0.79 ± 0.29	0.73 ± 0.31	0.98 ± 0.12
	Columnar	0.85 ± 0.15	0.77 ± 0.18	0.98 ± 0.05
	Mild dysplasia	0.88 ± 0.17	0.86 ± 0.16	0.96 ± 0.16
	Moderate dysplasia	0.91 ± 0.10	0.86 ± 0.14	0.97 ± 0.07
	Severe dysplasia	0.90 ± 0.12	0.89 ± 0.11	0.95 ± 0.13
	Carcinoma in situ	0.89 ± 0.15	0.90 ± 0.08	0.92 ± 0.17
	Average	**0.88 ± 0.15**	**0.93 ± 0.15**	**0.89 ± 0.15**
RGVF [21]	Superficial squamous	0.92 ± 0.12	0.88 ± 0.14	0.98 ± 0.02
	Intermediate squamous	0.95 ± 0.03	0.92 ± 0.06	0.98 ± 0.02
	Columnar	0.83 ± 0.16	0.76 ± 0.20	0.97 ± 0.08
	Mild dysplasia	0.92 ± 0.13	0.90 ± 0.16	0.96 ± 0.08
	Moderate dysplasia	0.89 ± 0.15	0.87 ± 0.17	0.94 ± 0.13
	Severe dysplasia	0.88 ± 0.15	0.90 ± 0.13	0.90 ± 0.19
	Carcinoma in situ	0.84 ± 0.18	0.88 ± 0.11	0.86 ± 0.24
	Average	**0.83 ± 0.20**	**0.96 ± 0.13**	**0.87 ± 0.19**
Patch-based FCM [46]	Superficial squamous	0.95 ± 0.12	0.75 ± 0.33	0.78 ± 0.29
	Intermediate squamous	0.98 ± 0.03	0.82 ± 0.25	0.86 ± 0.21
	Columnar	0.88 ± 0.20	0.78 ± 0.25	0.79 ± 0.19
	Mild dysplasia	0.80 ± 0.31	0.86 ± 0.26	0.79 ± 0.28
	Moderate dysplasia	0.81 ± 0.25	0.88 ± 0.19	0.81 ± 0.21
	Severe dysplasia	0.79 ± 0.28	0.88 ± 0.21	0.79 ± 0.25
	Carcinoma in situ	0.70 ± 0.29	0.88 ± 0.23	0.75 ± 0.25
	Average	**0.85 ± 0.21**	**0.83 ± 0.25**	**0.80 ± 0.24**
FCN-G [28]	**Average**	**N/A**	**N/A**	**0.92 ± 0.09**

information of the nuclear RoI, the refined segmentation results in the third row of Fig. 7 are well matched with the GTs.

Both accurate and stable abnormal nucleus segmentation are crucial for clinical diagnosis. Many previous methods involved limited use of prior knowledge, and although the results of some of them are highly accurate, their standard deviations are low. As shown in Table 3, our segmentation results for abnormal types of nucleus are both more accurate and more stable than those of the other methods.

V. CONCLUSION

Cervical nuclear segmentation plays an important role in automatic cytology screening and diagnosis for cervical cancer. In this paper, we presented a novel nuclear segmentation method in which a Mask-RCNN is combined with an LFCCRF. This technique involves adequate integration and utilization of prior and spatial information. The Mask-RCNN generally provides accurate boundary localization, mainly because of the robust semantic information obtained under the supervision of the pixel level prior information, and the LFCCRF explores the spatial information, including the position, intensity, and coarse segmentation results obtained by using the Mask-RCNN in the nuclear RoI, to refine the nuclear boundary. The experimental results on the Herlev dataset demonstrate that the proposed method clearly provides nuclear segmentation superior to that achievable by using the other methods. We believe that the results are reliable for subsequent analyses of automatic cytology screening and diagnosis for cervical cancer.

Although this work noticeably improved the cervical nucleus segmentation performance, the accuracy of the abnormal nuclear segmentation still needs to be enhanced owing to its clinical significance. In addition, the current work can be easily extended to cytological screening of other cancers.

ACKNOWLEDGMENT

The authors would like to thank for the Herlev dataset provided at http://mde-lab.aegean.gr/downloads.

REFERENCES

[1] *Comprehensive Cervical Cancer Control: A Guide to Essential Practice*. Geneva, Switzerland: World Health Organization, 2014.

[2] G. N. Papanicolaou, "A new procedure for staining vaginal smears," *Science*, vol. 95, pp. 438–439, Apr. 1942.

[3] D. Saslow *et al.*, "American cancer society, American society for colposcopy and cervical pathology, and american society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer," *CA, Cancer J. Clinicians*, vol. 62, no. 3, pp. 147–172, 2012.

[4] M. E. Plissiti and C. Nikou, "Cervical cell classification based exclusively on nucleus features," in *Proc. Int. Conf. Image Anal. Recognit.*, vol. 7325, 2012, pp. 483–490.

[5] D. Garcia-Gonzalez, M. Garcia-Silvente, and E. Aguirre, "A multiscale algorithm for nuclei extraction in pap smear images," *Expert Syst. Appl.*, vol. 64, pp. 512–522, Dec. 2016.

[6] A. Tareef, Y. Song, M.-Z. Lee, D. D. Feng, M. Chen, and W. Cai, "Morphological filtering and hierarchical deformation for partially overlapping cell segmentation," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Nov. 2015, pp. 1–7.

[7] T. Guan, D. Zhou, and Y. Liu, "Accurate segmentation of partially overlapping cervical cells based on dynamic sparse contour searching and GVF snake model," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1494–1504, Jul. 2015.

[8] A. Tareef *et al.*, "Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling," *Neurocomputing*, vol. 248, pp. 28–40, Jul. 2017.

[9] L. Zhao *et al.*, "Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF," *Comput. Biol. Med.*, vol. 71, pp. 46–56, Apr. 2016.

[10] D. M. Ushizima, A. G. C. Bianchi, and C. M. Carneiro, "Segmentation of subcellular compartments combining superpixel representation with Voronoi diagrams," in *Proc. Int. Symp. Biomed. Imag.*, 2015, pp. 1–2.

[11] A. Tareef, Y. Song, W. Cai, D. D. Feng, and M. Chen, "Automated three-stage nucleus and cytoplasm segmentation of overlapping cells," in *Proc. 13th Int. Conf. Control Automat. Robot. Vis.*, Dec. 2015, pp. 865–870.

[12] Z. Lu, G. Carneiro, and A. Bradley, "An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1261–1272, Apr. 2015.

[13] Z. Lu, G. Carneiro, and A. P. Bradley, "Automated nucleus and cytoplasm segmentation of overlapping cervical cells," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, vol. 8149, 2013, pp. 452–460.

[14] N. M. Harandi, S. Sadri, N. A. Moghaddam, and R. Amirfattahi, "An automated method for segmentation of epithelial cervical cells in images of thinprep," *J. Med. Syst.*, vol. 34, no. 6, pp. 1043–1058, 2010.

[15] M. S. Nosrati and G. Hamarneh, "A variational approach for overlapping cell segmentation," in *Proc. Overlapping Cervical Cytology Image Segmentation Challenge Conjunct IEEE Int. Symp. Biomed. Imag.*, 2014, pp. 1–2.

[16] J. Zhang, Z. Hu, G. Han, and X. He, "Segmentation of overlapping cells in cervical smears based on spatial relationship and overlapping translucency light transmission model," *Pattern Recognit.*, vol. 60, pp. 286–295, Dec. 2016.

[17] A. Gençtav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images," *Pattern Recognit.*, vol. 45, no. 12, pp. 4151–4168, 2012.

[18] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in pap smear images," *Pattern Recognit. Lett.*, vol. 32, no. 6, pp. 838–853, 2011.

[19] A. Kale and S. Aksoy, "Segmentation of cervical cell images," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2399–2402.

[20] M. Hu, X. Ping, and Y. Ding, "Automated cell nucleus segmentation using improved snake," in *Proc. Int. Conf. Image Process.*, vol. 4, Oct. 2004, pp. 2737–2740.

[21] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake," *Pattern Recognit.*, vol. 45, no. 4, pp. 1255–1264, 2012.

[22] C. Bergmeir, M. G. Silvente, and J. M. Benítez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a Web-based software framework," *Comput. Methods Programs Biomed.*, vol. 107, pp. 497–512, Sep. 2012.

[23] C. Jung, C. Kim, S. W. Chae, and S. Oh, "Unsupervised segmentation of overlapped nuclei using Bayesian classification," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2825–2832, Dec. 2010.

[24] C. Jung and C. Kim, "Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2600–2604, Oct. 2010.

[25] M. E. Plissiti and C. Nikou, "Overlapping cell nuclei segmentation using a spatially adaptive active physical model," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4568–4580, Nov. 2012.

[26] A. Tareef *et al.*, "Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation," *Neurocomputing*, vol. 221, pp. 94–107, Jan. 2017.

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

[28] L. Zhang, M. Sonka, L. Lu, R. M. Summers, and J. Yao, "Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei," in *Proc. 14th IEEE Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 406–409.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

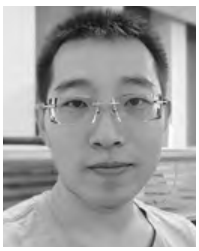
[30] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[31] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Nature Inspired Smart Inf. Syst.*, vol. 3, pp. 1–9, Oct. 2005.

[32] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, 2017, p. 3.

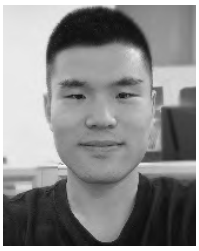
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[34] Y. A. Lecun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient BackProp," in *Neural Networks: Tricks of the Trade*. London, U.K.: Springer-Verlag, 1998.
[35] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[37] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[38] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[40] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.
[41] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
[42] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 716–724, Dec. 1994.
[43] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. OSDI*, 2016, pp. 265–283.
[44] F. Chollet. (2015). *Keras: Deep Learning Library for Theano and Tensorflow*. [Online]. Available: https://keras.io/k
[45] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[46] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in pap smears," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 539–556, 2014.

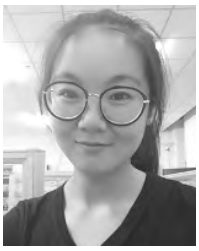


YIMING LIU is currently pursuing the master's degree with the School of Information and Communication Engineering, North University of China, Taiyuan, China. His research interests include image processing, deep learning, and computer vision.

PENGCHENG ZHANG received the Ph.D. degree in computer science and technology from Southeast University, Nanjing, China, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 2014. He is currently engaged in teaching and research with the North University of China. His research interests include medical image reconstruction, medical image analysis, dose calculation, and planning optimization.



QINGCHE SONG is currently pursuing the master's degree with the School of Information and Communication Engineering, North University of China, Taiyuan, China. His research interests include deep learning and image processing.



ANDI LI is currently pursuing the master's degree with the School of Information and Communication Engineering, North University of China, Taiyuan, China. Her research interests include CT imaging reconstruction and medical image processing.

PENG ZHANG is currently pursuing the master's degree in biomedical engineering with the North University of China, Taiyuan, China. His research interests include image reconstruction, biomedical optical spectroscopy, and deep learning.

ZHIGUO GUI received the Ph.D. degree in signal and information processing from the North University of China, Taiyuan, China, in 2004. He is currently a Professor with the North University of China. His research interests include signal and information processing, image processing and recognition, and image reconstruction.