

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

1

Contour-Seed Pairs Learning-Based Framework for Simultaneously Detecting and Segmenting Various Overlapping Cells/Nuclei in Microscopy Images

Jie Song, Liang Xiao, *Member, IEEE*, and Zhichao Lian, *Member, IEEE*

Abstract—In this paper, we propose a novel contour-seed pairs learning-based framework for robust and automated cell/nucleus segmentation. Automated granular object segmentation in microscopy images has significant clinical importance for pathology grading of the cell carcinoma and gene expression. The focus of past literature is dominated by either segmenting a certain type of cells/nuclei or simply splitting the clustered objects without contours inference of them. Our method addresses these issues by formulating the detection and segmentation tasks in terms of a unified regression problem, where a cascade sparse regression chain model is trained and then applied to return object locations and entire boundaries of clustered objects. In particular, we first learn a set of online convolutional features in each layer. Then, in the proposed cascade sparse regression chain, with the input from the learned features, we iteratively update the locations and clustered object boundaries until convergence. In this way, the boundary evidences of each individual object can be easily delineated and be further fed to a complete contour inference procedure optimized by the minimum description length principle. For any probe image, our method enables to analyze free-lying and overlapping cells with complex shapes. Experimental results show that the proposed method is very generic and performs well on contour inferences of various cell/nucleus types. Compared with the current segmentation techniques, our approach achieves state of the art performances on four challenging datasets, i.e., the kidney renal cell carcinoma histopathology dataset, drosophila Kc167 cellular dataset, differential interference contrast red blood cell dataset, and cervical cytology dataset.

Index Terms—Microscopy image segmentation, contour-seed pairs learning, cascade sparse regression chain model, complete contour inference

Manuscript received August 21, 2017; revised January 11, 2018 and April 24; accepted July 14, 2018. This work was supported in part by the Natural Science Foundation of China under Grant 61571230, Grant 11431015, and Grant 61171165, in part by the Fundamental Research Funds for the Central Universities under Grant 30915012204 and 30917011324, in part by the National Scientific Equipment Developing Project of China under Grant 2016YFF0103604, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20161500 and BK20150784, in part by the Research Innovation Program for College Graduates of Jiangsu Province under Grant KYLX15_0377, and in part by the Six Top Talents Project of Jiangsu Province under Grant 2012DZXX-036. (Corresponding author: Liang Xiao.)

J. Song and Z. Lian are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.

L. Xiao is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (xiaoliang@mail.njust.edu.cn).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org/>, provided by the author. The material includes additional experimental results. Contact xiaoliang@mail.njust.edu.cn for further questions about this work.

I. INTRODUCTION

AUTOMATED detection and segmentation of overlapping granular objects remains one of the most challenging issues in the analysis of microscopy images [1]. The microscopic image analysis typically employs fluorescence imaging of samples in a high-throughput format and reports quantitatively on parameters such as identification of each individual, changes of their morphology, and spatial distribution of targets [2]. The key procedure is to separate each individual cell/nucleus [1]-[5], and a motivating application is the development of molecularly targeted cancer therapies. It requires detecting the effect of screening samples on cell growth, differentiation, migration, and death [6]. However, as the pre-requisite step of micrograph analysis, fully automated segmentation is very problem-specific [5]. This makes the development of a generally applicable cell/nucleus segmentation method a huge challenge. Therefore, traditionally researchers have to manually segment various granular objects in microscopy images, which is time consuming, exhausting and prone to human error.

Granular structures of micrographs appear in many contexts and at many scales. Image segmentation for cell/nucleus analysis is generally a difficult problem due to large variability and complexity of the data. Specifically, the high variability in modern microscopy images is caused by a number of factors including differences in slide preparation and image acquisition. The complexity of data attributes to the variation in shape, size, intensity, and texture of microscopic objects during their life cycle. Typically, intensity and texture variations exist in inter and intra images due to the nonuniform absorption of the stains, which will reduce the contrast between background and objects, and further blur the boundaries of cells/nuclei. This, in turn, may result in identifying a single object as fragments. Furthermore, cells/nuclei are often organized in touching/overlapping clusters and have heterogeneous aspects. These touching/overlapping structures take up more space, which makes them more confluent. A high occlusion level in cell/nucleus clump decreases contrast among the objects, which also makes their boundaries more difficult to perceive. As a result, multiple overlapping objects are often segmented as a single object. All these factors make it difficult to accurately segment the granular objects from microscopy images. A promising solution for the dilemma is the incorporation of human knowledge into a learning-based system [7]-[9].

The detection and segmentation of microscopic structures, such as cells or nuclei are now a well-studied problem [1],

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

2

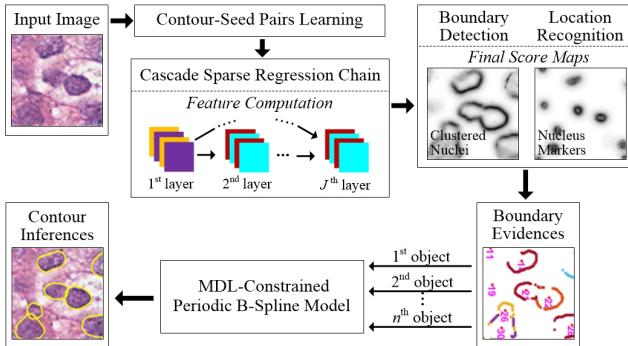


Fig. 1. The proposed CSRCM-based detection and segmentation framework.

[3]-[5], [10]-[14] for which a large number of approaches have existed in the literature and new methodologies continue to be investigated. Current systems can segment the clustered nuclei in histopathology images [7], [15]-[17], extract both nuclei and cytoplasm from RNA interference (RNAi) multi-channel images [18], [19], segment low-contrast blood cells in differential interference contrast (DIC) microscopy images [20], [21], and separate overlapping cells in cervical cytology specimens [9], [22], [23]. However, only recently has the comprehensive framework for various granular object segmentation been presented in different microscopy images [10]-[13].

To achieve the goal of joint cell/nucleus detection and segmentation, we propose a robust contour-seed pairs learning (CSPL) framework. Our method is designed to return a score with a well defined maximum along the boundaries, as described in Section III. The unified framework of the proposed method is shown in Fig. 1. First, we incorporate a sparse constraint to learn a set of features from online learned convolutional filters in each layer of chain before the cascade training. Second, in the proposed cascade sparse regression chain model, with the input from the learned features, we iteratively update the location and clustered objects' boundary scores by using gradient tree boosting technique [24]. When updating the scores, the local image appearance features are used in the first layer and added into the new ones learned in the subsequent layers [25]. The feature extraction and score estimation interact to reach convergence to produce the object markers and clustered objects' boundaries. In this way, the boundary evidences of each object can be easily delineated via cluster-boundary-to-marker association and be further fed to a complete contour inference procedure [26] optimized by the minimum description length (MDL) principle [27], [28]. By contrast, it is more robust to image noises or artifacts than the periodic B-spline model with shape prior constraint [29].

The major contributions of our method are as follows:

- **Unified framework:** We solve the detection and segmentation problem by formulating them in terms of a unified regression problem. Our framework jointly learns object locations and boundaries which is novel comparing to most of the existing methods that estimate them separately.
- **Cascade sparse regression chain model (CSRCM):** We train a regressor instead of a classifier to avoid the problems of producing multiple responses on the boundaries and poor localization accuracy. We improve upon the

regression of a single layer by introducing an additional refinement inspired by the auto-context algorithm, which was originally proposed for image segmentation.

- **Experiments:** The experiments evaluated on four types of disparate datasets show that the proposed method boosts the performances of both cell/nucleus detection and segmentation, comparing to the state-of-the-art algorithms.

The remaining of this paper is organized as follows: Section II reviews the related work based on the existing challenges. Section III describes the proposed CSRCM for the image detection and segmentation problem, while Section IV formulates the MDL criteria for contour inference. In Section V, we conduct extensive experiments on various microscopy datasets to substantiate the superiority of the proposed method over the other existing ones. Section VI concludes the paper.

II. RELATED RESEARCH

The methods described in the literature of microscopy image segmentation have mainly focused on three challenges. A comprehensive review on automated cell/nucleus detection and segmentation approaches on digital pathology and microscopy images can be found in [4].

A. Stained Nucleus Segmentation in Crowded Scenes

The first challenge involves local or global texture variation and appearance inhomogeneity, which often occur on stained nuclei in histopathology images and RNAi nuclei images. The segmentation in this situation can be based on watershed transform [30], [31], graph theory [16], [32], active contour [17], [33], spline fitting [15], and learning-based model [7], [34].

For low object-density images with isolated nuclei or infrequent overlaps between them, the classical watershed algorithm, and its many variants perform very well. For example, Veta *et al.* [30] have incorporated the gradient and orientation information into watershed framework for nucleus segmentation in breast cancer biopsy images. Graph-based approaches that model one image as a graph are also applicable to this case. For example, the dual graph for a Voronoi diagram, Delaunay triangulation, is used to segment clumps of nuclei in [32]. Recently, to improve the segmentation accuracy, simple learning-based methods have attracted much attention, such as using template matching [34] to delineate nuclear boundaries in histopathology dataset.

However, for the high-density images, nucleus segmentation might need more complicated and sophisticated algorithm designs to provide desired results. For example, Al-Kofahi *et al.* [16] have applied a graph cut algorithm with α -expansion and graph coloring to automated nucleus segmentation in breast histopathology images. On the other hand, active contours and spline fitting have been proved in many research papers to be effective in segmenting dense clustered nuclei. Fatakdawala *et al.* [33] have proposed an expectation maximization-driven geodesic active contour with an overlap resolution scheme. Ali *et al.* [17] have presented an adaptive active contour scheme that combines shape, boundary, region statistics, and mutual occlusion terms in a multi-level set formulation. Song *et al.* [15] have used optimized split point pairs-driven B-spline model to

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

3

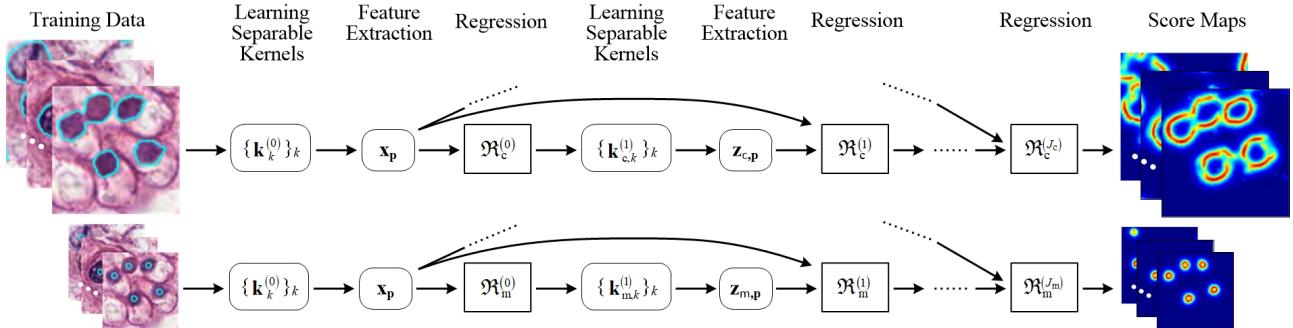


Fig. 2. Flowchart of the training procedure for our proposed cascade sparse regression chain model. The input image is convolved with a bank of separable kernels $\{k_k^{(0)}\}_k$ obtained from a set of convolutional filters $\{f_j^{(0)}\}_{j=1}^L$, which is learned online by sparse coding technique, to extract image appearance features. The features are used as input to regressors $\mathfrak{R}^{(0)}$. The outputs of the trees are then convolved with new learned kernel banks to extract context features. These features are fed together with the image appearance features to a second layer of regressors $\mathfrak{R}^{(1)}$. This process is iterated J times to produce the final score maps. Note that, location recognition (subscript 'm') and boundary detection of clustered objects (subscript 'c') are achieved in a unified regression framework but two separate chains.

infer overlapping nuclear contours in breast and kidney cancer biopsy images. In order to tackle the remaining limitations, deep learning-based techniques have been successfully applied for nucleus segmentation within breast cancer samples [7].

B. Both Nucleus and Cytoplasm Segmentation under Various Occlusion Level

In real life, many applications require the processing algorithm to handle both nuclei and cytoplasm, which co-exist within the same image (e.g., RNAi multi-channel images and cervical cytology specimens).

Nuclei are used as seed or marker regions to find initial contours when they are non-overlapped. Nuclei detection can be achieved by a thresholding method [35], [36], the maximally stable extremal region algorithm [22], and learning-based approaches, such as training a random decision forest [23] and learning a multiscale convolutional neural network [9]. Then, using the extracted nuclei as starting points, Chen *et al.* [35] have combined morphological algorithm with constraint factor graph cuts-based active contours to segment tightly clustered cells. In Ushizima *et al.* [36], segmentation of cytoplasm is achieved by calculating the Voronoi diagram of previously detected nuclei as constraints to the cytoplasm boundary. In order to provide the inference of overlapped part between cells, Nosrati and Hamarneh [23] have introduced a Voronoi energy term into the variational segmentation framework with star-shape prior. Recently, Lu *et al.* [22] have proposed a joint level set representation with geodesic-based shape prior.

However, overlapping areas may exist between the nuclei. To solve this problem, Yan *et al.* [19] have firstly introduced an enhanced distance transform watershed to obtain the initialization for segmentation of cells and Quelhas *et al.* [18] have used a sliding band filter to detect the nuclei. In [19], cells are then extracted by minimizing the energy functional of level set associated with the interaction model. In [18], cytoplasm shape estimation is guided by the nuclear detections followed by overlap correction and shape regularization steps to further refine the estimated cell shapes.

C. Low-Contrast Cell Segmentation with Variational Shapes

Segmenting the cells from DIC red blood cell images is quite

a challenge. This is mainly because: 1) Most DIC red blood cell images have low contrast with variational cell shapes and sizes. 2) The boundaries of some cells are vague or even missing. A certain amount of algorithms has been proposed over the years to tackle the problem of such cell segmentation. The method proposed in [37] is based on an iterative voting technique [38], which defines the initial condition for detecting the boundary using PDEs in a level set framework. Similar work is done by Chen *et al.* [21] embedded with local complex phase information and shape priors. Recently, several learning-based methods have also been applied to the cell segmentation of this category. For instance, Horn *et al.* [20] have focused on the learning of a probabilistic boundary model followed by an extended segmentation method using dynamic programming to identify the red blood cells in DIC microscopy images. Furthermore, Ronneberger *et al.* [8] have presented a called u-net framework to combine the features of different layers to calculate the final segmentation directly in the lowest layer using 11-layers u-shaped network architecture.

III. CASCADE SPARSE REGRESSION CHAIN MODEL

Our method aims to incorporate microscopic object detection and segmentation into a unified learning framework. A key technique is to learn a cascade sparse regression chain model. We train the proposed chain model for each task separately, but in a similar routine. In this section, a common framework is described for both of them.

A. Cascade Regression Chain

Our proposed cascade sparse regression chain consists of successive convolutional-pooling layers and regression layers (see Fig. 2), followed by non-maximum suppression. This is a fully different model from deep convolutional neural network [39], which produces the probability distributions in the last fully-connected layer. In addition, our model extracts superior joint features from the original image and estimated score maps by learning the input data distribution with a sparse prior. These, in turn, result in producing more accurate score values over pixels iteratively. As we can see from Fig. 3, the updated score maps of subsequent layers progressively correct the incorrect

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

4

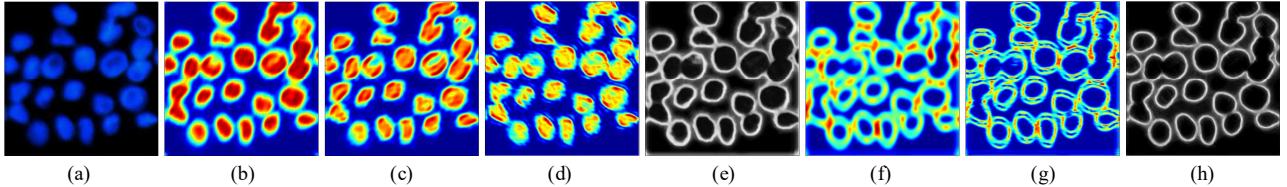


Fig. 4. (a) original image; (b) appearance feature map obtained when random possible locations of a neighborhood are considered, (c) followed by a subsampling, and (d) the one obtained without considering such a feature pool; (e) score map $\mathfrak{R}_c^{(0)}$ obtained from the result of Fig. 4(c); (f) context feature map computed from $\mathfrak{R}_c^{(0)}$ using our defined feature pool and (g) not using such a feature pool; (h) final score map $\mathfrak{R}_m^{(1)}$ obtained from the result of Fig. 4(f) (after subsampling). Note that the feature pool produced in the first layer is only composed of appearance features, which, however, also contains context ones from the second layer.

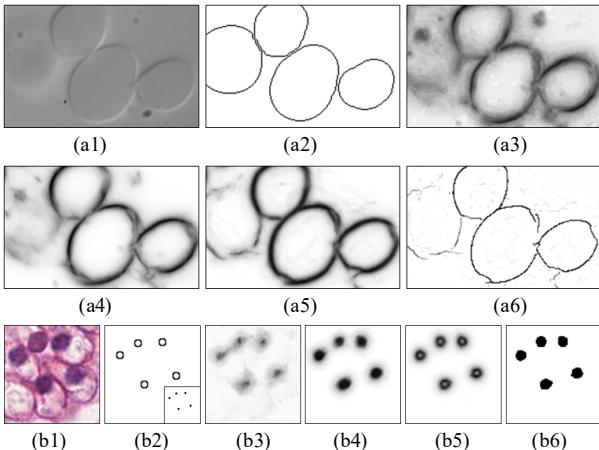


Fig. 3. Examples of predicting score maps of nucleus markers and clustered cells on testing images: (a1) and (b1), respectively; (a2), (b2) ground truth boundaries; (a3), (b3) score maps generated by learning two different regressors: $\mathfrak{R}_c^{(0)}$ and $\mathfrak{R}_m^{(0)}$, respectively; (a4), (b4) $\mathfrak{R}_c^{(1)}$ and $\mathfrak{R}_m^{(1)}$, respectively; (a5), (b5) $\mathfrak{R}_c^{(4)}$ and $\mathfrak{R}_m^{(5)}$, respectively. (a6) boundaries detected after non-maximum suppression on $\mathfrak{R}_c^{(4)}$, and (b6) thresholding result on $\mathfrak{R}_m^{(5)}$.

values on the background and discontinuity of the boundaries produced by the previous layers.

Our cascade sparse regression chain model (CSRCM) stacks multiple processing layers based on the auto-context algorithms [25] such that the output of j^{th} layers will serve as the input of $(j+1)^{\text{th}}$ layers. Specifically, during training of the subsequent layers, the appearance features \mathbf{x}_p^i are extracted from the original images and we use the score maps $\mathfrak{R}^{(j-1)}$ ($j=1, \dots, J+1$) obtained from the j^{th} regression layer as an additional source to extract the context features $\mathbf{z}_p^i(\mathfrak{R}^{(j-1)}) \in \mathbb{R}^D$. Then, the score maps are iteratively updated by using the new training samples $\{(\mathbf{x}_p^1, \mathbf{z}_p^1(\mathfrak{R}^{(j-1)}), y_p^1), \dots, (\mathbf{x}_p^N, \mathbf{z}_p^N(\mathfrak{R}^{(j-1)}), y_p^N)\}$, and fed into the following layers to produce the $\mathfrak{R}^{(j)}$, which is approximated by gradient boosting function [24]:

$$\mathfrak{R}^{(j)}\left(\mathbf{x}_p^i, \mathbf{z}_p^i(\mathfrak{R}^{(j-1)})\right) = \sum_{m=1}^M \alpha_m^{(j)} h_m^{(j)}\left(\mathbf{x}_p^i, \mathbf{z}_p^i(\mathfrak{R}^{(j-1)})\right) \quad (1)$$

The same boosting training procedure is repeated as the previous layers, and finally, a sequence of learned regressors $\{\{h_m^{(0)}\}_m, \dots, \{h_m^{(J)}\}_m\}$ will be generated. The proposed CSRCM is constructed as follows

$$\hat{\mathfrak{R}}^{(J)}\left(\mathbf{x}_p, \left\{ \mathbf{f}_{i_j}^{(J)} * \dots * \hat{\mathfrak{R}}^{(1)}\left(\mathbf{x}_p, \left\{ \mathbf{f}_{i_0}^{(0)} * \hat{\mathfrak{R}}^{(0)}\left(\mathbf{x}_p\right)\right\}_1^{L_0}\right)\right\}_1^{L_J}\right) \quad (2)$$

where the $\{\mathbf{f}_i\}_{i=1}^L$ is a convolutional filter bank learned in the unsupervised way by enforcing sparsity from a set of training images, as discussed in the next section, and $*$ denotes the convolution operator. Fig. 2 summarizes the proposed CSRCM and Fig. 3 shows examples of the score maps obtained by applying our CSRCM on testing micrographs.

The non-maximum suppression picks up the local maximum along a line perpendicular to the local orientation on the score maps. We estimate the orientation using the eigenvectors of optimally oriented flux matrix [40], which we found more robust than Hessian operator. Results obtained from this non-maximum suppression layer are shown in Fig. 3(a6).

B. Feature Extraction by Online Learned Filters

Our framework can utilize any kind of features from original image pixels and estimated scores, such as Haar-like [41], SIFT [42], and HOG [43] features, for regression. In this work, we learn convolutional features as in [44], [45] due to their ability to capture the patterns inherent in the input data. Recently, creating an overcomplete dictionary of features, sparse combinations of which can be used to represent images, has emerged as a powerful tool for object recognition [46], [47].

1) Low-Level Image Appearance Features: Suppose we have learned filters \mathbf{f}_i 's from given images and this was done with a learning procedure via sparse coding techniques [45]-[47]. We can then use this learned feature detector anywhere in the image \mathbf{I} (Fig. 4(a) for example) as kernels. Specifically, we can take the \mathbf{f}_i 's and convolve them with \mathbf{I} , thus obtaining a different feature activation value at each location \mathbf{p} in the image \mathbf{I} . In practice, the locations within $\mathbf{r} \in \mathbb{R}^2$ pixels away from the current pixel \mathbf{p} are randomly selected into the possible feature pool. This guarantees that additional image information will be considered, if they are indeed informative, but also produces a much larger pool of possible features. We handle this potentially large number by considering only a random subset of all possible locations. This also has the added benefit to reduce overfitting. Therefore, we simply compute our convolutional features \mathbf{x}_p as follows

$$\mathbf{x}_p = \left[(\mathbf{f}_1 * \mathbf{I})(\tilde{\mathbf{p}}), (\mathbf{f}_2 * \mathbf{I})(\tilde{\mathbf{p}}), \dots, (\mathbf{f}_L * \mathbf{I})(\tilde{\mathbf{p}}) \right]^T \quad (3)$$

where $\tilde{\mathbf{p}} = (\mathbf{p} + \mathbf{r})$, s.t. $|\mathbf{r}| \leq \text{rand}(T)$ for $T > 0$, and \mathbf{f}_i 's are learned online in the unsupervised manner by optimizing the function:

$$\arg \min_{\{\mathbf{f}_i\}, \{\mathbf{a}_i^{(j)}\}} \sum_j \left\| \mathbf{I}^{(j)} - \sum_i \mathbf{f}_i * \mathbf{a}_i^{(j)} \right\|_2^2 + \lambda \sum_i \left\| \mathbf{a}_i^{(j)} \right\|_1 \quad (4)$$

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

5

where $\{\mathbf{a}_i^{(j)}\}_i$ is the corresponding representation of each image $\mathbf{I}^{(j)}$. For color images, we use the $L^*u^*v^*$ decomposition and compute the convolutional features on three channels separately. We further use the tensor decomposition technique as in [44], [48] to obtain a decomposition of the original kernel bank in order to speed up convolutions. In this way, convolving the image with all the \mathbf{f}_i 's amounts to convolving it with the rank-1 separable kernels \mathbf{k}_k 's and then linearly combining the results. Therefore, each \mathbf{f}_i can be written as

$$\mathbf{f}_i = \sum_{k=1}^K c_i^k \mathbf{k}_k \quad (5)$$

where rank-1 kernels \mathbf{k}_k 's are shared among all the \mathbf{f}_i 's and only the coefficients c_i^k 's change. The \mathbf{k}_k 's and c_i^k 's are learned from

$$\min_{\{\mathbf{a}^k, \mathbf{b}^k, \mathbf{w}^k\}} \left\| \mathcal{F} - \sum_{k=1}^K \mathbf{a}^k \circ \mathbf{b}^k \circ \mathbf{c}^k \right\|_2^2 \quad (6)$$

where the tensor \mathcal{F} is constructed by stacking all the convolutional filters $\{\mathbf{f}_i\}_{1 \leq i \leq L}$, and the \mathbf{k}_k can be given by the components \mathbf{a}^k and \mathbf{b}^k of the canonical polyadic decomposition [48], i.e., $\mathbf{k}_k = \mathbf{a}^k \circ \mathbf{b}^k$, where \circ denotes the tensor product. The c_i^k in (5) corresponds to the i^{th} element of the vector \mathbf{c}^k .

We also aggregate statistics of the obtained features at various locations: $\text{pool}(\mathbf{x}_p)$. The function pool takes the mean activation within the neighborhood via the mean-pooling operator. This makes the feature detectors more robust to noise and the proposed model translation invariant¹. Therefore, the joint usages of random sampling strategy in (3) and the pooling operation not only avoid the overfitting problem but also give robustness to noise. The different feature results learned on a simple nuclei image are shown in Fig. 4(b), 4(c), and 4(d).

2) High-Level Map Context Features: Context features are learned from the scores, which are estimated using the outputs of the previous convolutional-pooling layer [25]. For our regression, these features are concatenated into feature vectors and added to the original ones \mathbf{x}_p to train a new regression layer. We also use the method of [44] to learn the separable kernels online to extract context features but from labels rather than scores. This will further drop the computational cost without compromising the descriptive power of the kernels. The computation of context features \mathbf{z}_p are then given by

$$\mathbf{z}_p(\mathfrak{R}^{(j-1)}) = \left\{ \left\{ \left((\mathbf{k}_k^{(j-1)} * \mathfrak{R}^{(j-1)})(\tilde{\mathbf{p}}) \right)_{k=1, \dots, K'} \right\}_{i=1, \dots, L'} \cdot \mathbf{c}_i \right\}_{i=1, \dots, L'} \quad (7)$$

where $\mathfrak{R}^{(j-1)}$ represents the updated scores in the j^{th} layers of our model (see $\mathfrak{R}^{(0)}$ in Fig. 4(e) and $\mathfrak{R}^{(1)}$ in Fig. 4(h)), and different context feature results are shown in Fig. 4(f) and 4(g), respectively. The proposed CSRCM can learn a boosted cascade of convolutional features which are invariant to irrelevant variations of samples while preserving relevant information. Learning the input data distribution with a sparse prior can enhance the model performance.

C. Regression-Based Score Map Generation

To estimate the score map, we learn a non-linear mapping

¹ <http://deeplearning.stanford.edu/wiki/index.php/Pooling>

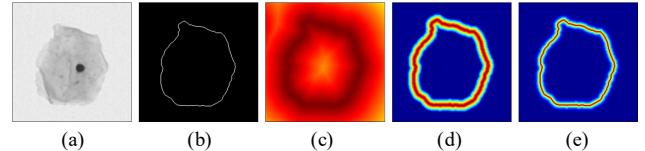


Fig. 5 The boundary labels \mathfrak{D} we want to learn for the given ground truth of an image. (a) original microscopy subimage; (b) the 1-pixel boundary ground truth; (c) the distance $-D(\mathbf{p})$; final boundary label (d) $y_p(a=1)$ and (e) $y_p(a=6)$.

from the input to the score of each pixel based on the idea of using regression in a supervised learning way. Recently, boosting techniques have been proven to be effective for object detection and segmentation, in natural and medical image analysis [49]-[51].

1) Gradient Boosting Regression: As often the case with microscopy images, many discriminative details are blurred due to the large variability and complexity of the data, so that exact locations of target boundary are uncertain. Therefore, the general classification-based approach [49], [52] cannot satisfy our need. The motivation for gradient boosting is a procedure that combines the outputs of learned base regressors to produce a powerful prediction [24]. Given 1-pixel boundary ground truth for markers or clustered objects, gradient boosting creates a $\mathfrak{R}: \mathbb{R}^D \rightarrow \mathbb{R}$ that maps learned features as inputs to the desired outputs of pixel scores by approximating the labels \mathfrak{D} :

$$\mathfrak{D}(\mathbf{p}) = e^{a(1 - \frac{D(\mathbf{p})}{r})} - 1 \quad (8)$$

if $D(\mathbf{p}) < r$, and zero otherwise, where $D(\mathbf{p})$ is the Euclidean distance transform of pixel \mathbf{p} being on the boundary and r is a threshold corresponding to \mathbf{r} in (3). The factor $a > 0$, controls the exponential decrease rate and is set to 6 in our experiment, as illustrated in Fig. 5. Suppose that we have a set of N training samples, which are denoted by $\{(\mathbf{x}_p^i, y_p^i)\}_{i=1, \dots, N}$ | $\mathbf{x}_p^i \in \mathbb{R}^D$, $y_p^i = \mathfrak{D}_i(\mathbf{p})$, $i = 1, \dots, N\}$. $\mathfrak{R}(\cdot)$ is constructed by

$$\mathfrak{R}(\mathbf{x}_p) = \sum_{m=1}^M \alpha_m h_m(\mathbf{x}_p) \quad (9)$$

where the $\alpha_m \in \mathbb{R}$ weights the contribution of each respective base regressor $h_m(\mathbf{x}_p): \mathbb{R}^D \rightarrow \mathbb{R}$. In particular, the gradient boosting tries to establish the $\mathfrak{R}(\cdot)$ that minimizes the average value of loss function on the training data by incrementally expanding the model in a greedy fashion:

$$\mathfrak{R}_m(\mathbf{x}_p) = \mathfrak{R}_{m-1}(\mathbf{x}_p) + \arg \min_{h \in \mathcal{H}} \sum_{i=1}^N L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i) + \alpha \cdot h(\mathbf{x}_p^i)) \quad (10)$$

where $L(\cdot)$ is the squared loss $L(y_p^i, \mathfrak{R}(\mathbf{x}_p^i)) = (y_p^i - \mathfrak{R}(\mathbf{x}_p^i))^2$ or ℓ_1 loss $L(y_p^i, \mathfrak{R}(\mathbf{x}_p^i)) = |y_p^i - \mathfrak{R}(\mathbf{x}_p^i)|$. The solution to the above minimization problem is to apply a steepest descent step:

$$\mathfrak{R}_m(\mathbf{x}_p) = \mathfrak{R}_{m-1}(\mathbf{x}_p) - \alpha_m \cdot \sum_{i=1}^N \nabla_{\mathfrak{R}_{m-1}} L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i)) \quad (11)$$

Then, for the squared loss, we fit a base regressor $h(\mathbf{x}_p)$ to the residual $r_i^m = -\partial L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i)) / \partial \mathfrak{R}_{m-1}(\mathbf{x}_p^i)$ at each (\mathbf{x}_p^i, y_p^i) for which the coefficient α is then calculated by line search:

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

6

Algorithm 1 Algorithm for Gradient Tree Boosting

Input: Training samples and labels $\{(\mathbf{x}_p^i, y_p^i)\}_{i=1,\dots,N}$, number of trees M , depth of tree D , and shrinkage factor $0 < \nu < 1$.

Initialization: $\mathfrak{R}_0(\cdot) = 0$

- 1: **for** $m = 1$ to M :
- 2: $w_i^m = \frac{\partial^2 L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i))}{\partial \mathfrak{R}_{m-1}^2(\mathbf{x}_p^i)}$ and $r_i^m = -\frac{1}{w_i^m} \frac{\partial L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i))}{\partial \mathfrak{R}_{m-1}(\mathbf{x}_p^i)}$
- 3: $\{\mathbf{x}_p^i\}_{i=1,\dots,Z} = \text{randSample}(\{\mathbf{x}_p^i\}_{i=1,\dots,N})$
- 4: $\mathcal{T} = \text{treeTrain}(\mathbf{x}_p^i, w_i^m, r_i^m, D), i=1,\dots,Z$
- 5: $\mathfrak{R}_m(\mathbf{x}_p^i) = \mathfrak{R}_{m-1}(\mathbf{x}_p^i) + \nu \cdot \sum_{l=1}^{|\mathcal{T}|} \gamma_l^m I(\mathbf{x}_p^i \in \mathcal{R}_l^m)$ where $(\mathcal{R}_l^m, \gamma_l^m) \in \mathcal{T}$
- 6: **end for**
- 7: **Procedure:** $\text{treeTrain}(\mathbf{x}_p^i, w_i^m, r_i^m, D)$
- 8: **if** $d < D$
- 9: $(\tau, \gamma_1^m, \gamma_2^m) = \text{splitLearning}(\mathbf{x}_p^i, w_i^m, r_i^m), i=1,\dots,Z$
- 10: $\mathcal{R}_1^m = \{\mathbf{x}_p^i | \chi(\mathbf{x}_p^i) < \tau\}, \mathcal{R}_2^m = \{\mathbf{x}_p^i | \chi(\mathbf{x}_p^i) \geq \tau\}$
- 11: $\mathcal{T}_1 = \text{treeTrain}(\{\mathbf{x}_p^i, w_i^m, r_i^m\}_{i|\chi(\mathbf{x}_p^i) < \tau}, \mathcal{R}_1^m, \gamma_1^m, d+1)$
- 12: $\mathcal{T}_2 = \text{treeTrain}(\{\mathbf{x}_p^i, w_i^m, r_i^m\}_{i|\chi(\mathbf{x}_p^i) \geq \tau}, \mathcal{R}_2^m, \gamma_2^m, d+1)$
- 13: return $\mathcal{T}_1 \cup \mathcal{T}_2$
- 14: **else**
- 15: return $\{(\mathcal{R}_l^m, \gamma_l^m)\}_{l=1,\dots,|\mathcal{T}|}$
- 16: **end if**
- 17: **Procedure:** $\text{splitLearning}(\mathbf{x}_p^i, w_i^m, r_i^m)$
- 18: Find τ and γ_1^m, γ_2^m through an exhaustive search on training samples:
- 19: $\hat{\tau} = \arg \min_{\tau, \gamma_1, \gamma_2} \varepsilon(\tau)$ in (13) where
 $\widehat{\gamma}_1^m = \text{ave}(r_i^m | \mathbf{x}_p^i \in \mathcal{R}_1^m(\tau)), \widehat{\gamma}_2^m = \text{ave}(r_i^m | \mathbf{x}_p^i \in \mathcal{R}_2^m(\tau))$
- 20: return $(\hat{\tau}, \widehat{\gamma}_1^m, \widehat{\gamma}_2^m)$

$$\alpha_m = \arg \min_{\alpha} \sum_{i=1}^N L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i) + \alpha \cdot r_i^m) \quad (12)$$

Hence, the model is updated as $\mathfrak{R}_m(\mathbf{x}_p) = \mathfrak{R}_{m-1}(\mathbf{x}_p) + \alpha_m \cdot h_m(\mathbf{x}_p)$.

2) *TreeBoost Based Algorithm:* Tree-based methods partition the feature space into L disjoint regions (leafs), and then fit a constant γ_l in each region [24]. During binary tree training, the regression trees as base regressors measure an error $\varepsilon(\tau)$ observed in each node, split the node by performing a binary test $\chi : \mathbb{R}^D \rightarrow \mathbb{R}$ with a threshold τ , and return values γ_1 and γ_2 . We seek the optimal split point τ , γ_1 , and γ_2 that minimize the error after split as follows

$$\varepsilon(\tau) = \sum_{i|\chi(\mathbf{x}_p^i) < \tau} w_i^m (r_i^m - \gamma_1^m)^2 + \sum_{i|\chi(\mathbf{x}_p^i) \geq \tau} w_i^m (r_i^m - \gamma_2^m)^2 \quad (13)$$

where w_i^m and r_i^m denote the weight and response of \mathbf{x}_p^i at the m^{th} iteration here, respectively. Throughout this paper, we use: $w_i^m = 2$ and $r_i^m = y_p^i - \mathfrak{R}_{m-1}(\mathbf{x}_p^i)$ for the squared loss, and $w_i^m = 1$ and $r_i^m = \text{sign}[y_p^i - \mathfrak{R}_{m-1}(\mathbf{x}_p^i)]$ for the ℓ_1 loss. Suppose that we have a partition into two regions $\mathcal{R}_1^m(\tau)$ and $\mathcal{R}_2^m(\tau)$, finding the optimal γ_1^m and γ_2^m in each region can be formulated as

$$\widehat{\gamma}_k^m = \arg \min_{\gamma_k^m} \sum_{\mathbf{x}_p^i \in \mathcal{R}_k^m(\tau)} L(y_p^i, \mathfrak{R}_{m-1}(\mathbf{x}_p^i) + \gamma_k^m) \quad (14)$$

where k indexes the region. It is simply the regression tree that best predicts the current residuals $r_i^m = y_p^i - \mathfrak{R}_{m-1}(\mathbf{x}_p^i)$, and $\widehat{\gamma}_k^m$ is weighted average of the residuals in each corresponding region. For any choice τ , the error minimization in (13) is solved by

$$\widehat{\gamma}_1^m = \text{ave}(r_i^m | \mathbf{x}_p^i \in \mathcal{R}_1^m(\tau)), \quad \widehat{\gamma}_2^m = \text{ave}(r_i^m | \mathbf{x}_p^i \in \mathcal{R}_2^m(\tau)) \quad (15)$$

Hence, the determination of the best τ is feasible by scanning through all of the inputs. The tree continues recursive growing as more splits are made and stops at a specified depth. Finally, the γ_l^m stored in the leaf node \mathcal{R}_l^m , where $l = 1, \dots, L$, is used as a score of the boundary. In our implementation, to inject the randomness for improved generalization, we also experimented with the stochastic regularization technique [24] and the results proved to be efficient. As a result, a sampled subset of all possible y_p and features \mathbf{x}_p is randomly selected at each splitting node.

After the tree training procedure, which is summarized in Algorithm 1, we update the model by integrating a shrinkage regularization to upgrade its generalization ability [24]:

$$\mathfrak{R}_m(\mathbf{x}_p^i) = \mathfrak{R}_{m-1}(\mathbf{x}_p^i) + \nu \cdot \sum_{l=1}^L \gamma_l^m I(\mathbf{x}_p^i \in \mathcal{R}_l^m) \quad (16)$$

where $\nu \in [0, 1]$ controls the learning rate of boosting procedure. $I(\cdot)$ is the indicator function. In our experiments, we set $\nu = 0.1$.

During our TreeBoost testing procedure, each new image feature is independently pushed through each trained regression tree by using the learned split parameter. When the feature arrives at a leaf node, the optimal γ_l^m is used to determine the score value of this testing sample at current tree. Fig. 3(a3) and 3(b3) show the generated boundary score map for markers and clustered cells, respectively, by applying a sequence of learned regression trees $h_m(\cdot)$ and corresponding weights α_m .

IV. COMPLETE CONTOUR INFERENCE

Suppose that we have n_1 markers from the \mathfrak{R}_m^c , denoted by $\mathbf{s} = \{\mathbf{s}_1, \dots, \mathbf{s}_{n_1}\}$, and n_2 boundary point coordinates from the corresponding \mathfrak{R}_c^j , denoted by $\mathbf{b} = \{\mathbf{b}_1, \dots, \mathbf{b}_{n_2}\}$, where $\mathbf{b}_i \in \mathbb{R}^2$, $\mathbf{s}_j \in \mathbb{R}^2$, and $n_2 \gg n_1$. We first employ the Euclidean distance to compute the association between \mathbf{b} and a certain \mathbf{s}_j , which aims to look for the boundary pixels close to the relevant marker. As discussed in [15], using only the distance measure $dis(\mathbf{b}, \mathbf{s}_j)$ to compute the boundary evidences may cause the miss-classification of boundary, because points on the boundary of i^{th} object might be assigned to its nearest neighbor, instead of i^{th} object. Therefore, we further incorporate an angle measure $ang(\mathbf{s}_j, \mathbf{b})$ to overcome this problem. More precisely, in a typical histopathology image, the regions occupied by nuclei have lower image intensities than the background. For this reason, if \mathbf{b}_i is a substance of j^{th} object's boundary, the gradient at \mathbf{b}_i diverges from \mathbf{s}_j and vice versa. Hence, the improved boundary classification is achieved by minimizing an angle between the direction of intensity gradient or negative intensity

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

7

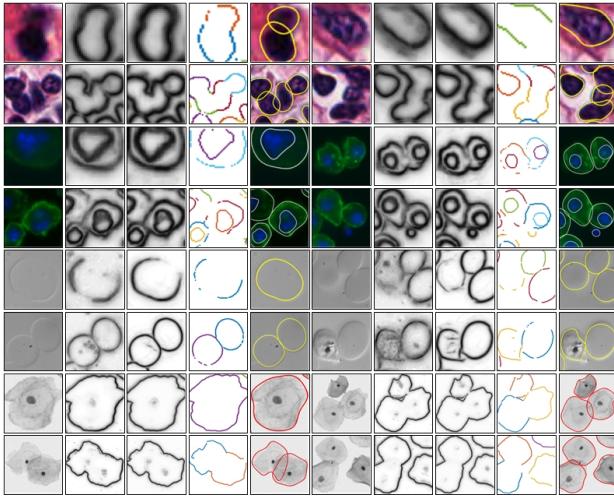


Fig. 6. Shape initialization and inference. The first and sixth columns display the selected test images from four disparate datasets. Other columns from left to right show the score maps obtained using our single-layer model and CSRCM, respectively, boundary evidences for each object based on cluster-boundary-to-marker association, and final inferred contours for individuals based on the MDL-constrained periodic B-spline model.

gradient at \mathbf{b}_i and the direction of line from \mathbf{s}_j to \mathbf{b}_i . As described in [15], the final cluster-boundary-to-marker association $\mathcal{A}(\mathbf{s}_j, \mathbf{b})$ is able to produce a set of reliable boundary evidences for individual objects (shape initialization) by weighting the above two measures:

$$\mathcal{A}(\mathbf{s}_j, \mathbf{b}) = \eta \cdot \text{dis}(\mathbf{b}, \mathbf{s}_j) + (1 - \eta) \cdot \text{ang}(\mathbf{s}_j, \mathbf{b}) \quad (17)$$

The 4th and 9th columns of Fig. 6 show the evidence results using the method above for the case when $\eta = 1/2$.

A. Contour Representation via Periodic B-Spline Model

Suppose that we have a set of n_i edge points as the boundary evidences for the i^{th} object in an image, denoted as $\{\mathbf{e}_{i,1}, \mathbf{e}_{i,2}, \dots, \mathbf{e}_{i,n}\}$, where $\mathbf{e}_{i,j}$ is a 2×1 vector; set $\mathbf{X}_i = \{\mathbf{e}_{i,j}, j=0, 1, \dots, n_i\}$. Since we know the partial possible boundary of each object, it is possible to mathematically represent its complete contour by using the periodic B-spline model [26]:

$$\mathbf{s}_i^{\Theta}(t) = \sum_{k=0}^{l-1} \beta_{k,d}(t) \mathbf{c}_{i,k} = \mathbf{B}_i \mathbf{c}_i \quad (18)$$

where \mathbf{B}_i is a $2n_i \times 2l$ matrix with $\beta_{k,d}(t_{i,0} + j(t_{i,l-1} - t_{i,0})/n_i) \mathbf{I}_2$ as its $(j, k)^{\text{th}}$ submatrix, \mathbf{c}_i is a $2l \times 1$ column vector containing the coordinates of control points, t is a parameter to identify a point on the curve, and Θ represents all the remaining parameters of our periodic B-spline model.

B. Shape Inference by Minimum Description Length Principle

We assume here that $t_{i,j}$ is known. Consider a set of pairs $\{(t_{i,j}, \mathbf{s}_i^{\Theta}(t_{i,j})), j = 0, 1, \dots, L-1\}$, and consider the problem of finding a periodic spline (of a certain degree, say $d=3$ in here, on a given periodic set of knots $\{t_{i,0} < t_{i,1} < \dots < t_{i,l}\}$, with $l \ll L$) that best fits this set. Since the knots are known, a set of l periodic B-splines $\{\beta_{k,3}(t), k = 0, \dots, l-1\}$ can be defined, so the problem is that of finding the optimal parameters Θ of the

model. Suppose that $\mathbf{e}_{i,j}$ is a noisy observation of $\mathbf{s}_i^{\Theta}(t)$ at a B-spline parameter value $t_{i,j}$, i.e., $\mathbf{e}_{i,j} = \mathbf{s}_i^{\Theta}(t_{i,j}) + \varepsilon_{i,j}$, where $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{2n_i})$. Hence, the Θ we want to optimize is represented as follows: $\Theta = \{\mathbf{c}_i^{(l)}, \sigma^2, l\}$.

The first key observation behind minimum description length (MDL) criteria is that looking for maximum likelihood estimates of $\mathbf{c}_i^{(l)}$ and σ^2 given l are equivalent to looking for the Shannon code for which the observations \mathbf{X}_i have the shortest code length [53]: $L(\mathbf{X}_i | \mathbf{c}_i^{(l)}, \sigma^2) = -\log p(\mathbf{X}_i | \mathbf{c}_i^{(l)}, \sigma^2)$; The second crucial aspect of MDL criteria is that the parameters themselves are also part of the code, denoted by $L(\mathbf{c}_i^{(l)}, \sigma^2, l)$. Therefore, the joint MDL estimate of $\mathbf{c}_i^{(l)}$, σ^2 , and l is

$$(\hat{l}, \widehat{\mathbf{c}}_i^{(l)}, \widehat{\sigma^2}) = \arg \min_l \left\{ L(\mathbf{c}_i^{(l)}, \sigma^2, l) - \max_{\mathbf{c}_i^{(l)}, \widehat{\sigma^2}} \left\{ \log p(\mathbf{X}_i | \mathbf{c}_i^{(l)}, \sigma^2) \right\} \right\} \quad (19)$$

where the data description length or, equivalently, the negative log-likelihood of $\mathbf{s}_i^{\Theta}(t)$ given \mathbf{X}_i is

$$-\log p(\mathbf{X}_i | \mathbf{c}_i^{(l)}, \sigma^2) = \frac{n_i}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \|\mathbf{X}_i - \mathbf{B}_i^{(l)} \mathbf{c}_i^{(l)}\|_2^2 \quad (20)$$

and the parameter description length $L(\mathbf{c}_i^{(l)}, \sigma^2, l)$ is given by

$$L(\mathbf{c}_i^{(l)}, \sigma^2, l) = \left(\log_2(h_i \cdot w_i) - \log_2 \left(\frac{\zeta}{\|\mathbf{B}_i^{(l)}\|_{\infty}} \right) \right) \cdot l \quad (21)$$

$$= (\log_2(h_i \cdot w_i)) \cdot l$$

where h_i and w_i denote the size of local window in where the i^{th} object located. Since each knot can be represented by an integer between 0 and $h_i \times w_i - 1$, $\log_2(h_i \cdot w_i)$ bits are necessary to encode it. On the other hand, each control point can be represented by $\log_2(\zeta / \|\mathbf{B}_i^{(l)}\|_{\infty})$ bits with the precision $\zeta / \|\mathbf{B}_i^{(l)}\|_{\infty}$ [27], [28], [53].

For a given l , the usual least squares criterion leads to

$$\widehat{\mathbf{c}}_i^{(l)} = \left((\mathbf{B}_i^{(l)})^T \mathbf{B}_i^{(l)} \right)^{-1} (\mathbf{B}_i^{(l)})^T \mathbf{X}_i \quad (22)$$

and the optimal residual error variance σ^2 is estimated by

$$\widehat{\sigma^2} = \frac{\|\mathbf{X}_i - \mathbf{B}_i^{(l)} \left((\mathbf{B}_i^{(l)})^T \mathbf{B}_i^{(l)} \right)^{-1} (\mathbf{B}_i^{(l)})^T \mathbf{X}_i\|_2^2}{n_i} \quad (23)$$

A key issue arising in parametric contour descriptions is the choice of parameterization order, e.g., the number of the B-spline control point, namely l . Since the parameterization itself guarantees regularity of the represented shape, we formulate adaptive contour location as a parameter estimation problem without regard for the association with a priori. Substituting (20) and (21) into (19), we can rewrite (19) as follows

$$\hat{l} = \arg \min_l \left\{ (\log_2(h_i \cdot w_i)) \cdot l + n_i \log_2 \left(\sqrt{\widehat{\sigma^2}(l)} \right) \right\} \quad (24)$$

where $\widehat{\sigma^2}(l)$ is a function of l and the minimum of l is typically found through exhaustive search [24].

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

8

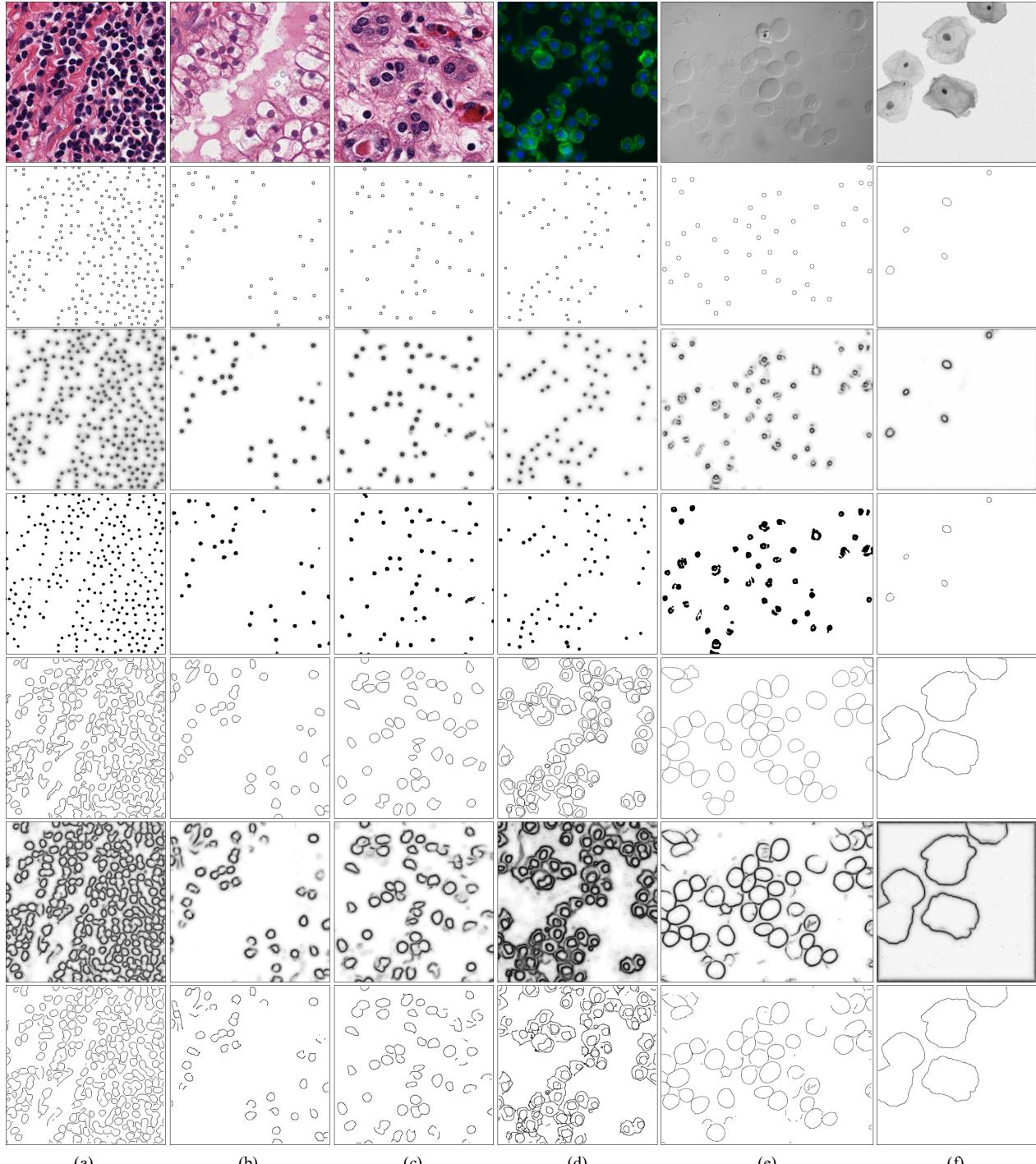


Fig. 7. Detection and segmentation results obtained from our CSPL method using the proposed CSRCM. (a)-(c) KIRC histopathology images, (d) drosophila Kc167 cellular image, (e) DIC human red blood cell image, and (f) cervical cytology specimen. In each case, we present from top to bottom the original images, the marker boundary ground truths, the final marker boundary score maps, the corresponding markers after thresholding, the boundary ground truths of clustered cells/nuclei, the final boundary score maps of clustered cells/nuclei, and the corresponding boundaries after non-maximum suppression. Note that we conduct the thresholding instead of non-maximum suppression to obtain the connect regions to represent the markers.

After calculating the optimal model parameters $\widehat{\Theta}$, we can obtain a complete inferred contour for each object shown in the 5th and 10th columns of Fig. 6. Given initial contours, the proposed framework performs complete contour representation with the periodic B-spline model and shape inference with the MDL criteria. As seen, the MDL-constrained periodic B-spline

curves snap to the latent boundaries so as to not only infer the desired objects' shapes, but estimate the satisfactory occlusion degrees among them. In the cases where the boundary evidences are not sufficient, the MDL criteria force the estimated contours to deviate from the correct objects' boundaries as less as possible.

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

9

TABLE I

OPTIMAL PARAMETER VALUES OF OUR CSRCM ON FOUR MICROSCOPY DATASETS. POSITIVE AND NEGATIVE LOCATIONS ARE THE LOCATIONS CLOSE TO THE BOUNDARIES AND FAR FROM THE BOUNDARIES, RESPECTIVELY

Dataset	Pos.+Neg.	Training No.	Tree No.	Tree Depth	Layer No.
KIRC	10^3+10^5	50000	25	4	3
Kc167	10^5+10^5	100000	100	7	3
RBC	10^5+10^5	100000	75	5	8
Pap Smear	$5 \times 10^4 + 5 \times 10^4$	25000	50	5	3

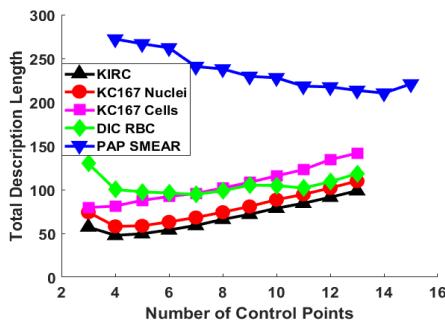


Fig. 8. Plots of the total description lengths with minimum at 4, 4, 3, 7, and 14 control points for KIRC nuclei (black curve), drosophila Kc167 nuclei (red curve), drosophila Kc167 cells (magenta curve), DIC red blood cells (green curve), and cervical cells (blue curve), respectively.

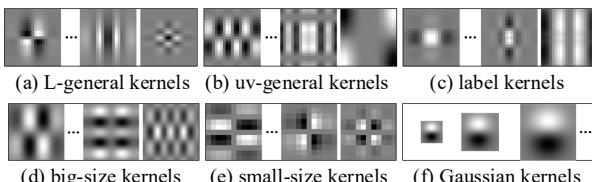


Fig. 9. Selections of the different classes of rank-1 kernels obtained by training the low rank tensor decomposition model on the datasets. The L-, uv-, and gts- denote L channel, u, v channels, and ground truth, respectively.

V.EXPERIMENTS AND RESULTS

In this section, we describe how we adapt our proposed framework for detection and segmentation purposes and the datasets used to evaluate the performance. To test our method, we use four challenging microscopy datasets containing quite disparate overlapping granular objects and discuss the optimal parameter settings of the proposed CSRCM. We also discuss the adaptive parameter selection of our fitting model for each of the datasets. We then evaluate the detection and segmentation performance and compare our results with other state of the art approaches. Finally, to further validate our regression method, we perform sensitivity analyses of the proposed CSRCM key parameters and also experiment with classification training and testing.

A. Datasets and Parameter Settings

We tested our method on four microscopy datasets depicted by Fig. 7. The clinical specialists manually delineated the cells/nuclei in all these microscopy images, which we used as the ground truth. The empirically optimal values for our CSRCM key parameters (i.e., the number of training samples, the number of trees, the depth of trees, and the number of layers or model iterations) were determined through the leave-one-out

cross validation [54] on small volumes of the datasets.

1) *Kidney Renal Cell Carcinoma Histopathology Dataset*: The micrographs of kidney renal cell carcinoma (KIRC) from the cancer genome atlas data portal (becklab.hms.harvard.edu/software) include different whole slide images, representing a range of histologic grades of KIRC. In our experiment, we utilized 18 images (400×400 pixels) for training and 19 images for testing with an average number of 96 nuclei per image.

2) *Drosophila Kc167 Cellular Dataset*: The drosophila Kc167 cellular dataset (<https://data.broadinstitute.org/bbbc/BBBC007/>) includes 14 micrographs (400×400 pixels) with an average of 98 cells per image. Each cell is stained for DNA to label nuclei and actin to show the cell body. The dataset is divided evenly for training and testing, and the first 7 images are used for training.

3) *Human Red Blood Cell Dataset*: This image set consists of five differential interference contrast (DIC) images of red blood cells (<https://data.broadinstitute.org/bbbc/BBBC009/>). To test the effectiveness of proposed method, a total of 5 cell images (800×600 pixels) are used, comprising normal and abnormal cell shapes and varying levels of contrast enhancement.

4) *Cervical Cytology Images form Pap Smear*: A sample of cells collected from the cervix are smeared on a glass slide and examined under a microscope [22]. The dataset we use contains 18 cytology subimages (512×512 pixels) with different degrees of overlap. In our experiment, 7 subimages are used for training and 11 subimages for testing

Table I lists the optimal parameter values of the proposed CSRCM for four microscopy datasets. During tree training, a quarter of all the locations were randomly utilized to learn the trees with 100 randomly selected features at central pixel and 500 context features per channel. More locations were needed to learn the weight α_m in (9) after computing the response of tree, and we therefore used all the locations for this purpose.

Fig. 8 plots the total description length of proposed MDL-based periodic B-spline model consisting of (20) and (21) for each dataset (with different object sizes and shapes) when the number of control points varies. The final contour estimations in the 5th and 10th columns of Fig. 6 correspond to the minimum control points of the plot in Fig. 8, representing an optimal tradeoff between the model's conciseness and its fitting quality.

Most key parameters in our feature learning were trained from (4) and (6) containing the number K of rank-1 kernels in (5), the size s of kernel, and the corresponding weights w to reconstruct the f_i 's in (5). The conventional kernels often suffer from an undesired merging of adjacent objects because of their intrinsic usage of an inappropriately broad kernel. As a result, neighboring structures cannot be properly resolved. Therefore, we added classes of Gaussian and size rank-1 kernels to our kernel banks. We assume that the parameter settings of convolutional-pooling layers are applicable to all the datasets. Therefore, the experimental settings are described as follows

- **General Kernels** with $K=36$, $s=21$, and w a 36×121 matrix;
- **Size Kernels** with $K=36$, $s=43$, and w a 36×89 matrix for big size; $K=25$, $s=11$, and w a 25×49 matrix for small size;
- **Gaussian Kernels** with $K=8$, $s=17$, and w a 8×4 matrix for small scale; $K=8$, $s=29$, and w a 8×4 matrix for middle

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

10

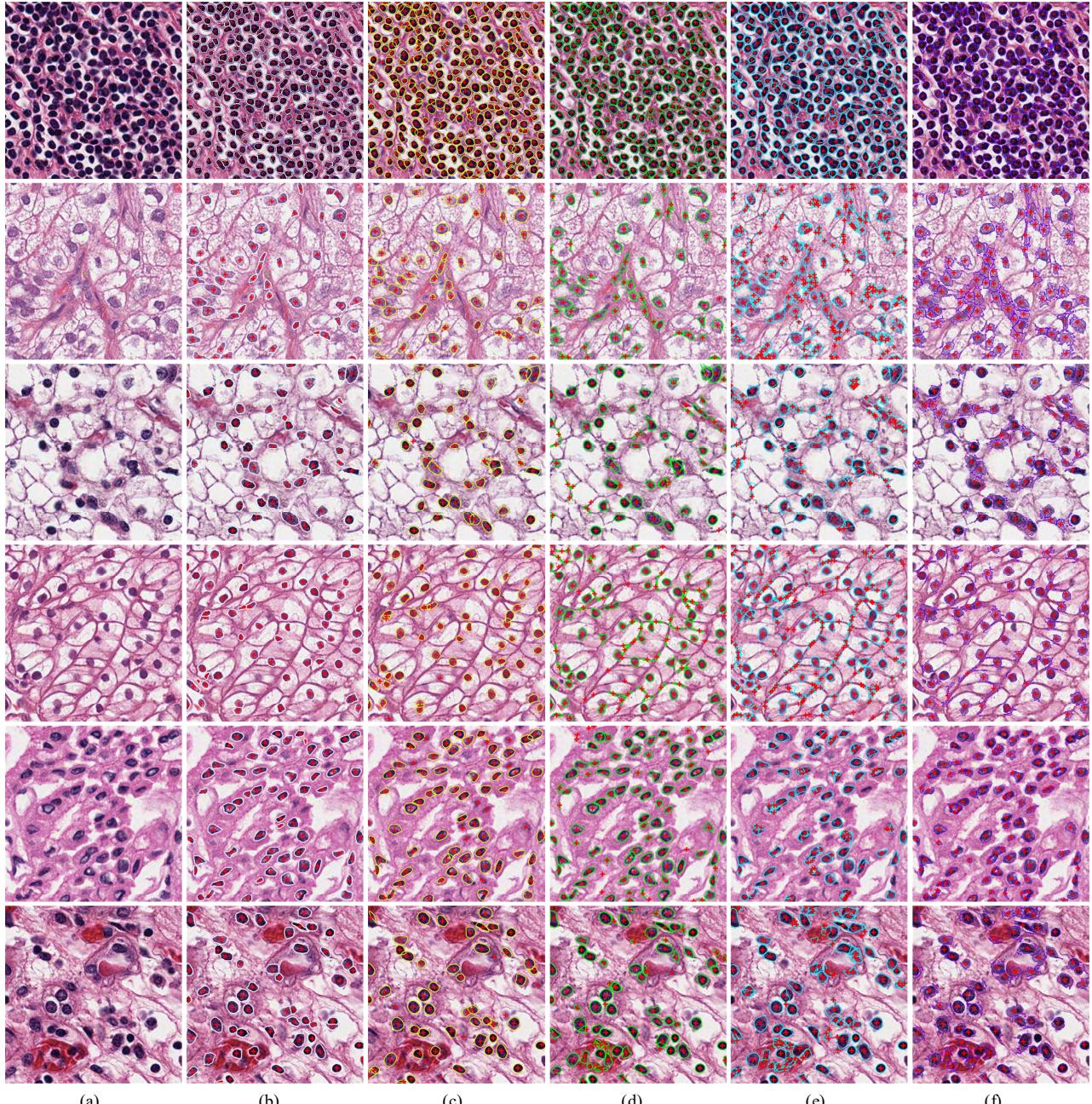


Fig. 10. Comparative detection and segmentation using different methods on the challenging KIRC histopathology dataset. (a) original images, (b) image outlined ground truth boundaries, (c) outlined by our method, (d) outlined by Song *et al.*'s approach [15], (e) outlined by Park *et al.*'s approach [29], and (f) outlined by Al-Kofahi *et al.*'s approach [16]. Note that Al-Kofahi *et al.* proposed a detection method by combining the LoG filter with adaptive scale selection. Park *et al.* used the morphological erosion to produce the markers. Song *et al.* combined Al-Kofahi *et al.*'s and Park *et al.*'s methods together to achieve final nucleus detection.

scale; $K=8$, $s=43$, and \mathbf{w} a 8×4 matrix for big scale. After obtaining the score maps, we also trained a small different set of rank-1 kernels for the subsequent model layers. The only difference is that this is done using instead the labels with parameters $K=36$, $s=21$ and \mathbf{w} a 36×160 matrix. Fig. 9 exemplifies our learned rank-1 kernel banks. In our implementation, a truncated 36×64 weight \mathbf{w} of L-general kernels was used to compute the features so as to inject the randomness for improved generalization.

B. Comparison with the Other State of the Art Methods

We compare our proposed method against seven of the most representative state of the art approaches for various cell/nucleus segmentation. They are Gaussian mixture model-based B-splines [29], optimized split point pairs-driven B-splines [15], nucleus segmentation using graph cut and graph coloring [16], nuclei and cytoplasm joint segmentation using the sliding band filter [18], joint level set representation with geodesic-based shape prior [22], a variational method with star-shape prior [23] and a convolutional networks-based segmentation model [8].

1) Evaluation Metrics: To better quantitatively evaluate our proposed method, we add four commonly used metrics:

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

11

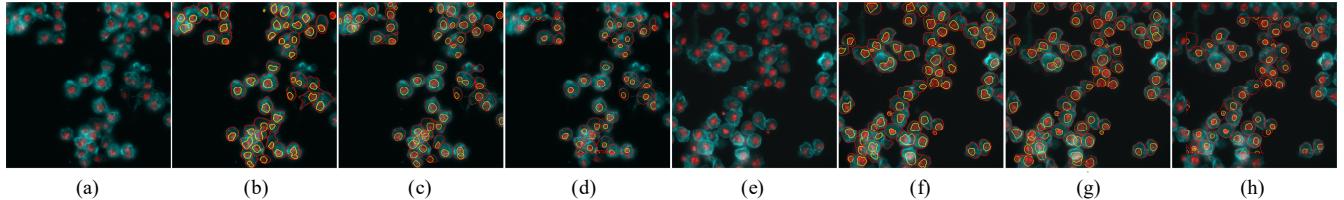


Fig. 11. Comparative segmentation using different methods on the drosophila Kc167 cellular dataset. (a), (e) original images, (b), (f) image outlined by ground truth boundaries, (c), (g) outlined by our method, and (d), (h) outlined by Quelhas *et al.*'s approach [18].

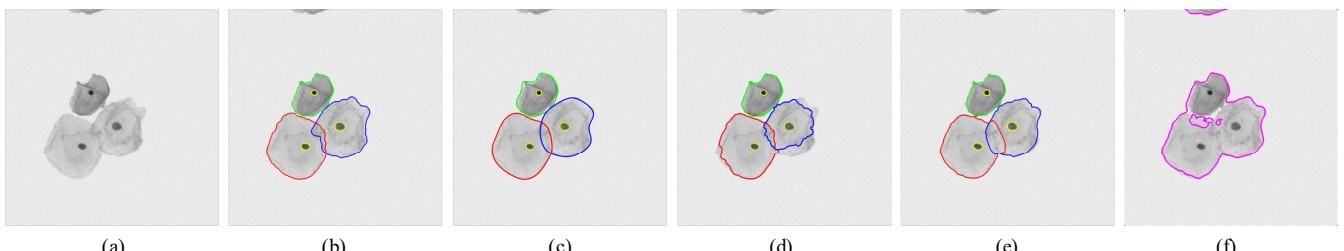
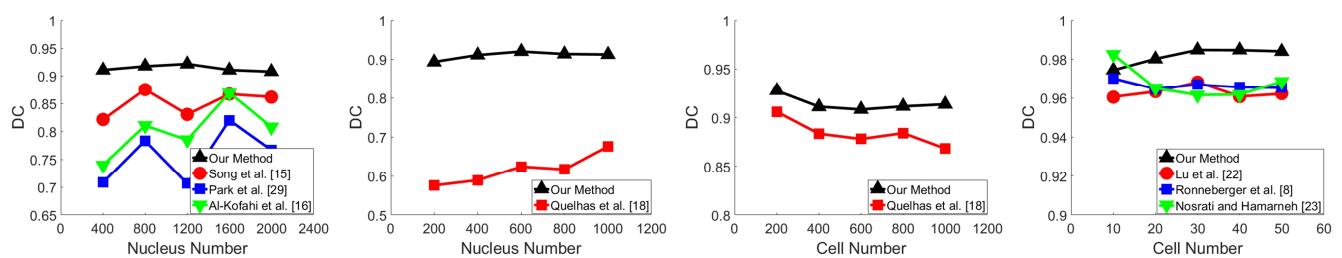


Fig. 12. Comparative segmentation using different methods on the cervical cytology dataset. (a) original image (b) image outlined by ground truth boundaries, (c) outlined by our method, (d) outlined by Lu *et al.*'s approach [22], (e) outlined by the approach of Nosrati and Hamarneh [23], and (f) outlined by Ronneberger *et al.*'s approach [8].



(a) KIRC histopathology dataset (b) Kc167 cellular dataset (DNA channel) (c) Kc167 cellular dataset (Actin channel) (d) Cervical cytology dataset
Fig. 13. Comparative pixel-wise area-based segmentation accuracy on the KIRC histopathology dataset, drosophila Kc167 cellular dataset, and cervical cytology dataset. The proposed method outperforms the other state-of-the-art approaches on each dataset and is robust to the number of cells/nuclei and type of dataset.

TABLE II
EVALUATION OF NUCLEUS DETECTION ON THE KIRC HISTOPATHOLOGY
DATASET IN TERMS OF MEAN DC, MEAN TPR, MEAN PPV, MEAN FPR, AND
MEAN ABD

Algorithm	DC	TPR	PPV	ABD
Al-Kofahi <i>et al.</i> [16]	0.6538	0.8858	0.5242	90.47
Park <i>et al.</i> [29]	0.6459	0.8972	0.5229	100.42
Song <i>et al.</i> [15]	0.7512	0.9042	0.6489	76.22
proposed	0.8422	0.9111	0.7871	36.52

TABLE III
COMPARATIVE PIXEL-WISE SEGMENTATION ACCURACY ON THE KIRC
HISTOPATHOLOGY DATASET

Algorithm	DC	TPR	PPV	FPR	ABD
Al-Kofahi <i>et al.</i> [16]	0.7216	0.9028	0.6079	0.0992	64.01
Park <i>et al.</i> [29]	0.6933	0.8408	0.6091	0.0961	78.02
Song <i>et al.</i> [15]	0.7964	0.8562	0.7542	0.0460	43.75
proposed	0.8998	0.9110	0.8895	0.0187	21.45

- Sensitivity: $TPR = \frac{TP}{TP+FN}$; $FPR = \frac{FP}{TN+FP}$; $PPV = \frac{TP}{TP+FP}$.
- Average boundary distance (ABD) is the average distance between the boundary of automatic segmented object (SEG) and that of the manual ground truth (GT)

$$ABD = \frac{1}{2} \left(\frac{\sum_{x \in SEG} d(x, GT)}{|SEG|} + \frac{\sum_{u \in GT} d(u, SEG)}{|GT|} \right) \quad (25)$$

where $d(x, GT)$ is the minimum distance of point x on the automatic segmented object boundary SEG to the points on the GT boundary, $d(u, SEG)$ is the minimum distance of point u on the GT boundary to the points on the automatic segmented object boundary SEG. The unit of ABD is pixel.

2) *Nucleus Detection*: Marker boundary score map generated from our proposed CSRCM provides cell/nucleus positions for the subsequent periodic B-spline fitting model. We conduct both qualitative and quantitative analyses to evaluate the detection. The ground truth of each cell/nucleus center is calculated as the mean of the annotated contour, and the detected centers are the centroids of the thresholding result of score map. Fig. 10 presents the comparative detection results between the proposed CSRCM and LoG filter with adaptive scale selection [16], morphological erosion [29], and combination of them [15] on the challenging KIRC histopathology dataset. As shown, the proposed method produces much better performance than the others, especially on the images with heavy background clutter.

We apply the method in [55] for quantitative analysis, which defines the ground truth regions as circular regions within 15

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

12

TABLE IV
QUANTITATIVE COMPARISON OF DIFFERENT KEY PARAMETERS OF PROPOSED MODEL ON THE KIRC HISTOPATHOLOGY DATASET

Train No.	DC	TPR	PPV	ABD	Tree No.	DC	TPR	PPV	ABD	Tree Depth	DC	TPR	PPV	ABD	Model Iter.	DC	TPR	PPV	ABD
6250	0.701	0.849	0.597	6.93	5	0.695	0.828	0.599	13.17	1	0.729	0.837	0.647	11.15	0	0.726	0.851	0.634	10.23
12500	0.722	0.847	0.629	7.38	15	0.735	0.848	0.649	6.22	2	0.742	0.840	0.665	8.28	1	0.746	0.852	0.665	5.22
25000	0.720	0.847	0.627	7.06	25	0.746	0.852	0.665	5.22	3	0.733	0.844	0.649	6.90	2	0.787	0.825	0.754	3.81
50000	0.746	0.852	0.665	5.22	50	0.745	0.850	0.664	7.06	4	0.746	0.852	0.665	5.22	3	0.786	0.827	0.750	5.96
100000	0.728	0.845	0.640	5.46	100	0.738	0.854	0.651	8.04	5	0.725	0.852	0.632	7.90	5	0.755	0.821	0.698	10.58

TABLE V
QUANTITATIVE COMPARISON OF DIFFERENT KEY PARAMETERS OF PROPOSED MODEL ON THE DROSOPHILA KC167 CELLULAR DATASET

Train No.	DC	TPR	PPV	ABD	Tree No.	DC	TPR	PPV	ABD	Tree Depth	DC	TPR	PPV	ABD	Model Iter.	DC	TPR	PPV	ABD
6250	0.602	0.762	0.498	18.01	5	0.584	0.755	0.480	18.95	1	0.581	0.753	0.475	17.95	0	0.599	0.764	0.494	17.90
12500	0.611	0.763	0.510	18.00	15	0.608	0.762	0.507	17.62	2	0.603	0.752	0.504	18.15	1	0.620	0.764	0.522	17.54
25000	0.614	0.764	0.513	17.98	25	0.620	0.764	0.522	17.54	3	0.612	0.756	0.515	17.70	2	0.663	0.738	0.602	16.26
50000	0.620	0.764	0.522	17.54	50	0.628	0.769	0.532	17.81	5	0.620	0.766	0.521	17.50	3	0.659	0.736	0.597	16.64
100000	0.620	0.763	0.523	17.58	100	0.632	0.770	0.537	17.80	7	0.623	0.768	0.525	17.35	5	0.654	0.736	0.589	16.55

TABLE VI

QUANTITATIVE COMPARISON OF DIFFERENT KEY PARAMETERS OF PROPOSED MODEL ON THE DIC RED BLOOD CELL DATASET. NOTE THAT THE LARGE VALUES OF ABD ARE MAINLY CAUSED BY THE INCORRECT VALUES ON THE BACKGROUND, WHICH DO NOT EXIST IN THE GROUND TRUTH

Train No.	DC	TPR	PPV	ABD	Tree No.	DC	TPR	PPV	ABD	Tree Depth	DC	TPR	PPV	ABD	Model Iter.	DC	TPR	PPV	ABD
6250	0.526	0.533	0.525	133.92	5	0.520	0.509	0.533	127.52	1	0.463	0.524	0.420	206.42	0	0.495	0.555	0.454	271.22
12500	0.551	0.546	0.556	128.94	15	0.541	0.527	0.558	118.05	2	0.511	0.519	0.506	163.09	1	0.559	0.541	0.579	109.82
25000	0.545	0.523	0.571	113.36	25	0.559	0.541	0.579	109.82	3	0.538	0.520	0.558	126.14	3	0.593	0.587	0.601	98.31
50000	0.559	0.541	0.579	109.82	50	0.579	0.559	0.602	100.64	4	0.559	0.541	0.579	109.82	5	0.601	0.606	0.597	97.91
100000	0.561	0.540	0.586	103.44	75	0.586	0.573	0.600	117.11	5	0.576	0.553	0.601	91.71	7	0.605	0.619	0.592	107.20

pixels for every nucleus center. A detected nucleus centroid is considered to be a true positive (TP) only if it lies within the ground truth region; otherwise, it is considered as a false positive (FP). Each TP is matched with the nearest ground truth nucleus center. The ground truth nucleus centers that are not matched by any detected results are considered to be false negatives (FN). We calculate pixel-wise Dice Coefficient (DC) measures [56]: $DC = \frac{2TP}{2TP+FP+FN}$ with a specific threshold,

where the minimal DC threshold is above 0.6. Table II lists the evaluation of nucleus detection for different methods. It is clear that our proposed method produces the best results. Specifically, [15] might need more sophisticated binarization design to provide desired results. [16] and [29] produce fair performance on some images but much lower detection accuracy on the whole KIRC histopathology dataset. The proposed approach outperforms the traditional image analysis methods due to the learned feature representations.

3) *Nucleus Segmentation*: Table III, Fig. 10, Fig.11, and Fig. 12 give the comparative segmentation results on three representative microscopy datasets. The contours created by the method of [29] do not well describe nuclear boundaries and it is prone to undersegmentation due to insufficient boundary information. Song *et al.* [15] are unable to maximally segment nuclear regions although they can handle weak boundaries on nucleus overlapping regions. The segmentation of Al-Kofahi *et al.*'s algorithm [16] also struggles with the case where the nuclei are aggregate because their work lies in the control of the filter scale. In addition, the performance of both [16] and [29] tend to suffer in severe case of heavy background clutter.

Overlap correction and shape regularization described by Quelhas *et al.* [18] still does not estimate the latent shape and preserve topology, and therefore miss some small object information because of intensity inhomogeneity and dense clusters. For cervical cell segmentation, although Lu *et al.* [22] can accurately locate every cell with parameter tuning, the computed boundary of each cell is not natural. For the clumps with different sizes of overlapped cells, they always fail to find a balance of iterations of the level set function evolution. The method in [23] produces a visual segmentation result that is precise in the overlapping regions, but less precise in the contours splitting the cells from the background, where inaccurate random forest probability map is the main reason for this issue. The approach proposed by [8] shows its superiority on segmenting natural clustered cell regions, but produce some small holes inside/outside individual cells. On the contrary, ours can effectively handle various cell/nucleus overlaps due to the learned locations and shape priors of clustered objects and constraints of topology preserving, and therefore provides more accurate segmentation results (see Fig. 10, Fig. 11, and Fig. 12). The comparative experiments on the more challenging histopathology dataset are provided in the supplemental file.

To quantitatively evaluate the performance of the proposed method, binary masks for each segmentation result and ground truth are generated. For each dataset, we randomly sampled and added a certain number of cells/nuclei from the ground truth, generated the binary cells/nuclei using different segmentation methods. Finally, we assess segmentation accuracy by using pixel-wise area-based approach. In this evaluation, the aim is to assess accuracy in terms of the areas of correctly segmented

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) < 13

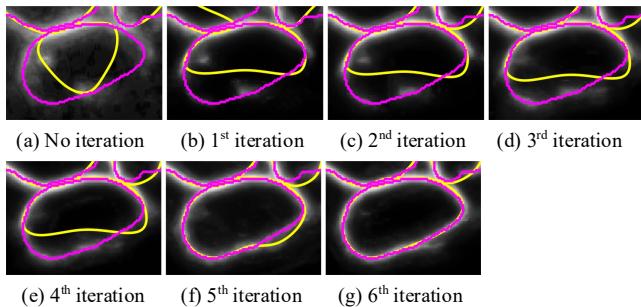


Fig. 14. Boundary score maps of the selected DIC red blood cells generated by the proposed cascade regression chain. The magenta curves indicate the manually-delineated cell boundaries and the yellow ones are the final complete contour inferences on the corresponding boundary score maps. From (a) to (g), with iterative refinements, a more and more accurate boundary score map can be obtained, guiding the fitting curves closer to the target latent boundaries.

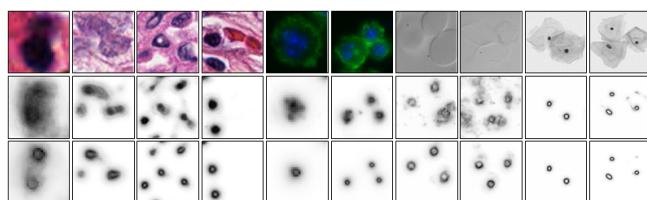


Fig. 15. Marker detections by using our proposed cascade regression chain on four different microscopy datasets. From top to bottom: original images, initial marker boundary score maps obtained from a single-layer regression model, and final score maps obtained from the proposed cascade regression chain.

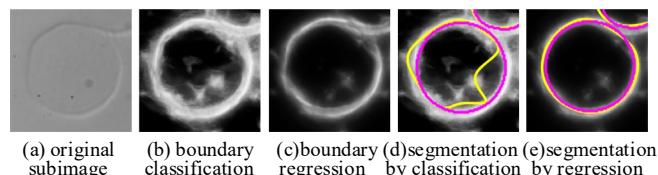


Fig. 16. Comparison of boundary classification results and regression results for the selected DIC red blood cells overlaid with manually-delineated contours shown as magenta curves and final inferred contours shown as yellow curves.

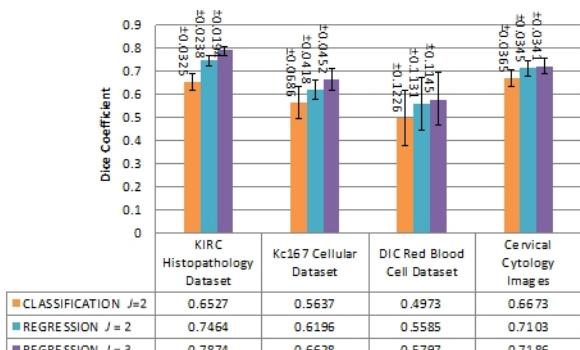


Fig. 17. Average DC values of the proposed method with respect to 2-layers ($J=2$) boundary classification, 2-layers boundary regression, and 3-layers ($J=3$) boundary regression.

cells/nuclei. We use one-to-one matches and consider the overlapping pixels of computed-annotated object pairs of these matches as true positives. Fig. 13 summarizes the performance of different methods. It can be clearly seen that, our results have a good true positive rate, false positive rate, and false negative rate, indicated by the DC value of 0.914, with an average 6.3%

improvement over the DC obtained by Song *et al.* at 0.851, 11.2% improvement over the DC obtained by Al-Kofahi *et al.* at 0.802, and 15.7% improvement over the DC obtained by Park *et al.* at 0.757 for the KIRC histopathology dataset. Our method also has a high average DC value of 0.91 compared to Quelhas *et al.*'s method with a low DC value of 0.62 and 0.88 for Kc167 nuclei and Kc167 cytoplasm, respectively. The last column of Fig. 13 shows that our method has a high pixel-wise DC value of 0.982 on cervical cytology images, whereas DC values obtained by Lu *et al.*'s, Nosrati and Hamarneh, and Ronneberger *et al.*'s approach are 0.963, 0.968, and 0.966, respectively. Note that as the number of objects increases, our method remains more robust. Unlike the others, it also retains its accuracy.

C. Impact of the Proposed Cascade Regression Chain Model Parameters

1) Effect of TreeBoost Parameters: We firstly evaluate the impact of training data size N on boundary detection accuracy. Following leave-one-out cross validation method [54], we conservatively set the number of trees as 25, the maximal tree depth as 4, and the model iteration as 1. Table IV, Table V, and Table VI show the detailed results from the parameter tuning process for the five submitted metrics on three disparate microscopy datasets. Here, we take the evaluation on KIRC histopathology dataset for an example. As expected, increasing the number of training samples generally improves the segmentation accuracy, as the average DC value increases from 0.701 ($N=6250$) to 0.746 ($N=50000$). Also, increasing the number of samples seems to make the boundary delineations more consistent as reflected by the reduced ABD value from 6.93 to 5.22. It is worth noting that the segmentation accuracy begins to degrade after 50000. Therefore, in this paper, we choose $N=50000$, which is enough to generate reasonable and accurate results. We next test the effects of other TreeBoost parameters including the number of trees M in (9) and the depth of tree D . We simply find that not the more trees the better, but the longer it would take to do the training. Specifically, the average DC value increases from 0.695 ($M=5$) to 0.746 ($M=25$), but reduces after $M=25$, from 0.746 to 0.738. This may be due to the case that excessive regression trees lead to too complicated hypotheses, and result in overfitting. In this paper, we finally choose 25 trees at each iteration for the KIRC histopathology dataset. In general, a low depth will be likely to underfitting, while a high value will be likely overfitting. In our case, we find that the DC value presents an upward tendency and ABD value gradually reduces as the maximum value (0.7464 and 5.220, respectively) appears on depth 4 setting.

2) Effectiveness of the Proposed Cascade Regression Chain: As mentioned in Section III, our cascade regression chain make an attempt to recursively select and fuse context information, as well as appearance, in a unified framework. The first layers-trained regression trees are based purely on the local image appearance. Objects with strong appearance cues are often assigned high scores. These scores then start to influence their neighbors in the subsequent layers, especially if there are strong correlations between them. In Table IV above, we describe the

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

14

results of boundary score maps estimated without and with auto-context algorithm at 1st, 2nd, 3rd, and 5th iterations, respectively. It can be seen that there is a clear improvement until the 2nd model iteration (3rd layer) with the DC value increasing from 0.726 to 0.787 and ABD value reducing from 10.23 to 3.81 pixels. However, having a much enlarged appearance-context feature pool actually degrade performance when we considered to fuse it from the 3rd model iteration, which make an even stronger impact than simultaneous score context.

To qualitatively evaluate our proposed cascade regression chain at each iteration, we perform a MDL-based periodic B-spline contour inference on the corresponding boundary score map. As shown in Fig. 14, with the increase of iteration, the estimated score map for the selected DIC red blood cells becomes much clearer and inferred complete contours come much closer to the manual-delineated cell boundaries. Besides, we also conduct the maker detection experiments on all the four datasets. Compared with the initial maps without embedding auto-context algorithm, it shows that our cascade regression chain is tolerant to variations in scale and geometry, has excellent noise immunity, and can detect overlapping objects with incomplete boundaries, as illustrated in Fig. 15.

In our experiments, we adopted the leave-one-out cross validation to determine the key parameter values of the proposed regression chain model. Empirically, we adjusted the number of the layers first since it poses significant impact on the performance. However, these parameters do not operate independently and the effects of their settings are linked. For example, smaller values of the number of layers (or model iterations) lead to larger values of tree number for the same training risk. In fact, the joint parameter estimation in the proposed regression chain model should be treated as a multi-objective nonlinear optimization problem. Although several particle swarm optimization (PSO) algorithms or variants [57] could be used to estimate both the chain size and the parameters of a TreeBoost model simultaneously, these approaches may take a long time to reach convergence and have very limited searching ability to find the global minimum of a cost function. Usually, the joint parameters estimation and model training with PSO may have higher computational complexity since the searching space is huge, and it is still a great challenge for us to obtain the jointly optimal parameters setting. In addition, how to build a unified criterion for joint parameters estimation remains an open problem for further research. In this work, although we empirically determined the optimum combinations of fitting for the datasets, the proposed approach is able to achieve state-of-the-art performance.

D. Boundary Regression versus Boundary Classification

To prove the importance of our regression approach compared to classification, we also train a set of classification trees to segment the clustered cells/nuclei from the microscopy images.

In the boundary classification method, we employ classification trees as the base learners by using the same sampling scheme, features, and tree structure as boundary regression, to estimate the probability of each pixel belonging to the target boundary. The only difference is that classification training is

implemented by using the function

$$\mathcal{D}(\mathbf{p}) = \begin{cases} 1 & \text{if } \mathbf{p} \text{ is on the boundary} \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

to be approximated and the exponential loss [24]. Since the pixels on the boundary itself for which $\mathcal{D}(\mathbf{p})$ should return 1 and their neighbors for which it should return 0 look very similar (see Fig. 16(a)), the probability map obtained even from 2-layers model still contains many incorrect large values on the background (see Fig. 16(b)). This is because the boundary classification strategy aims to pixel-wisely label the target boundary according to its local appearance. In contrast, our regression result (see Fig. 16(c)) employs the pixels near the target boundary to estimate the boundary location, which provides rich context information to overcome the limitation of indistinct boundary appearances. As a result, the probability map, in turn, misguides the final contour inferences, as shown in Fig. 16(d), compared with Fig. 16(e). Also, since the ground truth data itself can be inaccurate, to quantitatively compare the segmentation accuracies of these two strategies, we dilate the manually-delineated boundary by one pixel as well as classification result and regression result to take into account more overlapping parts. As we can see from Fig. 17, our boundary regression mechanism gives better segmentation performance for all the four datasets, in terms of higher DC values and lower standard deviations, than the boundary classification.

VI. CONCLUSION

In this paper, we proposed a novel cell/nucleus segmentation method using a contour-seed pairs learning framework that formulates the detection and segmentation tasks in terms of a unified regression problem. The proposed cascade sparse regression chain model (CSRCM) jointly performs location recognition of each object and boundary detection of clustered objects. The model first learns a set of superior features from online learned convolutional filters with a sparse prior in each layer of chain before the cascade training. Then, in the CSRCM, with the help from the joint convolutional features, the method iteratively updates the location and clustered objects' boundary scores until convergence. MDL-constrained periodic B-spline fitting models can effectively infer the complete contours of individual cells/nuclei, achieving an optimal balance between the conciseness of splines and their fitting quality. The experiments demonstrate that the intertwined relationship of extracted features and estimated scores boosts the performance of both cell/nucleus detection and segmentation. They also demonstrate the effectiveness of the proposed method comparing to state-of-the-art works on four types of disparate microscopy datasets.

In the future, we would expect comparable performance, particularly for the case of more general application to the field of image processing. We would like to investigate some other techniques to improve the discriminative power of generated feature representations and investigate how to design a better algorithm to simultaneously determine the optimal parameter values of the proposed method. Besides, we would like to apply

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

15

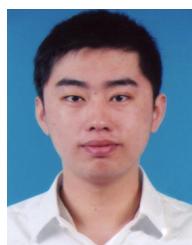
the proposed method to the automatic extraction of intrinsic measures for each cell/nucleus and their morphology analyses, for an objective evaluation and therapeutic outcome assessment of molecularly targeted cancer. Moreover, the handling of 2D microscopy images is our principle interests in this paper. However, the method could be applied to 3D structures for segmentation, which could be a promising future research direction.

REFERENCES

- [1] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 140–145, Sep. 2012.
- [2] C. S. Björnsson, G. Lin, Y. Al-Kofahi, A. Narayanaswamy, K. L. Smith, W. Shain, and B. Roysam, "Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue," *J. Neurosci. Methods*, vol. 170, no. 1, pp. 165–178, May 2008.
- [3] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [4] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, Feb. 2016.
- [5] H. Irshad, A. Veillard, L. Roux, and D. Racocceanu, "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, Apr. 2014.
- [6] R. Rubin, D. Strayer, E. Rubin, and J. McDonald, *Rubin's Pathology: Clinicopathologic Foundations of Medicine*. Baltimore, MD: Lippincott Williams & Wilkins, 2007.
- [7] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imaging*, vol. 35, no. 2, pp. 550–566, Feb. 2016.
- [8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, vol. 9351, pp. 234–241.
- [9] Y. Song, E.-L. Tan, X. Jiang, J.-Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, "Accurate cervical cell segmentation from overlapping clumps in pap smear images," *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 288–300, Jan. 2017.
- [10] M. Farhan, O. Yli-Harja, and A. Niemisto, "A novel method for splitting clumps of convex objects incorporating image intensity and using rectangular window-based concavity point-pair search," *Pattern Recognit.*, vol. 46, no. 3, pp. 741–751, Mar. 2013.
- [11] M. S. Nosrati and G. Hamarneh, "Segmentation of cells with partial occlusion and part configuration constraint using evolutionary computation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2013, vol. 8149, pp. 461–468.
- [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] H. Yang and N. Ahuja, "Automatic segmentation of granular objects in images: Combining local density clustering and gradient-barrier watershed," *Pattern Recognit.*, vol. 47, pp. 2266–2279, Jun. 2014.
- [14] C. Zhang, J. Yarkony, and F. A. Hamprecht, "Cell detection and segmentation using correlation clustering," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2014, vol. 8673, pp. 9–16.
- [15] J. Song, L. Xiao, and Z. Lian, "Boundary-to-marker evidence controlled segmentation and mdl-based contour inference for overlapping nuclei," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 451–464, Mar. 2017.
- [16] 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.
- [17] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE Trans. Med. Imaging*, vol. 31, no. 7, pp. 1448–1460, Aug. 2012.
- [18] P. Quelhas, M. Marcuzzo, A. M. Mendonça, and A. Campilho, "Cell nuclei and cytoplasm joint segmentation using the sliding band filter," *IEEE Trans. Med. Imaging*, vol. 29, no. 8, pp. 1463–1473, Aug. 2010.
- [19] P. Yan, X. Zhou, M. Shah, and S. T. C. Wong, "Automatic segmentation of high-throughput RNAi fluorescent cellular images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 109–117, Jan. 2008.
- [20] M. Horn and M. R. Berthold, "Towards active segmentation of cell images," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2011, pp. 177–181.
- [21] T. Chen, Y. Zhang, C. Wang, Z. Qu, F. Wang, T. Syeda-Mahmood, "Complex local phase based subjective surfaces (clapss) and its application to dic red blood cell image segmentation," *Neurocomputing*, vol. 99, pp. 98–110, Jan. 2013.
- [22] 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.
- [23] M. S. Nosrati and G. Hamarneh, "Segmentation of overlapping cervical cells: A variational method with star-shape prior," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2015, pp. 186–189.
- [24] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. Springer, 2001.
- [25] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, Oct. 2010.
- [26] M. Figueiredo, J. Leitao, and A. Jain, "Unsupervised contour representation and estimation using b-splines and a minimum description length criterion," *IEEE Trans. Image Proc.*, vol. 9, no. 6, pp. 1075–1086, 2000.
- [27] P. Grunwald, J. Myung, and M. Pitt, "Advances in minimum description length: Theory and applications," *Cambridge Univ. Press*, 2004.
- [28] D. Lalive, N. Barbot, and O. Boeffard, "Melodic contour estimation with b-spline models using a MDL criterion," in *Proc. SPECOM*, 2006, pp. 333–338.
- [29] 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.
- [30] M. Veta, A. Huisman, M. A. Viergever, P. J. van Diest, and J. P. W. Pluim, "Marker-controlled watershed segmentation of nuclei in H&E stained breast cancer biopsy images," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2011, pp. 618–621.
- [31] 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.
- [32] Q. Wen, H. Chang, and B. Parvin, "A Delaunay triangulation approach for segmenting clumps of nuclei," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2009, pp. 9–12.
- [33] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, and A. Madabhushi, "Expectation-maximization-driven geodesic active contour with overlap resolution (EMAGACOR): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1676–1689, Jul. 2010.
- [34] C. Chen, W. Wang, J. A. Ozolek, and G. K. Rohde, "A flexible and robust approach for segmenting cell nuclei from 2d microscopy images using supervised learning and template matching," *Cytometry A*, vol. 83, no. 5, pp. 495–507, May 2013.
- [35] C. Chen, H. Li, X. Zhou, and S. T. C. Wong, "Constraint factor graph cutbased active contour method for automated cellular image segmentation in RNAi screening," *J. Microsc.*, vol. 230, pp. 177–191, 2008.
- [36] D. M. Ushizima, A. G. C. Bianchi, and C. M. Carneiro, "Segmentation of subcellular compartments combining superpixel representation with voronoi diagrams," in *Overlapping Cervical Cytology Image Segmentation Challenge – ISBI 2014*, 2014.
- [37] A. Kuijper and B. Heise, "An automatic cell segmentation method for differential interference contrast microscopy," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.
- [38] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M. H. Barcellos-Hoff, "Iterative voting for inference of structural saliency and characterization of subcellular events," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 615–623, Mar. 2007.
- [39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [40] M. Law and A. Chung, "Three dimensional curvilinear structure detection using optimally oriented flux," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 368–382.
- [41] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. 511–518.
- [42] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) < 16

- [43] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.
- [44] A. Sironi, B. Tekin, R. Rigamonti, V. Lepetit, and P. Fua, "Learning separable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, 37, no. 1, pp. 94–106, Jan. 2015.
- [45] R. Rigamonti, M. A. Brown, V. Lepetit, "Are Sparse Representations Really Relevant for Image Classification?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1545–1552.
- [46] A. Coates and A. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 921–928.
- [47] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
- [48] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, pp. 455–500, 2009.
- [49] J. P. Vink, M. B. Van Leeuwen, C. H. M. Van Deurzen, and G. De Haan, "Efficient nucleus detector in histopathology images," *J. Microsc.*, vol. 249, no. 2, pp. 124–135, Feb. 2013.
- [50] A. Sironi, Engin Türetken, V. Lepetit, and P. Fua, "Multiscale centerline detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1327–1341, Jul. 2016.
- [51] J. Son, I. Jung, K. Park, and B. Han, "Tracking-by-segmentation with online gradient boosting decision tree," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3056–3064.
- [52] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [53] T. C. M. Lee, "An introduction to coding theory and the two-part minimum description length principle," *Int. Statist. Rev.*, vol. 69, no. 2, pp. 169–183, 2001.
- [54] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, 2012.
- [55] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect cells using non-overlapping extremal regions," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2012, vol. 7510, pp. 348–356.
- [56] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [57] H.-G. Han, W. Lu, Y. Hou, and J.-F. Qiao, "An adaptive-PSO-based self-organizing RBF neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 104–117, 2018.



Jie Song received the B.S. degree in computer science and technology from Nanjing University, Nanjing, China, in 2009 and the M.E. degree in computer science and engineering from Yunnan University, Yunnan, China, in 2013. He is currently pursuing the Ph.D. degree in computer science and technology at Nanjing University of Science and Technology, Nanjing, China. His research interests include biomedical image processing, machine learning, and pattern recognition.



Liang Xiao (M'11) was born in Hunan, China, in 1976. He received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Postdoctoral Research Fellow with the Pattern Recognition Laboratory, JUST. From 2009 to 2010, he was a Postdoctoral Fellow with Rensselaer Polytechnic Institute, Troy, NY, USA. Dr. Xiao is currently a Professor with the School of Computer Science and Engineering, JUST. Since 2013, he serves as the deputy director of Key Laboratory of Spectral Imaging Intelligent Perception-Jiangsu Province of China. Since 2014, he serves as the second director of the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education. His fields of interest include biomedical image processing, signal processing, image modeling and computer vision, machine learning, and pattern recognition.



Zhichao Lian received Bachelor degree and Master degree in computer science from Jilin University, Changchun, China, in 2005 and 2008 respectively. He received the Ph.D. degree from Nanyang Technological University in 2013. From 2012 to 2014, he was a Postdoctoral Associate in the Department of Statistics in Yale University. Dr. Lian is currently an Associate Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology in China. His research areas include biomedical image processing, pattern recognition, and neuro-imaging.