

# Accurate Cervical Cell Segmentation from Overlapping Clumps in Pap Smear Images

Youyi Song, Ee-Leng Tan, Xudong Jiang, *Senior Member, IEEE*, Jie-Zhi Cheng, *Member, IEEE*, Dong Ni, Siping Chen, Baiying Lei\*, *Member, IEEE*, and Tianfu Wang

**Abstract**—Accurate segmentation of cervical cells in Pap smear images is an important step in automatic pre-cancer identification in the uterine cervix. One of the major segmentation challenges is overlapping of cytoplasm, which has not been well-addressed in previous studies. To tackle the overlapping issue, this paper proposes a learning-based method with robust shape priors to segment individual cell in Pap smear images to support automatic monitoring of changes in cells, which is a vital prerequisite of early detection of cervical cancer. We define this splitting problem as a discrete labeling task for multiple cells with a suitable cost function. The labeling results are then fed into our dynamic multi-template deformation model for further boundary refinement. Multi-scale deep convolutional networks are adopted to learn the diverse cell appearance features. We also incorporated high-level shape information to guide segmentation where cell boundary might be weak or lost due to cell overlapping. An evaluation carried out using two different datasets demonstrates the superiority of our proposed method over the state-of-the-art methods in terms of segmentation accuracy.

**Index Terms**—Cervical cancer, dynamic multi-template deformation model, multi-scale convolutional networks, overlapping cells splitting, Pap smear screening.

## I. INTRODUCTION

CERVICAL cancer developed in the cervical transformation zone is mainly caused by infection of several

types of human papillomavirus [1]. This disease has a global health impact second to breast cancer in women. More than 0.53 million women are diagnosed with this type of cancer worldwide, and more than 0.28 million women's lives are taken by cervical cancer every year [2]. Cervical cancer is curable if it is detected timely and treated appropriately [3], but detecting this disease at pre-cancerous stage remains a challenging task [4]. Cytology-based screening methods, e.g., the Pap tests, have been the major approaches in clinical test, but these screening methods must be performed regularly [5]. In addition, examination of cervical cytological images is manually performed in current clinical practices. Such screening process is usually labor-intensive, time-consuming, and subjective [6] due to the high number of cells. Furthermore, cancerization of cells is a continuous process, and many abnormal cells are visually similar to normal cells. Therefore, locating these abnormal cells from a myriad of normal cells is extremely tedious and highly subjective to human errors [7], [8].

These issues can be addressed by automating the screening methods [9], which can offer efficient and accurate approaches to analyze vast amount of data and relieve the laborious tasks required by pathologists [10]–[15]. To automate the screening methods, accurately segmenting cervical cells is an essential task, which makes extracting quantitative features of each cell accessible. The structure information of cells (e.g., shape, staining intensity, texture, and number), which is extracted from each individual cell, is vital to the early detection of cervical cancer [16], [19].

Many approaches for cervical cells segmentation have been proposed in the literature, yet early works [21], [22], [24]–[26] perform segmentation only in local image patches. Although these works are extended to delineate cells in the fields of view (FOV) images [20], [23], [27]–[29], they are unable to segment cells with overlapped cytoplasm, which can be commonly observed in the images (see Fig. 1). Splitting of overlapped cytoplasm is the most difficult task for computerized segmentation of cervical cells. Some previous works [32], [34], [35]–[37], [41], [42] focus on splitting overlapped nuclei, but they are indirect solutions. In addition, splitting of overlapped cytoplasm is more complicated than splitting of overlapped nuclei since the boundary of cytoplasm is usually unsmooth. The shape of cytoplasm is non-convex, and the extent of overlapping of the cytoplasm is quite high. These previous works have limited performance in splitting of overlapped cytoplasm. In addition, some approaches [18], [39] employ shape priors to segment cell from overlapping clumps,

Manuscript received July 6, 2016; accepted September 1, 2016. Date of publication September 7, 2016; date of current version December 29, 2016. This work was supported partly by National Natural Science Foundation of China (Nos. 61402296, 61571304, 81571758, 61501305 and 61427806), National Key Research and Development Project (No. 2016YFC0104703), the (Key) Project of Department of Education of Guangdong Province (No. 2014GXM052) and Shenzhen Key Basic Research Project (Nos. JCYJ20150525092940986, JCYJ20150525092940982 and JCYJ20140509172 609164). *Asterisk indicates corresponding author*

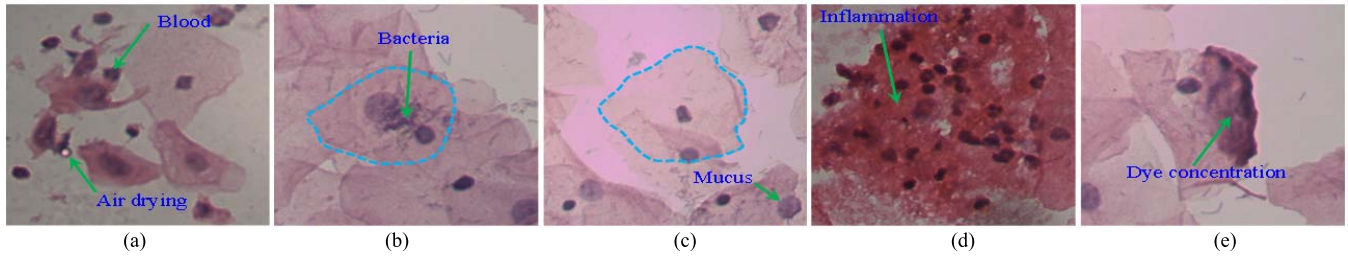
Y. Song, J.-Z. Cheng, D. Ni, and S. Chen are with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China.

E.-L. Tan and X. Jiang are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

\*B. Lei is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: leiby@szu.edu.cn).

T. Wang is with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: tfwang@szu.edu.cn).

Digital Object Identifier 10.1109/TMI.2016.2606380



**Fig. 1.** Illustration of the challenges of accurate cervical cell segmentation from overlapping clumps. (a) The complexity of the cell structure makes the segmenting real boundary of overlapping cytoplasm difficult as shown in the violet rectangles. (b-c) Highly overlapping of cells and illumination variances cause the weak boundary of cytoplasm. The light dotted lines show related cells' ground truth. (d) Inflammation regions are easily confused cytoplasm. (e) Dye concentration misleads the desired segmentation. In addition, as indicated by a green arrow, blood, air drying, bacteria and mucus also affect the segmentation results.

but the imposed shape prior and regular boundary patterns cannot provide sufficient shape details for the overlapping parts. Therefore, these conventional methods are unable to address the challenging issues (see Fig. 1).

By and large, segmenting cervical cell from overlapping clumps in the Pap smear images is a highly challenging task. Illumination variances and dye concentration caused by the staining procedure led to low color contrast among cell components. Other issues such as air drying, excessive blood, mucus, bacteria, and inflammation also render recognition of cytoplasm and nuclei regions even more challenging [17] (see Fig. 1 c-e). In addition, complex cells' shape and the high degrees of overlapping of cells led to a weak boundary of cytoplasm (see Fig. 1 a-b). Therefore, accurate segmentation of cervical cell from overlapping clumps becomes impossible [18].

In this paper, we propose a framework based on deep learning technique and a deformation model to accurately segment cervical cells from overlapping clumps. This framework not only explores the shape correlation of the overlapping cells, but also utilizes the cell's structural and contextual information. We employ a multi-scale convolutional neural network (CNN) to learn cervical image pixels so as to increase the color contrast among the cell components and remove image noise. Differing from the conventional CNN based method, the proposed CNN learns features at different scales. Therefore, contexts of multiple scales are integrated into local decisions to correctly classify each pixel into background, cytoplasm, or nucleus pixel.

For accurate splitting of the detected overlapping cytoplasm, a multiple cells labeling model is proposed in this paper. The labeling model casts the splitting problem into a discrete labeling task. The classified cytoplasm pixels are labeled with the corresponding nucleus' labels generated by CNN. The shape of cells is estimated using Gaussian kernels fitting, so the dependence of color information is relieved by inferring the shape cues. As color observations are usually deficient in cell extracted from overlapping clumps, these shape cues can assist in ensuring the correct splitting path is searched. The labeling results are then fed into the dynamic multi-template deformation model for further boundary refinement and inference. Case-specific shape constraint is incorporated in this deformation model to guide inference where the cell boundary is weak or even lost caused by the complexity of

cells shape and cells overlapping. Our framework has been evaluated using two different datasets, and the segmentation results are compared with the state-of-the-art approaches. Our comparison reveals that the proposed approach outperforms these approaches in terms of segmentation accuracy.

## II. RELATED WORK

### A. Cervical Cells Region Segmentation

The methods of segmenting cervical cells region can be divided into two categories based on the number of cells in the image: single cell view and FOV image segmentation. Single cell view image contains only one cell, and samples from FOV image at a 40 magnification and has a size around 10001000 pixels.

**Single Cell View Image Segmentation:** The widely used methods for single cell view image segmentation include thresholding [20], morphology operation [21],  $k$ -means [22], Hough transform [23], and watershed [17]. While these methods are efficient in terms of computation complexity, they only address simple issues. To enhance performance, Shys-Fan *et al.* [24] proposed a mean vector difference to differentiate edges between nuclei and cytoplasm by computing the probability of each pixel based on the gradient vector flow. However, this approach only utilizes intensity information and cannot guarantee the detected boundary is closed. Kuan *et al.* [25] used a snake model to obtain a continuous boundary. They also extended the gradient vector flow to multiple directions, which computes intensity gradient along radiating lines from nucleus centroid. This work was further enhanced by Tao *et al.* [26]. In this work, a dynamic sparse contour searching algorithm is designed to locate the weak contour points of cytoplasm in the overlapped regions.

**FOV Image Segmentation:** For segmenting cytoplasm and nuclei of multiple cells in one image, Genctav *et al.* [20] used a minimum error thresholding method to eliminate background. The nuclei candidates are then searched using multi-scale watershed. Bergmeir *et al.* [23] segmented nuclei regions under the assumption that nucleus has an elliptical shape. This method utilized Canny edge operator to detect edges of nucleus and cytoplasm. Then Hough transform is employed to localize nuclei based on the detected edges. However, these methods are only based on healthy cells rather than a mixture of healthy and pathological cells. As the shape and

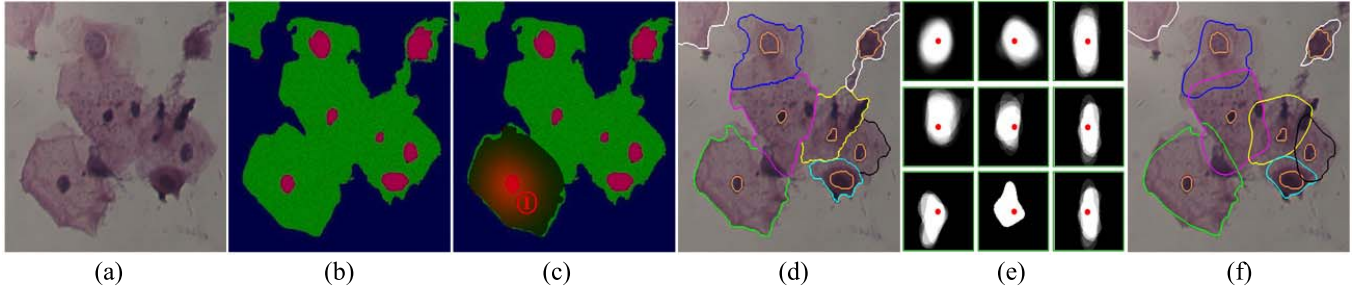


Fig. 2. Intermediate results of a sample image generated by the proposed approach for segmenting cervical cells from clump. (a) Input image. (b) Result of cell component segmentation by the multi-scale CNN. (c) Visualizing result of the overlapping-based unary cost of the cell indicated by the circled number. (d) Result of multiple cells labeling. (e) Generation of case-specific templates used as shape priors where the red dot denotes the centroid of the nucleus. (f) Result of cell boundary refinement and inference by the dynamic multi-template deformation model.

chromatin pattern of the abnormal cells are highly important for detecting cervical cancer, Zhang *et al.* [27], [28] proposed an approach using graph-cut to recognize abnormal cervical cell regions. However, they have only explored the statistic intensity information to construct the graph, which produces limited robustness. Also, their method often recognizes some variables (e.g., blood, bacteria, and mucus, see Fig. 1 a-c) as nuclei. To address the problems introduced by these types of noise, Song *et al.* [29] used a multi-scale CNN to learn image features and a graph model was constructed based on the output of the CNN. The superpixels rather than the pixels are used as the nodes of their graph to reduce the computational cost. The graph's energy is then optimized via graph cut techniques to produce the segmentation results.

### B. Touching Objects Splitting

To split overlapped objects, methods include watershed transform [30], [31], concave points searching [32], [33], graph optimization [34]–[36], level set [37]–[39], sliding band filter [40], and associating contour evidence [41]–[43] are widely used. The watershed method [30], [31] used markers to represent each individual object, and initiated the flooding process from these markers to localize the overlapping boundaries. In general, this method can produce good results if there are strong gradients at the boundary between the overlapped objects, but it cannot provide any inferences of the overlapped parts. Concave points searching methods [32], [33] extracted concave points according to the convexity property and then computed the splitting path between a pair of corresponding concave points. This method requires a convex shape of overlapped objects, which restricts its applications. Graph optimization based methods [34]–[36] constructed a graph on image pixels or candidates' edges and then minimized the graph energy to segment overlapped objects. This type of methods requires a suitable cost function to measure the similarity in different nodes. Also, the graph energy is easily stuck into local minima during the optimization process. Active contour model based methods [37]–[39] split the overlapped objects by constructing the shape prior constraint. This method is too computationally intensive when handling a large number of objects. In addition, the shape prior constraints are difficult to model because the shape of objects is heterogeneous.

Recently, a general framework [41], [42] was proposed to segment overlapped objects with an elliptical shape. The basic idea of this framework is to locate markers corresponding to the target objects. Each edge pixel's evidence is associated with these markers to locate the target's contours. Finally, the ellipse estimated method is utilized to infer the shape of each target based on located contours. This method can achieve promising results if the objects in the same overlapping clump have similar sizes.

## III. PROPOSED SEGMENTATION MODEL

The proposed approach for segmenting individual cervical cells has three main processes: cell component segmentation, multiple cells labeling, and cell boundary refinement and inference. Fig. 2 shows some intermediate results of a sample image generated by these processes. The detailed procedures are presented in the following sections.

### A. Cell Component Segmentation

Cell component segmentation aims to classify every pixel into the region of nuclei, cytoplasm or background. In our approach, each cell is represented by its nucleus. Thus, the accurate detection of nuclei is critical. In addition, the segmentation needs to be robust against image noise. For this challenging task, it is very difficult to figure out the effective features based on our human knowledge. Therefore, we utilize the powerful multi-scale CNN [44] to learn features from image samples. The network contains multi-scale CNNs, and each scale CNN has windows of multiple sizes centered at every pixel in the image. After it is properly trained, the CNN network encodes multiple types of information in multiple sizes of contextual windows such as appearance, texture, contextual, and shape information. The structure of our network is illustrated in Fig. 3.

Each input image  $I$ , multi-scale pyramid images  $I_s, \forall s \in \{1, \dots, S\}$ , are constructed using the dyadic Gaussian pyramids, where  $I_1$  and  $I$  are of the same size. For each scale  $s$ , the image's raw pixels are fed to the plane CNN ( $N_s$ ) with parameter  $\omega_s$  (The set of parameters  $\omega_s$  configures the plain CNN  $N_s$  to specify the architecture of the CNN for specific application), which can be described as:

$$N_s(I_s; \omega_s) = W_L H_{L-1}, \quad (1)$$



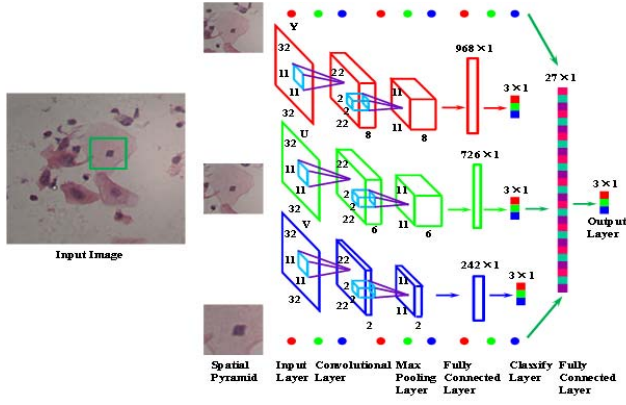


Fig. 3. CNN network structure for cell component segmentation. RGB color space is converted to YUV color space to improve the performance.

where the hidden unit at layer  $l$  is denoted as:

$$H_l = \text{pool}(\tanh(W_l H_{l-1} + b_l)), \quad (2)$$

where  $l \in \{1, \dots, L-1\}$ ,  $L$  is the number of layers of  $N_s$ ,  $b_l$  is a bias parameter,  $H_0 = I_s$ , and  $W_l$  is a Toeplitz matrix. The hidden unit vector  $H_l$  is denoted as a regular convolution between kernels from  $W_l$  and the previous hidden unit vector  $H_{l-1}$ , squashed through a nonlinear tanh function, and pooled spatially. The function  $\text{pool}$  takes the maximum activation within the neighborhood via the max-pooling operator. Both  $W_l$  and  $b_l$  are acquired by training.

Once the network is trained properly, the outputs of the  $S$  networks are up-sampled to have the same size as  $f_1$ , the output of  $N_1$ , and concatenated to produce feature vector:

$$\mathcal{F} = [f_1, \mu(f_2), \dots, \mu(f_S)], \quad (3)$$

where  $\mu$  is the up-sample function. Finally, a two-layer neural network is used to produce the categorization for all pixels.

### B. Multiple Cells Labeling

The segmentation of overlapped cervical cells is regarded as a labeling problem in our task. Assuming each image consists of  $M$  cells (Cells' location and the number of cells are obtained from the network's output in sub section III-A), the proposed method assigns each cytoplasmic pixel  $i$  with a label  $\ell_i$  indicating the corresponding cell (the cell and its label are denoted by the same symbol  $\ell_i$ ). Let  $p_i$  denote the observation of pixel  $i$ . The energy function of assigning cells' labels  $\ell = \{0, 1, \dots, M\}$  to all pixels (0 is the label of background) is defined as

$$E(\ell; \Theta, w) = \sum_{i \in \nu} \varphi_i(\ell_i; p_i, \Theta^i) + \sum_{(i,j) \in \varepsilon} \varphi_{i,j}(\ell_i, \ell_j; p_i, p_j), \quad (4)$$

where  $\nu$  denotes the set of pixels in an image,  $\varepsilon$  is the neighborhood set of pixel  $i$ ,  $w$  is the parameter vector of the model, and  $\Theta$  is the shape parameters of cytoplasm estimated by the Gaussian kernel. The function  $\varphi_i(\ell_i; p_i, \Theta^i)$  is usually named as the unary potential, and it is referred as the "overlapping-based unary cost" in this paper. This energy term

captures the cost of pixel  $i$  labeled as  $\ell_i$  with its shape  $\Theta^i$ . In other words, it estimates the probability of pixel  $i$  from cell  $\ell_i$ . Therefore, this energy term essentially ensures cells can be extracted from the overlapping clump.  $\varphi_{i,j}(\ell_i, \ell_j; p_i, p_j)$  is known as the pairwise potential, which promotes consistency between labels  $\ell_i$  and  $\ell_j$  of neighboring pixels  $i$  and  $j$ .

**Overlapping-based Unary Costs:** The overlapping-based unary potential  $\varphi_i(\ell_i; p_i, \Theta^i)$  represents the cost of the pixel  $i$  assigned with label  $\ell_i$ , which depends on the cell's shape  $\Theta^i$  and is defined as

$$\varphi_i(\ell_i; p_i, \Theta^i) = p_c(i) p^{\ell_i}(i) \prod_{\ell_k, \ell_k \in O} (1 - p^{\ell_k}(i)), \quad (5)$$

where  $p_c(i)$  is the probability of pixel  $i$  predicted as cytoplasm, and  $p^{\ell_i}(i)$  is the probability of pixel  $i$  belonging to cell  $\ell_i$ , which is decided by the cell's shape parameter  $\Theta^i$ .  $O$  denotes the overlap of cell  $\ell_i$  and cell  $\ell_k$ . To address the fact that a pixel may have multiple labels when the pixel located in overlapped region, we accumulate the shape likelihoods. In other words, to compute the cost of pixel  $i$  assigning to cell label  $\ell_i$ , we consider shapes of all cells in the overlapping clumps. This causes pixel  $i$  to be more likely to be assigned to label  $\ell_i$  when it has the stronger evidence for cell  $\ell_i$  and weaker evidence for other cell  $\ell_k$ .

**Cell Mask Estimation for Computing  $p^{\ell_i}(i)$ :** We assume that cells' shape can be fitted by the Gaussian kernel, and the center of the Gaussian kernel can be replaced by nucleus center. Let  $(x_0, y_0)$  denote nucleus center, which is quantified by nucleus centroid in this paper, cell's shape is represented as

$$s(x, y; \sigma_x, \sigma_y, \theta) = \mathcal{N} e^{-(a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2)}, \quad (6)$$

where  $\mathcal{N}$  is the normalization factor,  $a, b, c$  are coefficients to control the shape's orientation, isotropy, and scale via  $\theta, \sigma_x$ , and  $\sigma_y$ , respectively. These coefficients are computed by

$$\begin{aligned} a &= \frac{\cos^2 \theta}{2\sigma_x^2} + \frac{\sin^2 \theta}{2\sigma_y^2}; \quad b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}; \\ c &= \frac{\sin^2 \theta}{2\sigma_x^2} + \frac{\cos^2 \theta}{2\sigma_y^2}. \end{aligned} \quad (7)$$

Therefore, the cell's shape can be estimated by controlling parameters  $\theta, \sigma_x$ , and  $\sigma_y$ . We implement the estimating procedure using a pattern matching process. This process uses predefined Gaussian kernels to carry out the convolution operation with the touching clump and then locates the kernel with maximum convolution response on the overlapping clump [45], [46]. Table I illustrates the entire procedure of cell mask estimation. Finally, after the shape parameters  $a^{\ell_i}, b^{\ell_i}, c^{\ell_i}$  of cell  $\ell_i$  have been obtained, the shape probability of pixel  $i$  in the position  $(x, y)$  belonging to cell  $\ell_i$  is derived as

$$p^{\ell_i}(i) = e^{-(a^{\ell_i}(x-x_0^{\ell_i})^2 + 2b^{\ell_i}(x-x_0^{\ell_i})(y-y_0^{\ell_i}) + c^{\ell_i}(y-y_0^{\ell_i})^2)}. \quad (8)$$

**Pairwise Potential Design:** The pairwise potential is the sum of a smoothness potential  $\varphi_s(p_i, p_j)$  and a data-dependent edge potential  $\varphi_e(p_i, p_j)$ , which consists of the contextual

TABLE I  
ALGORITHM OF CELL MASK ESTIMATION

1. Calculating the cell's centroid $(x_0, y_0)$ .	
2. Setting the orientation change step $n_\theta$ and scale range step $n_s$ .	
3. Initializing the scale range: $\sigma^{min}$ and $\sigma^{max}$ .	
4. Setting kernels' scale: $\sigma^{min} < \sigma_x < \sigma_y < \sigma^{max}$ .	
5. Generating the set of Gaussian kernels based on parameters $\{\sigma_x, \sigma_y, \theta\}$ .	
6. Computing the orientation response $\mathcal{M}_{\theta_k}^{\ell_i}$ by:	
$\mathcal{M}_{\theta_k}^{\ell_i} = \sum_{\sigma_x \in \mathcal{H}_x} \sum_{\sigma_y \in \mathcal{H}_y} \iint S(x, y; \sigma_x, \sigma_y, \theta_k) p_c(x, y) dx dy,$	(T1. 1)
where $\mathcal{H}_x$ and $\mathcal{H}_y$ are the set of scale $\sigma_x$ and $\sigma_y$ , respectively. $p_c(x, y)$ is the probability of pixel located in $(x, y)$ belonging to cytoplasm.	
7. Estimating mask's orientation as:	
$\hat{\theta}_{\ell_i} = \frac{180}{n_\theta} \operatorname{argmax}_{\theta_k} \mathcal{M}_{\theta_k}^{\ell_i}$	(T1. 2)
8. Computing the scale level's response by:	
$C(\sigma_x, \sigma_y, \hat{\theta}_{\ell_i}) = \iint p_c(x, y) S(x, y; \sigma_x, \sigma_y, \hat{\theta}_{\ell_i}) dx dy,$	(T1. 3)
9. Estimating mask's scale by:	
$\{\hat{\sigma}_x^{\ell_i}, \hat{\sigma}_y^{\ell_i}\} = \operatorname{argmax}_{\{\sigma_x, \sigma_y\}} C(\sigma_x, \sigma_y, \hat{\theta}_{\ell_i}).$	(T1. 4)

dependencies among all pairs of neighboring pixels and is computed as

$$\varphi_{i,j}(\ell_i, \ell_j; p_i, p_j) = \varphi_s(p_i, p_j) + \varphi_e(p_i, p_j). \quad (9)$$

$\varphi_s(p_i, p_j)$  penalizes the dissimilarities between two neighboring pixels and is given as

$$\varphi_s(p_i, p_j) = \begin{cases} 0 & \text{if } l_i = l_j, \\ \alpha & \text{otherwise.} \end{cases} \quad (10)$$

The edge potential  $\varphi_e(p_i, p_j)$  captures the differences of the color between the adjacent pixels and is denoted as

$$\varphi_e(p_i, p_j) = 1 - \exp\left(-\frac{\|I(i) - I(j)\|}{2\sigma^2}\right), \quad (11)$$

where  $\|\cdot\|$  is the Euclidean norm,  $I(i)$  is the intensity value of pixel  $i$ ,  $\alpha$  and  $\sigma$  are parameters obtained from cross-validation on the training database.

**Parameter Learning and Inference:** To address the problem of minimizing the energy in Eq. (4) and obtaining the optimal labeling  $\tilde{\ell} = \operatorname{argmin}_{\ell} E(\ell; \Theta, w)$ , we learn the model's parameter  $w$  to estimate the most likely labeling  $\tilde{\ell}$ . The parameter  $w$  is obtained by minimizing the loss function using gradient descent technique. And once the parameter  $w$  is learned, we compute the joint distribution of all output variables for the inference task [50], [51].

### C. Cell Boundary Refinement and Inference

The proposed labeling algorithm can accurately label pixels to the corresponding cells when the pixel only has a single label. However, pixels in overlapped regions would have multiple labels. An example is illustrated in Fig. 4, where  $s_i \wedge s_j$  is the overlapping region ( $s_i \cup s_j$  is referred as clump in this paper), pixels in this region are assigned to labels  $\ell_i$  and  $\ell_j$ . Therefore, we propose a dynamic multi-template deformation model to overcome the shortcomings of the multiple cells labeling model.

Let  $\phi : \Omega \rightarrow \mathcal{R}$  denote the level set function defined on the image domain. To split touching clumps with  $M$  cells, the

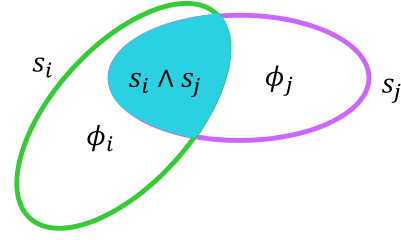


Fig. 4. Illustration of overlapping constraint. Pixels in overlapping region belong to both touching cells, while those in non-overlapping regions only belong to one cell.

energy function is defined as

$$\mathcal{F}(\{\phi_i\}_{i=1}^M) = \sum_{i=1}^M E(\phi_i) + \sum_{i=1}^M \sum_{j \in \varepsilon(i)} \psi(\phi_i, \phi_j), \quad (12)$$

where  $\mathbb{E}$  is the data energy generated from each cell independently,  $\varepsilon(i)$  denotes the neighborhood set of cell  $i$ .  $\psi$  is a binary function defined as an overlapping constraint (if no cell overlap with cell  $i$ , the binary function of cell  $i$  is set to 0).

**Cell-independent Energy:** The energy depends on appearance information and is calculated as

$$\mathbb{E}(\phi_i) = \omega \mathbb{E}_R(\phi_i) + \mathbb{E}_{Ext}(\phi_i), \quad (13)$$

where  $\omega$  is the weight parameter. The regularization term is computed by

$$\mathbb{E}_R(\phi_i) \triangleq \int_{\Omega} R(|\nabla \phi_i|) dx, \quad (14)$$

$$R(\phi_i) = \begin{cases} \frac{1}{(2\pi)^2} (1 - \cos(2\pi \phi_i)), & \text{if } \phi < 1, \\ \frac{1}{2} (\phi_i - 1)^2, & \text{if } \phi \geq 1. \end{cases} \quad (15)$$

The  $\mathbb{E}_{Ext}(\phi_i)$  is an external energy calculated from [52], [53]:

$$\mathbb{E}_{Ext}(\phi_i) = \sum_{m=1}^2 \lambda_m \int \left( \int K_m(x, y) M_m(\phi_i(y)) dy \right) dx, \quad (16)$$

$$K_m(x, y) = K_\sigma(x - y) |I(y) - \bar{u}_m|^2, \quad (17)$$

where  $M_1(\phi_i) = H(\phi_i)$  and  $M_2(\phi_i) = 1 - H(\phi_i)$ ,  $\lambda_m$  is a positive constant, and  $H$  is the Heaviside function.  $\bar{u}_1$  and  $\bar{u}_2$  are two values that approximate image intensities in  $\phi_i$  with positive and negative values around  $x$ , respectively.  $I(y)$  is the image intensity of pixels located in the region around  $x$ .

**Overlapping Constraint:**  $\psi$  is a binary function denoted as

$$\psi(\phi_i, \phi_j) = \int (H(\bar{s}_i)H(\bar{s}_j) - H(\phi_i)H(\phi_j)) \kappa dx, \quad (18)$$

$$\kappa = \max(\bar{s}_i(x), \bar{s}_j(x)), \quad (19)$$

where  $\bar{s}_i$  is the case-specific template of cell  $i$ . The binary function is designed to enhance consistency between shape template and level set function in overlapped regions.

**Case-Specific Shape Template Generation:** The template  $\bar{s}_i$  is generated by an adaptive hierarchical clustering algorithm. Given that the shapes  $s_i$  and  $s_j$  of two cells having

their centroids aligned to the same position, the similarity between  $s_i$  and  $s_j$  becomes

$$\begin{aligned} \mathbb{S}(s_i, s_j) &= 1 - \frac{\arg \min_{\theta \in \{1, 2, \dots, \pi\}} \iint (S(s_i(x, y)) - r_\theta(s_j(x, y)))^2 dx dy}{\iint S^2(s_i(x, y)) dx dy}, \end{aligned} \quad (20)$$

where  $r$  is the rotation matrix,  $S$  is the isotropic scaling factor obtained by computing the ratio of the mean radius between the shape  $s_i$  and  $s_j$ :  $S = \bar{r}_{s_i} / \bar{r}_{s_j}$ . To compute the similarity effectively, we use the polar coordinate space instead of the Cartesian space. We then employ an adaptive hierarchical clustering method [58] to cluster the cells' shape samples. Once the cluster centers are obtained, the cases-specific template  $\bar{s}_i$  of the cell  $s_i$  is obtained by

$$\bar{s}_i = \operatorname{argmin}_{k \in \{1, 2, \dots, K\}} \mathbb{S}(s_i, s_k). \quad (21)$$

**Energy Optimization and Evolution Process:** The energy function in Eq. (12) is optimized using the gradient flow technique. The results of multiple cells labeling is used to initialize the level set model. In addition, the shape templates are updated regularly, and the identified shape template is used to initialize the level set representation. This approach not only helps the level set model to jump out local minima in the process of energy optimization, but also achieves accurate segmentation results.

## IV. EXPERIMENTAL RESULTS

### A. Dataset Description

Two different datasets are utilized to evaluate the proposed method: 1) **ISBI 2015 Challenge Dataset**<sup>1</sup>: It is provided by the ISBI 2015 challenge consisting of eight cervical cytology images with  $1024 \times 1024$  pixels (the ground truth of images is also obtained from the challenge's website). Each image has 20 ~ 60 Pap stained cervical cells, which are distributed in 11 clumps with 3.3 cells per clump on average. 2) **Shenzhen University (SZU) Dataset**: This dataset is collected from Shenzhen Sixth People's Hospital, Shenzhen, China, and consists of 21 cervical cytology images with  $1360 \times 1024$  pixels. Each image has 30 ~ 80 H&E stained cervical cells, which are distributed in seven clumps with 6.1 cells per clump on average. Ground truth for this dataset is obtained by the pathologists.

### B. Implementation Details

**Code Development:** Our algorithm was implemented in MATLAB R2013a, and all experiments were performed on a machine with an Intel Core I5-3470 CPU (3.20 GHz) and 8 GB memory. The multi-scale CNN was implemented using the available toolbox.<sup>2</sup> Multiple cells labeling was performed

by using the UGM package,<sup>3</sup> and the level set framework on DRLSE<sup>4</sup> was adopted.

**Samples Selection for CNN Training:** 5650 background, 8590 cytoplasm, and 8560 nucleus pixels are taken from eight images in the SZU dataset to establish our training set. Patches with 3 different scales centered at the chosen pixels are used as the input of our CNN. Each patch has a size of  $32 \times 32$  pixels. The patch of the largest scale samples the  $32 \times 32$  pixels from  $128 \times 128$  pixels. We also rotate the sampling window to generate more training samples, samples are generated after window rotated from 0 to 180 degrees with a step of 15 degrees. In addition, we transform training samples into the YUV color space to boost the learning performance. Once the network has been trained properly, the same processing with training sample is implemented on images from SZU dataset to segment cell components. The images from ISBI 2015 challenge dataset are gray images, hence only the Y channel is used to segment cells.

### C. Performance Metric

The performance metrics proposed by ISBI 2015 challenge are used to evaluate our methods. One of the metrics is average Dice Coefficient (the Dice Coefficient is defined as:  $DC = 2|R_s \cap R_g| / (|R_s| + |R_g|)$ , where  $R_s$  is the segmented result and  $R_g$  is the ground truth). The average  $DC$  is measured over the "good" segmentation, which requires the cell with a  $DC$  value above 0.7. Another metric is the false negative rate ( $FN_O$ ) which counts the percentage of cells with  $DC \leq 0.7$ . The remaining metrics are the true positive rate ( $TP_P = TP / (TP + FN)$ ) and false positive rate ( $FP_P = FP / (FP + FN)$ ), where  $TP$  is the number of correctly segmented pixels,  $FP$  is the number of incorrectly segmented pixels, and  $FN$  is the number of missed segmented pixels.

### D. Parameter Sensitivity Analysis

The size of CNN input and number of CNN scales affect the learning efficiency and effectiveness of the proposed multi-scale CNN. Other parameters also impact the network model and even change the model's architecture. There is no scientific theory for selecting these parameters. Therefore, the current parameters of CNN models are selected manually. It is impractical to show sensitivity and effects of these parameters, because most of the combinations of parameters' value are unable to make CNN to well address this issue. Therefore, we only give the architecture of our CNN. In addition, the parameters  $\alpha$  and  $\sigma$  in Eq. (10) and Eq. (11), respectively, are decided by cross-validation strategy. We use four images from ISBI 2015 challenge dataset to determine their values of these two parameters. Hence, in this section, we only analyze and discuss the remaining parameters of our model: orientation change step  $n_\theta$  and scale range step  $n_s$  in cell mask estimation, the number of clusters in case-specific template generation, and parameters in level set function evolution. All parameters and their values of our model are illustrated in Table II.

<sup>1</sup>Dataset can obtain from the challenge's website: [http://cs.adelaide.edu.au/~zhi/isbi15\\_challenge/index.html](http://cs.adelaide.edu.au/~zhi/isbi15_challenge/index.html).

<sup>2</sup>Code available at: <http://cn.mathworks.com/matlabcentral/fileexchange/24291-cnn-convolutional-neural-network-class>.

<sup>3</sup>Code released at: <https://www.cs.ubc.ca/~schmidtm/Software/UGM.html>.

<sup>4</sup>Code provided at: <http://www.imagecomputing.org/~cmli/DRLSE/>.



TABLE II  
PARAMETERS AND THEIR SETTING OF THE PROPOSED MODEL

Parameter	Value(s)	Notes
$S$	3	Number of scales of multi-scale CNN.
$\omega_s$	Shown in Fig. 3	The configuration of plain CNN.
$n_\theta$	12	Orientation change step used to estimate cell mask.
$n_s$	$(\sigma^{max} - \sigma^{min})/10$	Scale range step used to estimate cell mask.
$\alpha$	0.1	Constant used to quantify the dissimilarities between two neighborhood pixels.
$\sigma$	0.2	Constant used to depict the color difference between the adjacent pixels.
$I_i$	5	Inner iteration of energy optimization and evolution process of deformation model.
$I_o$	10	Outer iteration of energy optimization and evolution process of deformation model.
$\omega$	0.02	Constant used to control the importance of the regularization and external energy terms.
$\Delta t$	2	Time step of energy optimization and evolution process of deformation model.
$T_s$	4	Number of shape template retrieves.
$I_c$	30	Iteration used to generate shape template.

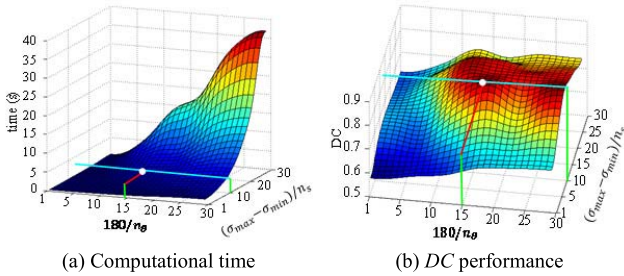


Fig. 5. Computational time and  $DC$  performance affected by values of parameters  $n_\theta$  and  $n_s$  of the cell mask estimation.

**Sensitivity of  $n_\theta$  and  $n_s$ :** The estimated masks are more accurate when the orientation change step  $n_\theta$  and scale range step  $n_s$  in the cell mask estimation are of lower values since more Gaussian kernels are generated. However, the computational cost is also increased. Fig. 5 shows the impact on processing time and segmentation results evaluated via  $DC$  by varying these two parameters. It is clear that the computational cost increases dramatically with the decrease of  $n_\theta$  and  $n_s$ , and no further improvement of  $DC$  is observed when  $180/n_\theta$  hovers around 15 and  $(\sigma^{max} - \sigma^{min})/n_s$  stabilizes at 10 ( $\sigma^{max}$  and  $\sigma^{min}$  are determined by the distance between the nucleus and its contour, we set  $\sigma^{max}$  as the value of maximum distance and  $\sigma^{min}$  as the half value of minimum distance). Therefore, we set  $n_\theta$  as 12 and  $n_s$  as  $(\sigma^{max} - \sigma^{min})/10$ , where  $DC$  peaks at 0.84 and the processing time is only 3.2s.

**Number of Clusters:** 320 cells with ground truth are chosen and clustered to generate case-specific shape templates. These cells are chosen manually from eight images. Only two images of them come from ISBI 2015 challenge dataset, which only contain normal cells, and other images come from SZU dataset, which contain both normal and abnormal cells. To ensure that the generated shape templates represent most of the cells' shapes, we carefully select cells with different

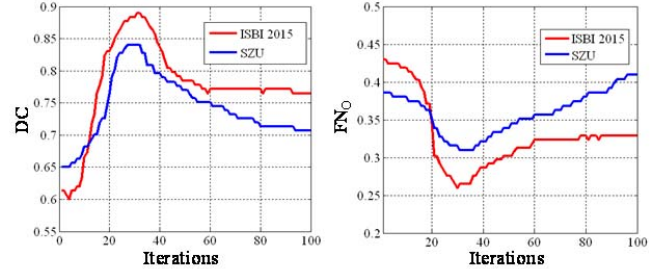


Fig. 6. Performance affected by increasing the number of iterations in the adaptive hierarchical clustering to generate case-specific templates in cell boundary refinement and inference.

shape variations for our training samples. We also control the number of cells with a similar shape. In addition, it should be noted that the reported performance of the proposed method is tested on the remaining images, which contains a total of 1200 cells. The number of cluster is determined by the iterations. Fig. 6 shows the results evaluated by  $DC$  and  $FNO$  at different iterations. We can see that when the iterations are under 10, the number of templates is above 212, and the performance is poor. As different parts of cell boundary have unequal ability to recover the cell's shape, more errors will be introduced when there is overlapping at the most effective regions. Templates should be generalized as possible so as to address the high overlapping cases. In addition, the number of templates decreases as iteration increases. All samples are clustered into one template when the iterations are above 96. The performance in this situation is not promising since the generated template would be unable to address shape variation of the cells. Consequently, we set the number of iterations as 30, which generates 42 templates leading to the best performance.

**Parameter Setting in Level Set Function Evolution:** As the intensity at cell's boundary changes indiscriminately when the cell is overlapped, we can only make minor modifications in each evolution. We set the inner iterations as five and outer iterations as 10. To speed up the curve evolution, we set the time step  $\Delta t = 2$ . The weight parameter  $\omega$  in Eq. (16) is set to 0.02 to fulfill the Courant-Friedrichs-Lewy condition. On the other hand, we retrieve new shape template based on the evolution result, and then the retrieved result is utilized to initialize the level set function. As most of the cells' templates are unchanged after retrieving more than four times, we set  $T_s = 4$ .

## E. Results

**Qualitative Results:** Fig. 7 shows some examples of segmentation results produced by the proposed model. The upper row is from ISBI 2015 dataset with overlapped cells ranges from 2 to 8, while the bottom row is from SZU dataset with the number of overlapped cells ranges from 2 to 10. Fig. 7 illustrates the strength of our model to split overlapped cells. The proposed method produces promising performance, even when the overlapping degree is high. Meanwhile, the proposed method has excellent edge capture ability under illumination variances, dye concentration, and uneven dyeing.

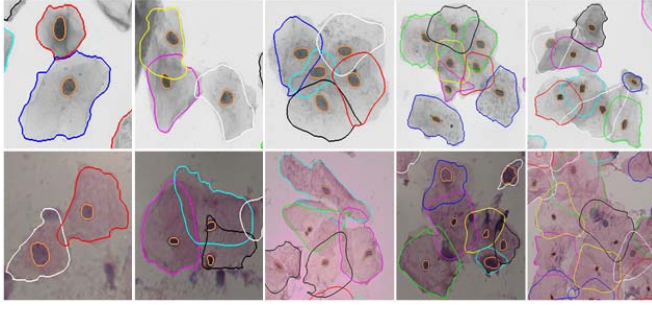


Fig. 7. Samples of the segmentation result for qualitative evaluation of proposed approach. The upper row is from ISBI 2015 dataset, while the bottom row derives from SZU dataset.

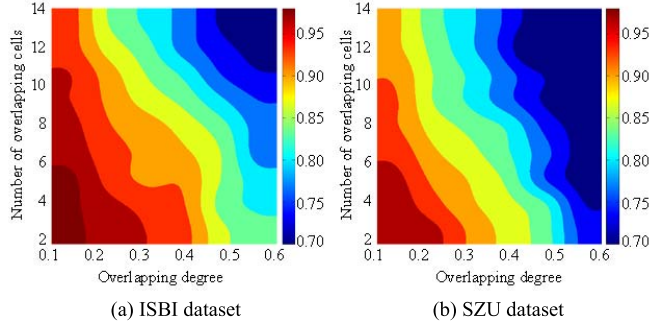


Fig. 8. Quantitative results evaluated by average  $DC$  against the overlapping degree and the number of overlapping cells. (a) The result from ISBI 2015 dataset. (b) The result from SZU dataset.

**Quantitative Results:** 200 cell clumps with different overlapping degree in each dataset are randomly selected, and the number of cells ranges from 2 to 14 per image. The overlapping degree between two cells is defined as:  $\max(|A \cap B|/|A|, |A \cap B|/|B|)$ , where  $A$  and  $B$  are the ground truth of cells and  $|\cdot|$  denotes the area of the ground truth. Fig. 8 shows the  $DC$  results with the different overlapping degree and the number of overlapped cells. The result with ISBI 2015 dataset is better than those with SZU dataset. This is attributed to the fact that the images from SZU dataset contain many abnormal cells, while the ISBI dataset contains only the healthy cells. Abnormal cells have more complicated shapes and variations in color, which increases the difficulty in splitting of these cells accurately. Nevertheless, the proposed method obtains  $DC$  value over 0.8 when the overlapping degree is under 0.5 with more challenging SZU dataset. Although the  $DC$  value decreases with the increasing number of overlapped cells, it is still above 0.75 when the number of cells is less than 14.

**Comparison with Other Methods:** For performance comparison, we conducted experiments using the joint level set based approach (JLS) [39] and the elliptical shape fitting method (ESF) [41]. For fair comparison, these algorithms work on the outputs of the CNN network to remove the bias generated from different image datasets. Table III summarizes the results of  $DC$ ,  $TP_P$ ,  $FP_P$ , and  $FN_O$  produced by these two methods and the proposed method. Compared with JLS, the proposed method produces a slightly higher  $DC$  value on ISBI 2015 dataset, while the improvement is more significant

TABLE III  
RESULTS FOR VARIOUS METHODS OF  
OVERLAPPING CELLS SPLITTING

	ISBI 2015 dataset				SZU dataset			
	$DC$	$TP_P$	$FP_P$	$FN_O$	$DC$	$TP_P$	$FP_P$	$FN_O$
ESF [41]	0.82	0.87	0.008	0.41	0.79	0.81	0.009	0.45
JLS [39]	0.88	0.92	0.002	<b>0.21</b>	0.81	0.83	0.007	<b>0.23</b>
Ours	<b>0.89</b>	0.92	0.002	0.26	<b>0.84</b>	<b>0.88</b>	<b>0.004</b>	0.31

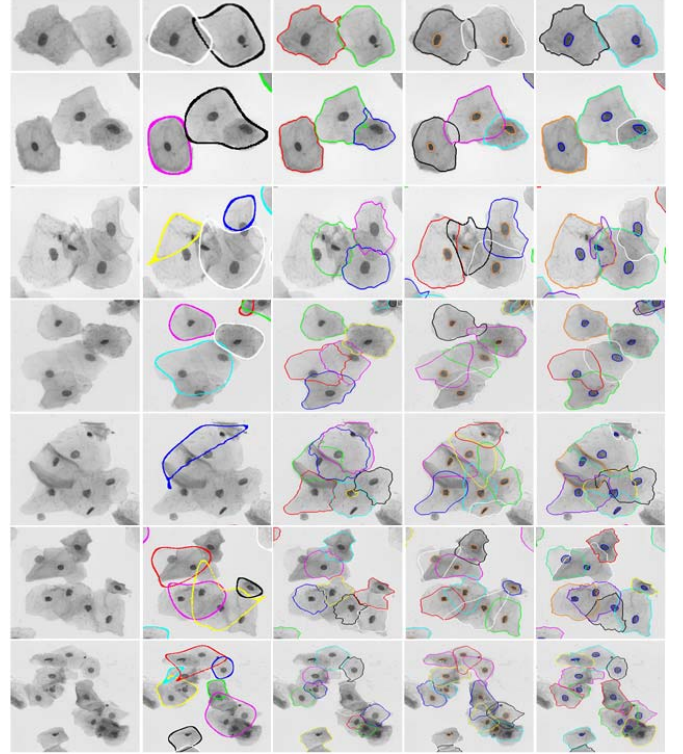


Fig. 9. Sample results from ISBI 2015 dataset. From left to right column: input images, outlined by JLS [39], ESF [41], the proposed method and the manual ground truth.

on SZU dataset. The similar trend can be found in  $TP_P$ . The proposed method shows significant superior performance in terms of  $FP_P$ . Compared with ESF, the proposed method shows a clear improvement in  $DC$  on both datasets. The advantages are also apparent for  $TP_P$  and  $FP_P$ .

Figs. 9 and 10 present the results of seven samples produced by JLS, ESF, and the proposed method from ISBI 2015 dataset and SZU dataset, respectively. The first row of Figs. 9 and 10 contain images with only two overlapped cells, and the overlapping degree is low in both figures. All three methods work well in this case, while ESF produces the best real boundary. ESF also shows its superiority in shape recovery when the contours' information is sufficient. From the second to seventh samples, the number of overlapped cells and overlapping degree are both increased. ESF gradually loses its performance due to insufficient boundary information. In general, ESF does not perform well for recovering cell's shape when most parts of the cell boundary are overlapped.

On the other hand, although JLS can accurately locate every cell, the computed boundary of cells is not natural.



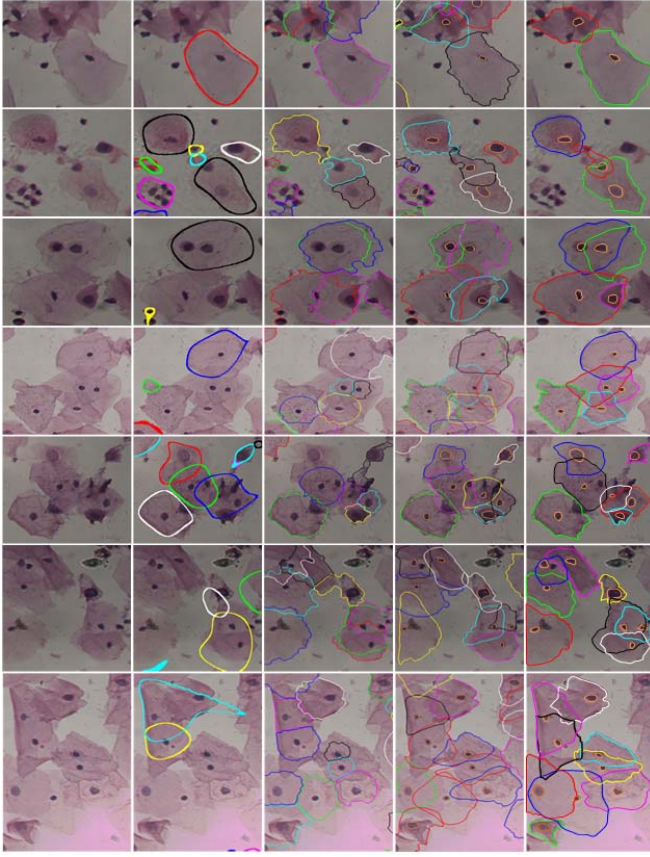


Fig. 10. Sample results from SZU dataset. From left to right column: input images, outlined by JLS [39], ESF [41], the proposed method and the manual ground truth.

TABLE IV

RESULTS FOR VARIOUS METHODS OF NUCLEI REGION SEGMENTATION

	ISBI 2015 dataset			SZU dataset		
	PPV	NPV	F1	PPV	NPV	F1
[25]	0.64	0.81	0.715	0.52	0.77	0.598
[36]	0.83	0.74	0.782	0.79	0.67	0.710
[28]	0.86	0.89	0.875	0.85	0.90	0.873
Ours	<b>0.95</b>	<b>0.93</b>	<b>0.940</b>	<b>0.94</b>	<b>0.92</b>	<b>0.912</b>

For the clumps with different sizes of overlapped cells, JLS always fails to find a balance of iterations of the level set function evolution. In contrast, the proposed method can locate every cell when the nucleus is detected accurately. Moreover, the boundary of cells is more natural because the case-specific shape template is incorporated by the proposed method. In addition, Figs. 9 and 10 show that the proposed method is still effective even when cells from clump are highly overlapped, and every level set function can obtain sufficient evolution simultaneously even at high degrees of overlapping.

We also compared the nuclei segmentation results with that in [25], [28], [36] based on the same criteria reported in [32]. Specifically, we compute the positive predictive value (PPV), negative predictive value (NPV), and F1 measurement. Table IV shows that the proposed approach with multi-scale CNN achieves the best performance in nuclei segmentation for both datasets.

TABLE V

RESULTS (MEAN AND STANDARD DEVIATION OF DC) OF CELL REGION SEGMENTATION PRODUCED BY MULTI-SCALE CNN AND SINGLE SCALE CNN

	ISBI 2015 dataset		SZU dataset	
	Cytoplasm	Nuclei	Cytoplasm	Nuclei
Scale (32×32)	0.79±0.17	0.82±0.12	0.76±0.21	0.83±0.12
Scale (64×64)	0.84±0.17	0.86±0.14	0.83±0.18	0.86±0.09
Scale (128×128)	0.81±0.21	0.84±0.10	0.80±0.20	0.84±0.11
Multi-scale	<b>0.91±0.14</b>	<b>0.93±0.06</b>	<b>0.90±0.17</b>	<b>0.92±0.08</b>

Finally, we evaluate the running time of the three methods, which are related to the number of overlapped cells. We observed that for a clump of 10 cells overlapping each other, ESF, JLS, and the proposed methods require 16 minutes, 58 minutes, and 14 minutes, respectively. It should be noted that our code is neither optimized nor speeded up via any multi-thread, GPU acceleration or parallel programming.

## V. DISCUSSIONS

### A. Effect of Multi-Scale CNN for Cell Segmentation

Multi-scale CNN not only makes contextual information more accessible, but also makes the management of its parameters easier than conventional CNN (plain CNN). The size of conventional CNN input is a problem. Small size provides a poor observation, which is insufficient to learn accurately. By contrast, large size increases the network parameters, and tuning or managing these parameters become more difficult. However, multi-scale CNN can address this issue. The input of our network is only  $32 \times 32$  pixels, but it can offer an observation with  $128 \times 128$  pixels. Moreover, the information from different scales is combined effectively by the multi-scale network. A large scale of networks' input provides not only contextual information, but also noise. However, this powerful combination ability can remove the noise and merge the contextual information. We compare the performance of multi-scale CNN and single scale CNN for cell segmentation (the performance is evaluated by Dice coefficient). As illustrated in Table V, multi-scale CNN gets more powerful ability in cell segmentation than single scale CNN.

### B. Effect of Multiple Cells Labeling for Clump Splitting

Multi-scale CNN can predict the pixels from cytoplasm or nuclei given a cytoplasm pixel, but it cannot predict which cell this pixel comes from. Segmenting objects from clump exactly is to estimate which cell the pixel belong to. In this paper, we propose a multiple cells labeling model to split overlapping objects. The performance is also compared with other state-of-the-art splitting methods. All compared methods are implemented on the output of multi-scale CNN. Table VI shows their results using the same metric.

From Table VI, we can see that the results of  $TP_P$  and  $FP_P$  are the same for all methods. The reason is that these two metrics only evaluate whether pixels are segmented correctly. In other words, they only concern whether the component of pixels is recognized accurately. In fact, as splitting methods only locate the splitting lines among different objects

TABLE VI

RESULTS OF CLUMPS SPLITTING PRODUCED BY VARIOUS METHODS

	ISBI 2015 dataset				SZU dataset			
	<i>DC</i>	<i>TP<sub>p</sub></i>	<i>FP<sub>p</sub></i>	<i>FN<sub>o</sub></i>	<i>DC</i>	<i>TP<sub>p</sub></i>	<i>FP<sub>p</sub></i>	<i>FN<sub>o</sub></i>
[30]	0.58	0.92	0.002	0.61	0.56	0.91	0.004	0.67
[31]	0.67	0.92	0.002	0.56	0.66	0.91	0.004	0.62
Ours	<b>0.71</b>	0.92	0.002	<b>0.38</b>	<b>0.68</b>	0.91	0.004	<b>0.43</b>

TABLE VII

RESULTS OF CLUMPS SPLITTING BY VARIOUS SHAPE INFERENCE METHODS IMPLEMENTED ON RESULTS OF MULTIPLE CELLS LABELING (NUMBER IN THE BRACKET DENOTES THE IMPROVED RATIO)

	ISBI 2015 dataset		SZU dataset	
	<i>DC</i>	<i>FN<sub>o</sub></i>	<i>DC</i>	<i>FN<sub>o</sub></i>
[41]	0.85 (3.5%)	0.31 (24.4%)	0.81 (2.5%)	0.32 (28.9%)
[39]	0.89 (1.1%)	0.18 (14.3%)	0.82 (1.2%)	0.20 (13.0%)
Ours	0.89 (N/A)	0.26 (N/A)	0.84 (N/A)	0.31 (N/A)

in the clump, the component of pixels is unchanged after they are performed. Hence, the values of these two metrics are determined by the multi-scale CNN. On the one hand, the *FN<sub>o</sub>* of all methods has a high value even if our method is the lowest. These results indicate many splitting objects with a *DC* value under 0.7. However, our method obtains the best results in *DC* and *FN<sub>o</sub>*. This superior performance suggests that integrating shape and appearance information is essential to split overlapping objects. Using shape information of objects can ensure that small objects cannot be dominated by big objects when they are overlapping, which can deal with objects with huge shape variations. However, splitting methods with distance information [30], [31] are limited for this situation. And the appearance information combined with the space and color information can promote the shape consistency between splitting line and object's contour. The objects can be segmented accurately even for objects with a high overlapping degree. In contrast, splitting methods based on region growing [30], [31] are unable to regulate the association between splitting line and object's contour.

### C. Effect of Dynamic Multi-Template Deformation Model for Shape Inference

Pixels in the overlapping region caused by any pair of cells belong to both cells, but splitting methods only divide them into the adjacent cell. Hence, cells are segmented incompletely because some parts of the cell are lost. These results are insufficient to accurately analyze object's structure information such as shape analysis. Therefore, the inference or refinement of cells' boundary is critical to find these lost parts of cells. However, the intensity of these lost parts is quite weak in images as they are overlapped by other cells. We address this issue by combining the splitting results, space relationship of overlapping cells, and case-specific shape template.

Table VII summarizes the performance of our deformation model and the state-of-the-art methods using our splitting results. Only the inference module of these methods is implemented for fair comparison and refinement. Although some cytoplasm pixels are wrongly segmented as background pixels, all three methods produce error that do not result in significant

difference between *TP<sub>p</sub>* and *FP<sub>p</sub>*. Therefore, we remove the results of these two metrics from Table VII, but provide the performance difference when other compared methods are not directly implemented on our splitting results. It is observed that shape inference and information provided by object's contour [41] can produce good result by integrating the knowledge of object's shape.

In addition, even if the object's shape can be estimated by contour's information, the estimated shape deviates from the real one due to the ignorance of objects' knowledge. However, human can recognize objects easily without the sufficient observation because of learned knowledge. Learning the objects' knowledge such as the case-specific shape template from images can lead to a more promising result of shape inference. Similar to [41], the method in [39] utilizes contour information to build the shape prior, but the guidance of their shape prior to the overlapped region is quite weak. In fact, the overlapped region needs more shape intervention since other information (e.g. color information) is usually deficient in these regions. Therefore, results in [39] are less than our method, the *DC* value in SZU dataset is also less than that of our method. Moreover, the case-specific shape template will be more abundant and accurate if it can learn from a large image data. This means the proposed method will achieve a better performance with the increasing of images. In addition, the performance of these two methods is boosted using our splitting method.

### D. Methodology Extensions

Quantitative analysis of objects, such as shape and staining pattern, is of vital importance not only for the analysis of medical microscope images, but also for other kinds of microscope images [54], [55]. However, this analysis in many microscope images usually suffers from two main problems: 1) the low contrast between objects and background; 2) numerous overlapped objects in microscope images. Therefore, an effective and efficient tool must be able to accurately segment the object from overlapping clump. It is of great importance for automatic quantitative analysis of microscope images.

The proposed method is an effective tool to address these two issues. Our approach can be applied in other applications. CNN has been proved in many research papers to be effective in recognizing objects including both medical images and natural images. To improve the performance of recognizing objects with low contrast, we develop a multi-scale CNN, which combines information in different scales to enhance objects' contrast. This network can automatically learn objects' feature regardless of imaging model, which can be extended to some other applications.

To split clumps with high overlapping degree, we propose a shape estimation method based on Gaussian kernel fitting. As most objects' shape can be roughly fitted by the combination of multiple ellipses, the proposed method can be applied in other similar scenario. The space relationship of overlapped objects and their appearance against the local splitting line are two evidences to predict their boundary in the overlapping region. Combining these evidences effectively is essential for many other splitting tasks.

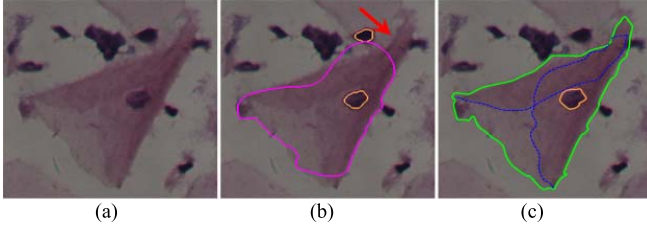


Fig. 11. Example of failure in our model with folded cells. (a) An original sample. (b) The results of the proposed method, red arrow indicates the segmentation error in the region that the orientation gradient of boundary changes sharply as the cell is folded. (c) The manual ground truth, the green contour indicates the desired boundary without the folded cell, the dot blue lines are the real boundaries caused by the folded cell.

It is difficult and even impossible to recover the overlapped parts without a high-level shape prior. Therefore, we propose a dynamic multi-template deformation model for boundary refinement and inference by learning case-specific shape template. Our deformation model integrates the overlapping constraint, appearances of all the clumps, and learned shape template to guide the inference of overlapped parts. Our inference method is only affected by the case-specific shape template, but shape template is learned by the training data. Therefore, this model is also applicable in some other fields where splitting of overlapped objects needs to be performed.

#### E. Limitations

The proposed method is based on the assumption that each nucleus represents a cell. Therefore, if there is overlapping at the nucleus, degraded performance would be expected from the proposed method. In fact, many works [32], [56], [57] have been developed to address the splitting of overlapped nuclei. Although the proposed method can address this issue, the processing time of splitting overlapped nuclei is about 12 minutes for an image with  $1360 \times 1024$  pixels, which contains 23 clumps and each clump contains four nuclei in average. Therefore, it is impractical to utilize the proposed approach to split overlapped nuclei and cytoplasm simultaneously without any code optimization or parallel pooling. The proposed method does not make use of the folded cell (see Fig. 11 for an example of folded cell). In this case, the assumption that the shape of cells can be approximated by the ellipse would become invalid. Even if a good performance can be obtained in terms of the evaluation metrics, significant error would occurred in the shape analysis using the segmentation results. The primary reason is that the real boundary of folded cells is inside the cell.

#### F. Future Directions

Future directions of our work mainly focus on the real-time implementation of the proposed method using code optimization via GPU acceleration. Moreover, we shall expand the image set and incorporate more image variations especially more abnormal cells to demonstrate the robustness of the proposed method. The training samples will be extended to integrate more shape templates so that more shape information can be obtained to recover overlapped parts of cells.

Future work also addresses the problem of folded cells, which is more challenging than splitting of the overlapped objects. Although some approaches can be utilized to address this issue (e.g., controlling the staining procedure for each slide), methods based on image processing will be explored. The interesting topic of the relationship between color information and folded degree will be investigated as well.

## VI. CONCLUSIONS

In this paper, we proposed a new approach to segment overlapped cervical cells in Pap smear images. The proposed model effectively addresses the segmentation issue, even for images with a large number of overlapped cells. We evaluated the proposed method with two different datasets. The experimental results demonstrated that the proposed approach outperforms state-of-the-arts methods. Furthermore, the proposed method is also effective in segmenting abnormal cells, even for images having large number of overlapped cells and high degrees of overlapping.

## APPENDIX

**Learning  $w$  for Multiple Cells Labeling:** The energy in Eq. (4) is minimized by learning the model's parameter  $w$  in this paper. Given that the training data  $X = \{(p^n, \ell^n)\}$ ,  $n = 1, \dots, N$ , according to the Maximum a Posteriori principle, the posterior probability of  $w$  is written as

$$p(w | X) = p(w) \prod_{n=1}^N \frac{p(\ell^n | x^n; w)}{p(\ell^n | x^n)}, \quad (A.1)$$

$$p(\ell | X; w) = \frac{1}{Z(w)} \exp(-E(\ell; \Theta, w)), \quad (A.2)$$

where  $Z(w) = \sum_{\ell} \exp(-E(\ell; \Theta, w))$  is the partition function. Here, we use a zero mean Gaussian prior  $p(w)$  distribution over the space of weight vectors and the Bayes' rule with the assumption of independent and identical distribution on  $X$ . Assuming the likelihood  $p(\ell^n | x^n; w)$  is independent of  $w$ , the regularized loss function would be

$$L(w) = \lambda \|w\|^2 - \sum_{n=1}^N \log p(\ell^n | x^n; w). \quad (A.3)$$

We estimate the parameter  $w$  by minimizing the loss function using gradient descent technique, and Bayesian rule is utilized to avoid fixing the parameter  $\lambda$  [47]–[49].

**Energy Optimization for Cell Boundary Refinement:** The energy function in Eq. (12) is optimized using the gradient flow technique. For each level set function  $\phi_i(x, y, t)$ , the gradient flow equation is computed by

$$\frac{\partial \phi_i}{\partial t} = - \frac{\partial \mathcal{F}(\{\phi_i\}_{i=1}^M)}{\partial \phi_i}, \quad (A.4)$$

$$\frac{\partial \mathcal{F}(\{\phi_i\}_{i=1}^M)}{\partial \phi_i} = \frac{\partial \mathbb{E}(\phi_i)}{\partial \phi_i} + \sum_{j \in \varepsilon(i)} \frac{\partial \psi(\phi_i, \phi_j)}{\partial \phi_i}. \quad (A.5)$$

The derivative of cell independent energy is calculated as

$$\begin{aligned} \frac{\partial \mathbb{E}}{\partial \phi_i} = & -\omega \operatorname{div} \left( \frac{R'(|\nabla \phi_i|) \nabla \phi_i}{|\nabla \phi_i|} \right) \\ & + \delta(\phi_i) \left( \sum_{m=1}^2 (-1)^m \lambda_m G_m \right) \end{aligned} \quad (A.6)$$



where  $\mathcal{G}_m$  is

$$\mathcal{G}_m(x) = \int K_\sigma(y-x) |I(x) - \bar{u}_m(y)|^2 dx, m = 1, 2. \quad (\text{A.7})$$

and  $\delta$  is a smoothed Dirac delta function and derivative of the Heaviside function  $H$ . This function is approximated by a smoothing function and defined as

$$H(x) = \frac{1}{2} \left[ 1 + \frac{2}{\pi} \arctan \left( \frac{x}{\varsigma} \right) \right], \quad (\text{A.8})$$

where  $\varsigma$  is a constant set to 1.5. Therefore, we have

$$\delta(x) = H'(x) = \frac{1}{\pi} \left( \frac{\varsigma}{x^2 + \varsigma^2} \right). \quad (\text{A.9})$$

Based on Euler-Lagrange equations, we have

$$\bar{u}_m = \frac{K_\sigma(x) * [M_i(\phi(x))I(x)]}{K_\sigma(x) * M_i(\phi(x))}, \quad m = 1, 2 \quad (\text{A.10})$$

Finally, we have the derivative of the overlapping constraint as

$$\frac{\partial \psi(\phi_i, \phi_j)}{\partial \phi_i} = \delta(\phi_i) H(\phi_j) \kappa. \quad (\text{A.11})$$

## REFERENCES

- [1] M. Schiffman, P. E. Castle, J. Jeronimo, A. C. Rodriguez, and S. Wacholder, "Human papillomavirus and cervical cancer," *Lancet*, vol. 370, no. 9590, pp. 890–907, 2007.
- [2] WHO/ICO. (2013). *Information Centre on HPV and Cervical Cancer (HPV Information Centre), Human Papillomavirus and Related Diseases Report in China*. [Online]. Available: [www.who.int/hpvcenter](http://www.who.int/hpvcenter)
- [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. Clin.*, vol. 62, no. 3, pp. 147–172, 2012.
- [4] Comprehensive Cervical Cancer Control: A Guide to Essential Practice, WHO, 2006.
- [5] G. N. Papanicolaou, "A new procedure for staining vaginal smears," *Science*, vol. 95, pp. 438–439, Apr. 1942.
- [6] G. G. Birdsong, "Automated screening of cervical cytology specimens," *Human Pathol.*, vol. 27, no. 5, pp. 468–481, 1996.
- [7] E. Bengtsson, "Computerized cell image analysis: Past, present, and future," *Image Analysis*. Berlin, Germany: Springer, 2007, pp. 395–407.
- [8] E. Bengtsson, "Recognizing signs of malignancy—The quest for computer assisted cancer screening and diagnosis systems," in *Proc. IEEE Conf. Comput. Intell. Comput. (ICCIC)*, Dec. 2010, pp. 1–6.
- [9] D. C. Wilbur *et al.*, "The becton dickinson focalpoint GS imaging system clinical trials demonstrate significantly improved sensitivity for the detection of important cervical lesions," *Am. J. Clin. Pathol.*, vol. 132, no. 5, pp. 767–775, 2009.
- [10] R. Sparks and A. Madabhushi, "Explicit shape descriptors: Novel morphologic features for histopathology classification," *Med. Image Anal.*, vol. 17, no. 8, pp. 997–1009, Dec. 2013.
- [11] L. F. Handfield, B. Strome, Y. T. Chong, and A. M. Moses, "Local statistics allow quantification of cell-to-cell variability from high-throughput microscope images," *Bioinformatics*, vol. 31, no. 6, pp. 940–947, Mar. 2015.
- [12] S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns," *Bioinformatics*, vol. 30, no. 18, pp. 2644–2651, Sep. 2014.
- [13] J. Z. Cheng *et al.*, "Computer-aided US diagnosis of breast lesions by using cell-based contour grouping," *Radiology*, vol. 255, no. 3, pp. 746–754, 2010.
- [14] J. Z. Cheng *et al.*, "ACCOMP: Augmented cell competition algorithm for breast lesion demarcation in sonography," *Med. Phys.*, vol. 37, no. 12, pp. 6240–6252, 2010.
- [15] J. Z. Cheng, C. M. Chen, Y. H. Chou, C. S. Chen, C. M. Tiu, and K. W. Chen, "Cell-based two-region competition algorithm with a map framework for boundary delineation of a series of 2D ultrasound images," *Ultrasound Med. Biol.*, vol. 33, no. 10, pp. 1640–1650, 2007.
- [16] M. E. Plissiti and C. Nikou, "A review of automated techniques for cervical cell image analysis and classification," in *Biomedical Imaging and Computational Modeling in Biomechanics*, vol. 4. Amsterdam, The Netherlands: Springer-Verlag, 2013, pp. 1–18.
- [17] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 2, pp. 233–241, Feb. 2011.
- [18] M. S. Nosrati and G. Hamarneh, "Segmentation of overlapping cervical cells: A variational method with star-shape prior," in *Proc. Int. Symp. Biomed. Imag. (ISBI)*, 2015, pp. 186–189.
- [19] 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.
- [20] 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.
- [21] P. Bamford and B. Lovell, "A water immersion algorithm for cytological image segmentation," in *Proc. Workshop APRS Image Seg.*, 1996, pp. 75–79.
- [22] N. A. M. Isa, "Automated edge detection technique for Pap smear images using moving K-means clustering and modified seed based region growing algorithm," *Int. J. Comput. Int. Manage.*, vol. 13, no. 3, pp. 45–49, 2005.
- [23] C. Bergmeir, M. G. Silvente, and J. M. Benitez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a Web-based software framework," *Comput. Methods Prog. Biomed.*, vol. 107, no. 3, pp. 497–512, Sep. 2012.
- [24] S. F. Yang-Mao, Y. K. Chan, and Y. P. Chu, "Edge enhancement nucleus and cytoplasm contour detector of cervical smear images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 353–366, Apr. 2008.
- [25] K. Li, Z. Lu, W. Y. Liu, and J. P. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF Snake," *Pattern Recognit.*, vol. 45, no. 4, pp. 1255–1264, 2012.
- [26] T. Guan, D. X. Zhou, and Y. H. Liu, "Accurate segmentation of partially overlapping cervical cells based on dynamic sparse contour searching and GVF snake model," *IEEE J. Biomed. Health Inf.*, vol. 19, no. 4, pp. 1494–1504, Apr. 2015.
- [27] L. Zhang *et al.*, "Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts," *Comput. Med. Imag. Graph.*, vol. 38, no. 5, pp. 369–380, 2014.
- [28] L. Zhang *et al.*, "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," *Cytometry A*, vol. 85, no. 3, pp. 214–230, 2014.
- [29] Y. Y. Song, L. Zhang, S. P. Chen, D. Ni, B. Y. Lei, and T. F. 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.
- [30] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 741–748, Mar. 2009.
- [31] H. G. Yang and N. Ahuja, "Automatic segmentation of granular objects in images: Combining local density clustering and gradient-barrier watershed," *Pattern Recognit.*, vol. 47, no. 6, pp. 2266–2279, 2014.
- [32] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1661–1677, Sep. 2011.
- [33] T. Janssens, L. Antanas, S. Derde, I. Vanhorebeek, G. V. D. Berghe, and F. G. Grandas, "Charisma: An integrated approach to automatic H&E-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting," *Med. Image Anal.*, vol. 17, no. 8, pp. 1206–1219, 2013.
- [34] S. Arslan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, "Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1121–1131, Jun. 2013.
- [35] F. Moussavi, Y. Wang, P. Lorenzen, J. Oakley, D. Russakoff, and S. Gould, "A unified graphical models framework for automated mitosis detection in human embryos," *IEEE Trans. Med. Imag.*, vol. 33, no. 7, pp. 1551–1562, Jul. 2014.

- [36] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.
- [37] S. Ali and A. Madabhushi, "An integrated region-boundary-shape-based active contour for multiple object overlap resolution in histological image," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1448–1460, Jul. 2012.
- [38] J. P. Bergeest and K. Rohr, "Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals," *Med. Image Anal.*, vol. 16, no. 7, pp. 1436–1444, 2012.
- [39] Z. Lu, G. Carneiro, and A. P. 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.
- [40] P. Quelhas, M. Marcuzzo, A. M. Mendonça, and A. Campilho, "Cell nuclei and cytoplasm joint segmentation using the sliding band filter," *IEEE Trans. Med. Imag.*, vol. 29, no. 8, pp. 1463–1473, Aug. 2010.
- [41] C. Park, J. Z. Huang, J. X. Ji, and Y. Ding, "Segmentation, inference and classification of partially overlapping nanoparticles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 669–681, Mar. 2013.
- [42] S. Zafari, T. Eerola, J. Sampo, H. Kälviäinen, and H. Haario, "Segmentation of overlapping elliptical objects in silhouette images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5942–5952, Dec. 2015.
- [43] Y. Y. Song, D. Ni, L. He, S. P. Chen, B. Y. Lei, and T. F. Wang, "Segmentation and splitting of touching vaginal bacteria based on superpixel and effective distance," in *Proc. IEEE Workshop Int. Conf. Data Mining*, 2014, pp. 976–981.
- [44] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [45] H. Kong, H. C. Akakin, and S. E. Sarma, "A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1719–1733, Dec. 2013.
- [46] H. Kong, S. E. Sarma, and F. Tang, "Generalizing Laplacian of gaussian filters for vanishing-point detection," *IEEE Trans. Intell. Transp.*, vol. 14, no. 1, pp. 408–418, Jan. 2013.
- [47] S. Nowozin, P. V. Gehler, and C. H. Lampert, "On parameter learning in CRF-based approaches to object class image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 98–111.
- [48] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, pp. 1–305, Jan. 2008.
- [49] S. Nowozin and C. H. Lampert, "Structured learning and prediction in computer vision," *Found. Trends Comput. Graph. Vis.*, vol. 6, pp. 185–365, Mar. 2011.
- [50] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 531–552, Mar. 2011.
- [51] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.
- [52] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.
- [53] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.
- [54] Y. Pan *et al.*, "Size-dependent cytotoxicity of gold nanoparticles," *Small*, vol. 3, no. 11, pp. 1941–1949, 2007.
- [55] C. L. Nehl, H. W. Liao, and J. H. Hafner, "Optical properties of star-shaped gold nanoparticles," *Nano Lett.*, vol. 6, no. 4, pp. 683–688, 2006.
- [56] 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.
- [57] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.
- [58] H. G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Trans. Geosci. Remote*, vol. 46, no. 7, pp. 2097–2111, Jul. 2011.