

# An Automatic Learning-Based Framework for Robust Nucleus Segmentation

Fuyong Xing\*, *Student Member, IEEE*, Yuanpu Xie, and Lin Yang, *Member, IEEE*

**Abstract**—Computer-aided image analysis of histopathology specimens could potentially provide support for early detection and improved characterization of diseases such as brain tumor, pancreatic neuroendocrine tumor (NET), and breast cancer. Automated nucleus segmentation is a prerequisite for various quantitative analyses including automatic morphological feature computation. However, it remains to be a challenging problem due to the complex nature of histopathology images. In this paper, we propose a learning-based framework for robust and automatic nucleus segmentation with shape preservation. Given a nucleus image, it begins with a deep convolutional neural network (CNN) model to generate a probability map, on which an iterative region merging approach is performed for shape initializations. Next, a novel segmentation algorithm is exploited to separate individual nuclei combining a robust selection-based sparse shape model and a local repulsive deformable model. One of the significant benefits of the proposed framework is that it is applicable to different staining histopathology images. Due to the feature learning characteristic of the deep CNN and the high level shape prior modeling, the proposed method is general enough to perform well across multiple scenarios. We have tested the proposed algorithm on three large-scale pathology image datasets using a range of different tissue and stain preparations, and the comparative experiments with recent state of the arts demonstrate the superior performance of the proposed approach.

**Index Terms**—Deep convolutional neural network, nucleus segmentation, sparse representation.

## I. INTRODUCTION

A persistent difficulty in evaluating digitized pathologic specimens is its subjective nature. Recent literature [1], [2] shows that computer-aid image analysis can significantly improve the objectivity and reproducibility. An automatic, high throughput image analysis system usually requires accurate and robust nucleus segmentation as the pre-requisite step. However, it is difficult to separate nuclei from each other as well as the background due to nucleus occlusion or touching, shape

Manuscript received June 13, 2015; revised September 08, 2015; accepted September 12, 2015. Date of publication September 23, 2015; date of current version February 01, 2016. This research is funded, by part, by NIH R01 AR065479-02. Asterisk indicates corresponding author.

\*F. Xing is with the Department of Electrical and Computer Engineering and the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: f.xing@ufl.edu).

Y. Xie is with the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611 USA.

L. Yang is with the Department of Electrical and Computer Engineering and the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2015.2481436

variation, intra-nucleus inhomogeneity, background clutter, and poor contrast (see Fig. 1). In order to tackle these challenges, many state-of-the-art approaches have been applied so far to medical image segmentation.

Among the established methods, watershed-based segmentation is one of the most widely used methods in medical image analysis. However, it is prone to over-segmentation. To address this limitation, marker-based watershed has been proposed and achieved great success in nucleus/cell segmentation. Lin *et al.* [3] have proposed a gradient weighted-distance transform method to locate nucleus centroids, which are viewed as markers for watershed segmentation of nuclei in 3D fluorescence microscopy images. Yang *et al.* [4] have exploited conditional erosion to detect nucleus markers for watershed segmentation in fluorescent nuclei images, and another mathematical morphology based marker detection is reported in [5], which supports watershed transform in RNA interference (RNAi) images. In addition, Cheng and Rajapakes [6] have employed an adaptive H-minima transform [7] to extract nucleus markers for the subsequent watershed segmentation in *Drosophila Kc167* cell images, Schmitt and Hasse [8] have detected the markers by localizing the gravitation centers of mass clumped cells with kernel-based iterative radial voting, and Mao *et al.* [9] have presented a two-step supervised learning approach to locate nuclei in bladder cancer images. The choice of marker detection algorithms for watershed segmentation highly depends on the specific applications, and it is difficult to apply a general rule. Meanwhile, the aforementioned detection methods are not robust to image noise or require careful parameter selection, and thus might need further processing to merge or split falsely segmented regions from watershed.

Another popular type of nucleus or cell segmentation methods is based on graphs or Voronoi diagrams. Al-Kofahi *et al.* [10] have first detected nuclei seeds using Euclidean distance map-based Laplacian-of-Gaussian, and subsequently applied a graph cut algorithm with  $\alpha$ -expansion [11] and graph coloring to final nucleus segmentation in histopathology images. However, it might require user's interaction to merge or split false initial segmentation. Bernardis and Yu [12] have proposed a graph partitioning algorithm, which defines two grouping cues composed of feature similarity-based short-range attraction and feature dissimilarity-based long-range repulsion, for cell segmentation in fluorescence microscopy images. However, this method does not well delineate cell boundaries. Yang *et al.* [13] have applied a concave vertex graph to automatic image analysis of histopathology specimens. By minimizing a cost function based on cellular characteristics, the shortest path in the constructed graph is calculated to separate the touching

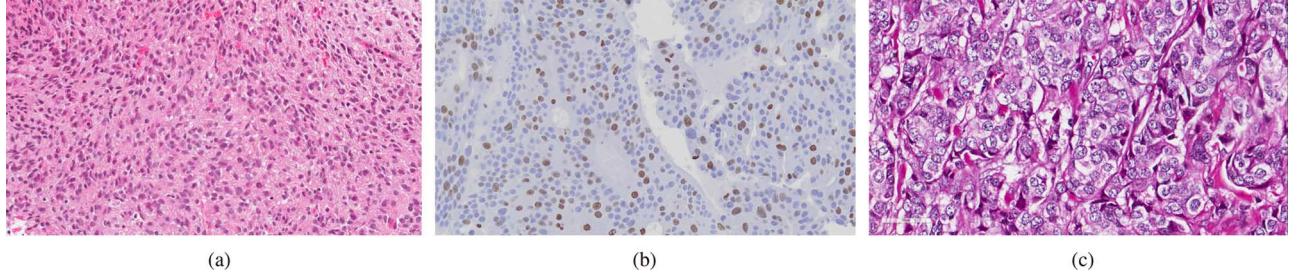


Fig. 1. Sample images of brain tumor (a), pancreatic neuroendocrine tumor (b), and breast cancer (c). Many nuclei touch with each other, exhibit weak nucleus boundaries, or have significant shape variations. The background clutter in the breast cancer image presents significant challenges as well.

cells. The dual graph for a Voronoi diagram, Delaunay triangulation, is used to segment clumps of nuclei in [14], where points of maximum curvature are used as vertices for Delaunay triangulation, and a constraint satisfaction network is applied to decomposition of the edge hypotheses. Another concave vertex-based touching cell splitting algorithm is presented in [15], which iteratively segments cells based on dominant concave region detection. These methods heavily rely on accurate concave point detection, which might be not easy to achieve in challenging histopathology images. Wu *et al.* [16] have split interacting fibroblast cells by seeking the shortest path in an adjacent graph, but it might exhibit high time cost when dealing with large-scale images. Similarly, Jones *et al.* [17] have proposed a shortest path searching algorithm within a specifically defined manifold for Drosophila cell segmentation, but it fails to handle weak cell boundaries in challenging cases.

Active contour or surface models are also commonly used for medical image segmentation. Fatakdawala *et al.* [18] have proposed an expectation maximization (EM) driven geodesic active contour [19] with an overlap resolution (EMaGACOR) scheme, which can automatically detect and segment lymphocytes on histopathology images. Similarly, Chang *et al.* [20] have applied a level set model with color and scale constraints to nuclei segmentation, and Padfield *et al.* [21] have added a size constraint to the geodesic model [19] to extract G2 phase nuclei of eukaryotic cells. In [22], level set techniques are used to detect, segment, and track leukocytes, which are constrained with the shape, size, and internal intensity homogeneity of the objects. However, these methods do not explicitly handle multiple touching objects, and need further processing to split nucleus or cell clumps. Yan *et al.* [23] have proposed a repulsive level set model to penalize cell overlapping by introducing an interactive scheme between cells, Qi *et al.* [24] have adopted a seed detection-based repulsive level set model to segment overlapping breast cancer cells in tissue microarray (TMA) images, and Mosaliganti *et al.* [25] have used a coupled level set model to segment cells in 3D fluorescent images, which respects underlying intensity and shape distributions. However, these implicit deformable models usually exhibit high computational cost for a large number of cells. Some improved level set implementation methods have been proposed for fast segmentation, including narrow band [26], graph-vertex coloring [27], convex optimization [28], and more advanced techniques [29]. Despite significant efficiency improvement, usually implicit active contour models are slower than corresponding parametric models [30]. In addition, although topology preserving level set [31],

[32] is proposed, generally implicit models that allow one initial contour to split or merge do not take advantage of known topological constraints. This is not preferred when the object positions are detected in advance. The parametric gradient vector flow (GVF) model [33], on the other hand, is combined with color gradient and  $l_2e$  robust estimation in [34] for unsupervised segmentation of lymphocyte cellular images. However, this method does not handle touching cells. Cai *et al.* [35] have presented a repulsive GVF snake model to segment and track axons in fluorescence microscopy images, but it assumes the rough shapes and areas of objects are known a priori, and this assumption does not hold for automatic nucleus or cell segmentation in histopathology images. A more general coupled parametric active contour model is presented in [30], which introduces a region-based term into the traditional deformable model [36] to prevent contour overlapping. However, it calculates the repulsion from all the other contours in the image, which might be computationally expensive for a large number of objects.

Histopathology images often exhibit significant variations in staining preparations, and thus simple color convolution usually under- or over-estimates the nuclei [37]. Furthermore, nucleus touching or overlapping presents more challenges for color convolutions. In order to reduce the sensitivity to batch effects and tumor heterogeneity, color normalization is often used before nucleus or cell segmentation. Kothari *et al.* [38] have exploited color maps to normalize the color space from sample images to reference images, and afterward separated the nuclei from the background with a linear discriminant analysis classifier. This color segmentation is quite effective to handle histopathology images but it is not designed to handle touching objects. Chang *et al.* [37] have applied the same color normalization technique to testing images, and achieved nucleus regions within a multi-reference graph cut framework. Thereafter, concave points are detected to build Delaunay triangulation graphs, to which geometric reasoning techniques are introduced to partition overlapped nuclei. This method relies on robust concave point detection. For our dataset, it is challenging to achieve clean masks for accurate concavity detection.

Nucleus/cell segmentation methods that only rely on image appearances may be insufficient to handle cells with missing or misleading boundaries, which often occur on touching/overlapping cells. In order to tackle these limitations, high level shape priors are introduced to level set models [39], [40] for better segmentation of partially occluded objects. Leventon *et al.* [41] have computed the shape prior by applying principle component analysis (PCA) to the signed distance functions of

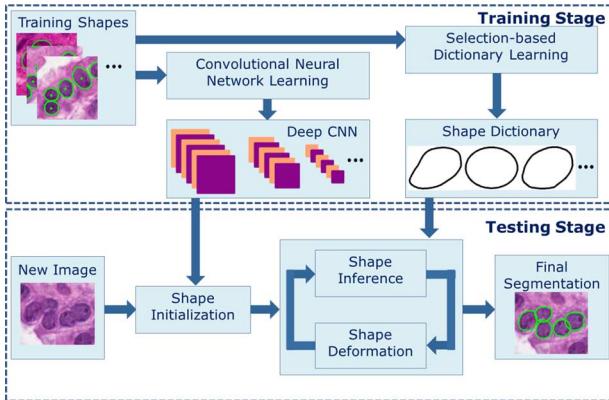


Fig. 2. The flowchart of the proposed framework.

the geodesic active contours, and similarly, another PCA-based shape prior is incorporated into an active contour model for medical image segmentation [42]. Cremers *et al.* [43] have used a related signed distance function [44] to represent a shape and introduced a labeling function to constrain the active domain of the shape prior. Recently, Ali *et al.* [45] have incorporated object shape priors into a level set formulation to handle occlusion, and it does not require point correspondence for shape representation, which are required for the active shape model [46]. However, all the level set based implementations exhibit high computational cost, and most of them are defined for single object segmentation and need additional efforts for simultaneous segmentation of multiple touching objects. Cai *et al.* [47] have applied shape constraints to the GVF model [33] to segment neurons, and Park *et al.* [48] have formulated the shape priors with a Gaussian mixture model and applied expectation maximum to shape inference for nanoparticle segmentation. However, both methods require clean edge maps to obtain desired segmentation, which are not easy to achieve in pathology images with background clutter. Sparse representation [49], [50] has recently been successfully applied to shape prior modeling for lung, liver, and cerebellum segmentation. In comparison with the PCA-based active shape model, sparse representation-based shape modeling is robust to false appearance cues and preserve local details, even though they are not statistically important [49]. These approaches are designed to segment single organ instead of multiple objects, and are not able to be directly applied to nucleus segmentation.

In this paper, we propose a robust learning-based framework (see Fig. 2) for automatic nucleus segmentation by combining both bottom-up and top-down information. In the training stage, we learn a deep convolutional neural network (CNN) model and a compact nucleus shape repository with a novel selection-based dictionary learning algorithm. In the testing stage, the CNN model is applied to the images to generate probability maps followed by iterative region merging for shape initialization, one per nucleus. Thereafter, the proposed nucleus segmentation algorithm performs shape deformation using a local repulsive deformable model and shape inference using the shape prior derived from the sparse shape model.

A preliminary version of our work [51] is presented in the International Conference of Medical Image Computing and Computer Assisted Intervention-MICCAI 2013. Compared with the

short conference version, each step of the proposed approach is explained in detail in this paper. Specifically, a deep convolutional neural network that is robust to inhomogeneous intensity and image noise is newly introduced to initialize shapes for nucleus segmentation. In addition, a novel repulsive deformable model with locality constraint is applied to touching nucleus segmentation with the combination of a selection-based nucleus shape prior model, and it is more robust to image noises than the balloon snake model in [51]. Finally, the proposed framework is extensively tested on three types of different datasets, and the experiments are redesigned to compare with six state of the arts. An extensive evaluation of CNN-based probability map generation (shape initialization) and parameter selection for the proposed algorithm are investigated and provided in this paper. In summary, the contributions of this paper are: 1) A deep CNN-based shape initialization is introduced to generate initial shapes, which learns hierarchical feature representation from raw images (YUV color space) and is relatively insensitive to image noise and inhomogeneous intensity; 2) An integrated framework consisting of selection-based sparse shape prior and efficient repulsive deformable model is proposed for contour evolution, which can effectively separate touching or overlapping nuclei; 3) The proposed framework is extensively tested on multiple types of clinical datasets using a range of different tissues and stain preparations.

## II. SHAPE INITIALIZATION

In digitized histopathology images, it is not infrequent that tumor nuclei form dense clumps and overlap with one another. In order to achieve nucleus segmentation with shape preservation that facilitates the subsequent morphological feature extraction, we incorporate a top-down shape prior model and a bottom-up deformable model into a unified framework. In order to facilitate the contour evolution step, we need to obtain robust initializations. For this purpose, we first learn a deep convolutional neural network (CNN) to generate a probability map which presents nucleus regions, and then apply an iterative region merging algorithm to the map to create the initial shape markers. The flowchart of the shape initialization is shown in Fig. 3.

#### A. Probability Map Generation

In our approach, the probability map is calculated from the classification results. Given an input image, each pixel is assigned a probability indicating how likely it belongs to part of a nucleus. This problem is solved by training a classifier based on a deep CNN using supervised learning. Recently, CNN has been proven to be effective for object detection, segmentation, and classification in nature and medical image analysis [52]–[56]. In particular, Ciresan *et al.* [57] have used a deep neural network algorithm to detect mitosis in breast cancer histology images, which can correctly separate mitosis from non-mitosis. Instead of relying on hand-crafted features which need to be specifically designed, CNN can automatically learn multi-level hierarchies of features which are invariant to irrelevant variations of samples while preserving relevant information [58].

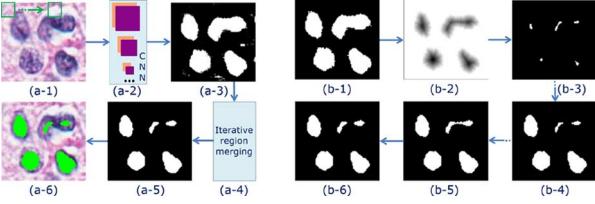


Fig. 3. Left: The flowchart of the shape initialization. Given an input image (a-1), small patches are cropped with a sliding window technique and fed into the learned CNN model (a-2) to generate a probability map (a-3), on which an iterative region merging algorithm (a-4) is used for nucleus shape marker generation (a-5, a-6). Right: The procedure of iterative region merging. Based on the noise-eliminated probability map (b-1), the inverse of the distance map (b-2) is calculated and used to generate initial markers (b-3) with H-minima transform. These initial markers iteratively expand based on distance values until markers merge into each other; during the merging procedure, the markers who will merge in next iteration are recorded and smoothed as the final shape markers (b-6). For those non-touching nuclei, the predicted nucleus regions are smoothed with a morphology operation and viewed as final shape markers. The image (b-4) in the bottom right is an intermediate result of marker growing. The image (b-5) is the result of the last iterative region merging, which merge two distinct markers. Therefore, the markers before this iteration are recorded as initial shapes.

TABLE I  
THE STRUCTURE OF THE CNN USED FOR THE BRAIN TUMOR  
DATASET IN OUR ALGORITHM

Layer No.	Layer Type	Output Size	Filter Size
1	Input	$55 \times 55 \times 3$	-
2	Convolutional	$50 \times 50 \times 48$	$6 \times 6$
3	Max Pooling	$25 \times 25 \times 48$	$2 \times 2$
4	Convolutional	$22 \times 22 \times 48$	$4 \times 4$
5	Max Pooling	$11 \times 11 \times 48$	$2 \times 2$
6	Fully-connected	$1024 \times 1$	-
7	Fully-connected	$1024 \times 1$	-
8	Output	$2 \times 1$	-

This learned feature representation exhibits both strong expressive and discriminative powers that can enhance the classification performance.

1) *Convolutional Neural Network*: A CNN [59] usually consists of successive pairs of convolutional and max-pooling layers, followed by several fully-connected layers. The deep architecture extracts hierarchical features from the input images, and the last layer, which is usually chosen as a sigmoid or softmax activation function, outputs probability distributions over classes based on the learned features [60], [61]. The CNN structure used for brain tumor in our algorithm is listed in Table I, and the architectures for NET and breast cancer datasets are similar except that the input patch sizes are  $39 \times 39$  (64 output feature maps) and  $45 \times 45$ , respectively.

In our deep architecture, multiple layers are stacked such that the output of the  $(l - 1)$ -th layer will serve as the input of the  $l$ -th layer. A convolutional layer learns a set of convolutional filters which will be used to calculate output feature maps, with all units in a feature map sharing the same weights. Suppose the  $l$ -th layer is a convolutional layer, a distinct set of feature maps will be selected from the output of the  $(l - 1)$ -th layer as the input to the  $l$ -th layer to compute the  $j$ -th feature map  $P_j^l$ . This procedure can be written as

$$P_j^l = f(\sum_i P_i^{l-1} * O_{ij}^l + \beta_j^l), \quad (1)$$

where  $i \in [0, N_{in}^{l-1} - 1]$ , and  $j \in [0, N_{out}^l - 1]$  denote the indices of the selected input feature maps and output feature maps respectively, with  $N_{in}^{l-1}$  and  $N_{out}^l$  representing the number of the input and output feature maps of the  $l$ -th layer. The  $\beta_j^l$  and  $O_{ij}^l$  represent learnable biases and convolutional kernels corresponding to the  $i$ -th input feature map and the  $j$ -th output feature map. The function  $f$ , defined as  $f(x) = \max(0, x)$ , represents the rectified linear units (ReLU) [62], which can significantly improve the time cost at the training stage [60].

The max pooling layer summarizes the activities and picks up the max values over a neighborhood region in each feature map [63], which not only reduces feature dimensionality but also introduces local shift and translation invariance to the neural network. This operation is usually performed separately in each feature map. The convolutional-pooling layers are stacked to learn local hierarchical features, based on which the fully-connected layers learn more higher level feature representation for classification. Different from the convolutional layers sharing weights, each neuron in fully-connected layers makes connections to all the neurons in the previous layer. The last layer is a softmax layer (fully-connected), which produces probability distributions over classes [60], [61], and the outputs of the softmax function can be interpreted as category probabilities [64]. In our application, it is a binary classification problem such that the last layer has only two neurons.

2) *CNN Training and Testing*: Our CNN is trained using raw pixel values (the YUV color space) of small image patches with certain size, centered on the pixel itself. Therefore, each pixel corresponds to a vectorized feature. We manually annotate the geometric centers of nuclei on the training images. The patches whose centers are closer than several pixels (chosen as 4) to the corresponding annotated nucleus centers are labeled as positive, otherwise negative. In order to achieve rotation invariance, all positive patches are rotated by  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ . This can also produce more positive training samples, which are significant when there exist less real nucleus patches than non-nucleus patches in the training images. Fig. 4 shows the learned filters and the feature maps in the first layer of the CNN model on the NET dataset.

After training, we apply the learned CNN model to small patches cropped from a new testing image using a pixel-wise sliding window with the same size as training patches. Those patches partly lying outside the image boundary are handled with zero padding. For each testing image, the CNN model with the sliding-window technique creates a probability map where each pixel is assigned a probability being close to nucleus centers, and those lying on the background would be with lower probabilities. This feed-forward procedure of probability map generation is carried out using the CUDA parallel computing platform on graphic processing units (GPUs), and the computational time can be improved at runtime. Fig. 5 shows the comparative foreground segmentation between a simple color deconvolution method [65] and the used CNN model. As we can see, the CNN model is more robust to intensity inhomogeneity and background clutter, and it can achieve much better performance than the color deconvolution method on the challenging datasets such as breast cancer.

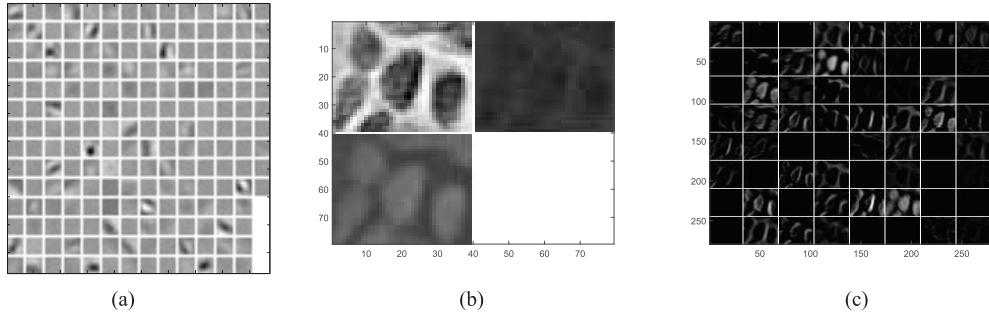


Fig. 4. (a) Learned kernels in the first layer of the used CNN model on the NET dataset. (b) One sample NET image patch (the YUV color space). (c) Several randomly selected feature maps in the first layer. The feature maps are generated by feeding the sample input image to the CNN model.

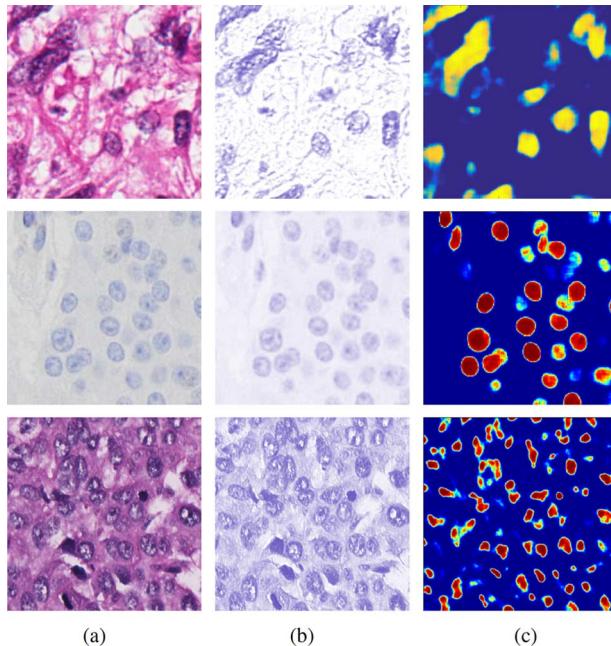


Fig. 5. Comparative foreground segmentation between the color deconvolution [65] (b) and the used CNN model (c) on brain tumor (top), NET (middle), and breast cancer (bottom) datasets (a). The CNN model is more robust to intensity inhomogeneity and background clutter than the color deconvolution method.

In our CNN model, the last layer of the CNN model is a softmax layer with two neurons, which output two probabilities (sum to 1) that the center pixel of the input image patch is positive and negative, respectively. The larger the probability value is, the more likely the input patch belongs to the corresponding category. A positive probability value larger than 0.5 indicates that the patch is less likely to be negative (the corresponding negative probability is smaller than 0.5). Therefore, we apply a threshold  $q_1 = 0.5$  to the positive probability map, which denotes how likely each pixel to be positive, to remove those pixels with probabilities lower than threshold  $q_1$  before further analysis. In addition, there may be some small noise-caused regions. We use the region size as a criterion and eliminate those regions with size less than a threshold ( $q_2$ ) before performing the subsequent shape initialization. We select the value of  $q_2$  by considering the original scale in terms of stain preparation. The NET image data is captured at  $20\times$  magnification with about 344 nm/pixel, and the other two are generated at  $40\times$  magnification with around 250 nm/pixel. Based on observations, we select  $q_2$  as 30, 70 and 70 pixels for NET, brain tumor, and breast cancer, respectively, and they are proper to eliminate image noises, which have much smaller areas than real nuclei. Therefore, the regions whose area are smaller than  $q_2$  are viewed as noises and eliminated without further consideration.

For non-touching or non-overlapping nuclei, the local maxima on the obtained probability map can be viewed as markers for the subsequent nucleus segmentation. However, it is challenging to detect correct extrema for overlapping nuclei. In [6], [9], H-minima transform [7] is exploited to suppress undesired minima and calculate initial markers for the subsequent segmentation. Inspired by this idea, we propose an iterative region merging algorithm to calculate the initial shapes to the best extent possible. Based on the probability map, we first calculate the distance map [66], where each pixel corresponds to a value measuring the shortest distance from the background. Next, H-minima transform is applied to the inverse of the distance map to achieve desired minima. For each connected region on the probability map, we iteratively expand the markers based on the distance values until all the markers merge; during the merging procedure, we record the markers who will merge in next iteration. Finally, we exploit a simple morphology operation to smooth the markers, which can effectively preserve the nucleus shapes and are used as initial shapes for the subsequent nucleus segmentation. For better illustration, we visualize the procedure of iterative region merging in the right panel of Fig. 3. For the non-touching or non-overlapping nuclei, the predicted nucleus regions are smoothed to form the final shape markers. In this scenario, it can reduce the running time due to no iterative region growing for these nuclei.

### B. Shape Initialization

Compared with the method presented in [6] that stops all the markers' growing for each connected region when the first merging occurs, our algorithm only prevents those will merge in next iteration from growing but keeps others expanding. The markers grow based on the distance values on the complement image, where the pixels with the same values actually form a level curve. In this way, the final markers will exhibit the largest areas before merging, and thus they can preserve nucleus shapes to the greatest extent without unnecessary early termination of growing. This is especially important for the touching/

**Algorithm 1:** Iterative region merging

---

**Input:** Probability map and initial depth  $dt_0$  for H-minima transform  
**Output:** Initial shapes

---

```

1. Do  $HT(I, dt_0)$  and get  $N_i(dt_0)$ 
2. for connected region  $c_i \in C$ 
3.    $dt = dt_0$ 
4.   if  $N_i(dt) > 1$ 
5.     while  $N_i(dt) > 1$ 
6.        $dt = dt + 1$ 
7.       Do  $MG(c_i, dt)$  and get  $N_i(dt)$ 
8.       for connected region  $c_{ij} \in c_i$ 
9.         Record corresponding markers as initial shapes
10.        which will merge into  $c_{ij}$  in next iteration
11.       end for
12.     end while
13.   else
14.     Record corresponding markers as initial shapes
15.   end if
16.   Marker smoothing
17. end for
```

---

overlapping nuclei exhibiting large size variations. Let  $C = \{c_i\}$  denote the set of connected regions,  $HT(I, dt)$  mean the H-minima transform on the inverse of the distance map  $I$  with threshold  $dt$ ,  $MG(c_i, dt)$  represent the marker growing operation on  $c_i$  with step size  $dt$ ,  $N_i(dt)$  be the number of minima in  $c_i$ , and  $c_{ij}$  be the  $j$ -th region component in  $c_i$ , the details of the iterative merging algorithm are listed in Algorithm 1.

### III. NUCLEUS SEGMENTATION

Given the initial shapes (contours) calculated from Section II, we propose to combine bottom-up and top-down information together to achieve nucleus delineation considering the fact that nucleus boundaries are often weak or even missing. In addition, the proposed algorithm can handle misleading cues due to inhomogeneous intensity or background clutter in the digitized specimens. Sparse shape model has shown to be more effective than PCA-based shape prior due to its insensitiveness to object occlusion [49], [50]. However, using all training shapes is inefficient during sparse reconstruction on a large dataset at run-time. KSVD [67] is a popular dictionary learning algorithm, but it is not designed as a discriminative and selection-based dictionary learning method with respect to classification and segmentation. In this work, we propose a novel and robust selection-based dictionary learning algorithm for nucleus shape modeling. Different from KSVD, this method directly selects the most representative nucleus shapes from the training dataset as dictionary bases. The robustness of the dictionary learning method is achieved by minimizing an integrated square error with a sparse constraint. In order to simultaneously and efficiently segment multiple nuclei, we combine a top-down shape prior model and a bottom-up deformable model with locality and repulsion constraints. The proposed algorithm alternately performs shape deformation using the efficient local repulsive deformable model, and shape inference using the shape prior derived from the sparse shape model. The flowchart of nucleus segmentation is shown in Fig. 2.

#### A. Selection-Based Sparse Shape Model

In this paper, nucleus shape  $v \in R^{2m}$  is represented by the concatenated 2D coordinates of  $m$  landmarks which are auto-

matically detected by the rules: 1) The two endpoints of the major axis of the shape are selected as major landmarks. 2) All the other landmarks are evenly interpolated along the shape. Given  $N$  nucleus shapes  $\{v_i\}_{i=1}^N$  aligned by Procrustes analysis, sparsity-based shape modeling aims to find a compact shape dictionary  $B = [b_1 \ b_2, \dots \ b_K]$  ( $\{b_k \in R^{2m}\}_{k=1}^K$  are bases) and a sparse coefficient  $\alpha$  such that any aligned shape  $v$  can be represented with a few bases:  $v = B\alpha + \epsilon$ , where  $\epsilon$  is the residual. The dictionary learning can be formulated as

$$\min_{B, \{\alpha_i\}} \sum_{i=1}^N \|v_i - B\alpha_i\|_2 + \lambda \|\alpha_i\|_1. \quad (2)$$

For a large nucleus shape dataset, it is intuitive to select a subset of the data as a shape repository that can sufficiently represent the whole dataset. This summarization can help remove outliers that are not the true representatives of the dataset and might reduce the computational time for runtime optimization due to the decreased object-space dimension. Based on these considerations, we propose a novel selection-based dictionary learning method for sparse representation by minimizing a locality-constrained integrated squared error (ISE). Scott [68] has shown that minimizing the ISE is equal to minimizing the objective function:  $\int g(x|\theta)^2 dx - 2/N \sum_{i=1}^N g(x_i|\theta)$ , where  $g(x|\theta)$  is a parametric model with parameter  $\theta$  and  $N$  is the number of data points  $\{x_i\}_{i=1}^N$ . In the sparse shape model, we have  $v_i = B\alpha_i + \epsilon_i$ , where  $\epsilon_i$  is the residual for the  $i$ -th shape. Therefore, we can model the residual density with function  $g(\epsilon|\theta)$  and minimize the objective function as follows

$$\begin{aligned} \min_{\theta} J(\theta) = & \min_{\theta} \left[ \left( \int g(\epsilon|\theta)^2 d\epsilon - \frac{2}{N} \sum_{i=1}^N g(\epsilon_i|\theta) \right) \right. \\ & \left. + \lambda \sum_{i=1}^N \sum_{k=1}^K |\alpha_{ik}| \|v_i - b_k\|^2 \right], \text{ s.t. } 1^T \alpha_i = 1, \forall i, \end{aligned} \quad (3)$$

where  $\epsilon_i = v_i - B\alpha_i$  and  $\alpha_i = [\alpha_{i1} \ \alpha_{i2} \ \dots \ \alpha_{iK}]^T$ . The first two terms form the  $L_2 E$  criteria, which is robust to outliers [68]. The last term constrains local representation of bases with weighted sparse codes, and is used to encourage each nucleus to be sufficiently represented by its neighboring dictionary bases for similarity preserving, which is essential in the sparse reconstruction. The constraint  $1^T \alpha_i = 1, \forall i$ , ensures the shift-invariance. The residual is modeled with multivariate normal distribution:  $\epsilon_i \sim N(0, \sigma^2 I_{2m})$ . In this way  $g(\epsilon_i|\theta) = \xi \phi(\epsilon_i|0, \sigma^2 I_{2m})$ , where  $\xi$  denotes the percentage of the inlier shapes that need to be estimated and  $\phi$  is the probability density function of multivariate normal distribution. Based on (3), the dictionary  $B$  and sparse coefficients  $\{\alpha_i\}_{i=1}^N$  can be calculated by estimating  $\theta = \{\xi, B, \alpha_1, \alpha_2, \dots, \alpha_N, \sigma^2\}$ .

Equation (3) can be solved by performing dictionary basis selection and coefficient computation alternately. As  $J(\theta)$  in (3) is differentiable with respect to  $\{b_k\}_{k=1}^K$ , projection based-gradient descent is utilized for minimization to update the bases, which are directly-selected shapes within each iteration. For coefficient calculation, we keep the dictionary fixed. Based on

the sparse reconstruction criterion, the sparse coding objective function can be rewritten as:

$$\min_{\{\alpha_i\}_{i \in A}} \left[ \sum_{i \in A} \|v_i - B\alpha_i\|^2 + \lambda \sum_{k=1}^K |\alpha_{ik}| \|v_i - b_k\|^2 \right], \\ s.t. \quad 1^T \alpha_i = 1, i \in A, \quad (4)$$

where  $A$  is the set of indices corresponding to estimated inlier shapes. Locality-constrained linear coding (LLC) [69] is applied to (4) for coefficient computation, where the neighboring bases are defined in terms of the Euclidean distances between the shape and dictionary bases.

Let  $B^0$  be the initial dictionary where the bases are randomly selected from the dataset, LLC is used in (4) to compute the current coefficients. Meanwhile, the active set  $A$  is updated with the indices corresponding to the  $N \cdot \xi$  shapes with the smallest reconstruction errors  $\|\epsilon_i\|_2^2$  in each iteration. At the  $t$ -th step,  $\theta^t = \{\xi^t, B^t, \alpha_1^t, \alpha_2^t, \dots, \alpha_N^t, \sigma^{t2}\}$ , and  $B^t = [b_1^t, b_2^t, \dots, b_K^t]$ , denote the gradient of  $J(\theta)$  in (3):  $\partial J(\theta)/\partial b_k, \partial J(\theta)/\partial \sigma^2$ , and  $\partial J(\theta)/\partial \xi$ , as  $J_{b_k}, J_{\sigma^2}$  and  $J_\xi$ , respectively. The basis  $b_k$  is updated by selecting the shape  $v_l$  which has the largest correlation between the displacement and the current  $b_k^t$ :

$$COR(v_l, b_k^t, J_{b_k}) = \frac{(v_l - b_k^t)^T (-J_{b_k})}{\|(v_l - b_k^t)\|_2 \| -J_{b_k} \|_2}. \quad (5)$$

Let  $A^t$  represent a set of indices corresponding to the current estimated inliers, the current reconstruction error  $E^t$  and the ISE error  $F^t$  are defined as

$$E^t = \frac{\sum_{i=1, i \in A^t}^N \|\epsilon_i^t\|_2^2}{\sum_{i=1}^N z(i \in A^t)}, \quad (6)$$

$$F^t = \xi^{t2} \phi(0|0, 2\sigma^{t2} I_{2m}) - \frac{2\xi^t}{N} \sum_{i=1}^N \phi(\epsilon_i^t | 0, \sigma^{t2} I_{2m}), \quad (7)$$

where  $z(x)$  is the indicator function. Assume that  $E_{\min}^t$  is the current reconstruction error,  $E_{rep}^t$  is the reconstruction error after replacing the  $k$ -th basis with  $v_l$ ,  $F_{\min}^t$  is the current ISE error, and  $F_{rep}^t$  is the ISE error after replacing the  $k$ -th basis with  $v_l$ , then the replacement will be performed only if  $E_{\min}^t > E_{rep}^t$  and  $F_{\min}^t > F_{rep}^t$ . The  $\sigma^2$  and  $\xi$  are updated in the negative gradient directions:

$$(\sigma^{t+1})^2 = \sigma^{t2} - \Delta h_{\sigma^2} \cdot J_{\sigma^2}, \quad \xi^{t+1} = \xi^t - \Delta h_\xi \cdot J_\xi, \quad (8)$$

where  $\Delta h_{\sigma^2}$  and  $\Delta h_\xi$  represent the learning rates. The details are listed in Algorithm 2.

Due to large shape variations of objects, it might be not suitable to learn a single shape prior model for all the shape instances. Inspired by [70], [71], we build multiple subpopulation shape prior models based on clustered shapes. Specifically, we group the aligned training shapes into several clusters with k-means, and learn one shape prior model for each group. For runtime optimization, we align the testing shape to each mean shape and select the one with smallest alignment error to be the

---

**Algorithm 2:** Selection-based dictionary learning using ISE

---

**Input:** Nucleus shapes  $\{v_1, v_2, \dots, v_N\}$

**Output:** Dictionary  $B = [b_1 \ b_2 \ \dots \ b_K]$

---

1. Initialize  $B$
  2. **for**  $t = 1 : T$  ( $T$  is maximum value of iteration)
  3.      $\alpha_i \leftarrow$  solution of (4)
  4.     **for**  $k = 1 : K$
  5.         Compute  $COR(v_l, b_k^t, J_{b_k})$  using (5)
  6.         Compute  $E_{\min}^t, E_{rep}^t, F_{\min}^t, F_{rep}^t$  with (6,7)
  7.         **if**  $(E_{\min}^t > E_{rep}^t \text{ and } F_{\min}^t > F_{rep}^t)$
  8.             Update  $b_k$  and  $A$
  9.         **end**
  10.      **end**
  11.     Update  $\sigma^2$  and  $w$  using (8)
  12. **end**
- 

active model, and perform shape inference using (4) with the learned dictionary.

### B. Shape Deformation

In this step, we propose an efficient shape deformation method based on the Chan-Vese model for nucleus segmentation. The Chan-Vese model [72] is formulated based on the well-known Mumford-Shah functional [73], and consists of two region-based data fitting terms, one for foreground and the other for background, and several regularization terms. Denote by  $I$  the image and  $v$  the contour, the energy function  $J(v)$  is

$$J(v) = \int_{\Omega} (I(\mathbf{x}) - u)^2 d\mathbf{x} + \int_{\Omega_0} (I(\mathbf{x}) - u_0)^2 d\mathbf{x} \\ + \gamma |v| + \eta |\Omega|, \quad (9)$$

where  $\Omega$  and  $\Omega_0$  denote the regions inside and outside the contour  $v$ ,  $u$  and  $u_0$  represent the average intensity of  $\Omega$  and  $\Omega_0$ , respectively, and the last two terms are the length of  $v$  with weight  $\gamma$  and the area of  $\Omega$  with parameter  $\eta$ , respectively.

Due to intensity inhomogeneity and noise inside nuclei, it might be not sufficient to only apply the original Chan-Vese model in (9) to nucleus segmentation on digitized specimens. In order to enhance the robustness, we add an edge detector into (9) combining with the region-based data fitting term to better move contours towards nucleus boundaries. In addition, the model in (9) is not designed to separate touching or overlapping objects. Given multiple initial contours (shapes), the active contours may cross one another during their evolutions. To address this limitation, a repulsive term is introduced into (9)[23], [30], [45]. Instead of moving each contour independently, the introduced interaction is modeled concurrently among all contours during their evolutions. Consider an image  $I$  containing  $N$  nuclei, denoted by  $v_i (i = 1, \dots, N)$  with  $v_i$  representing the  $i$ -th contour. The energy function  $J$  for nucleus segmentation

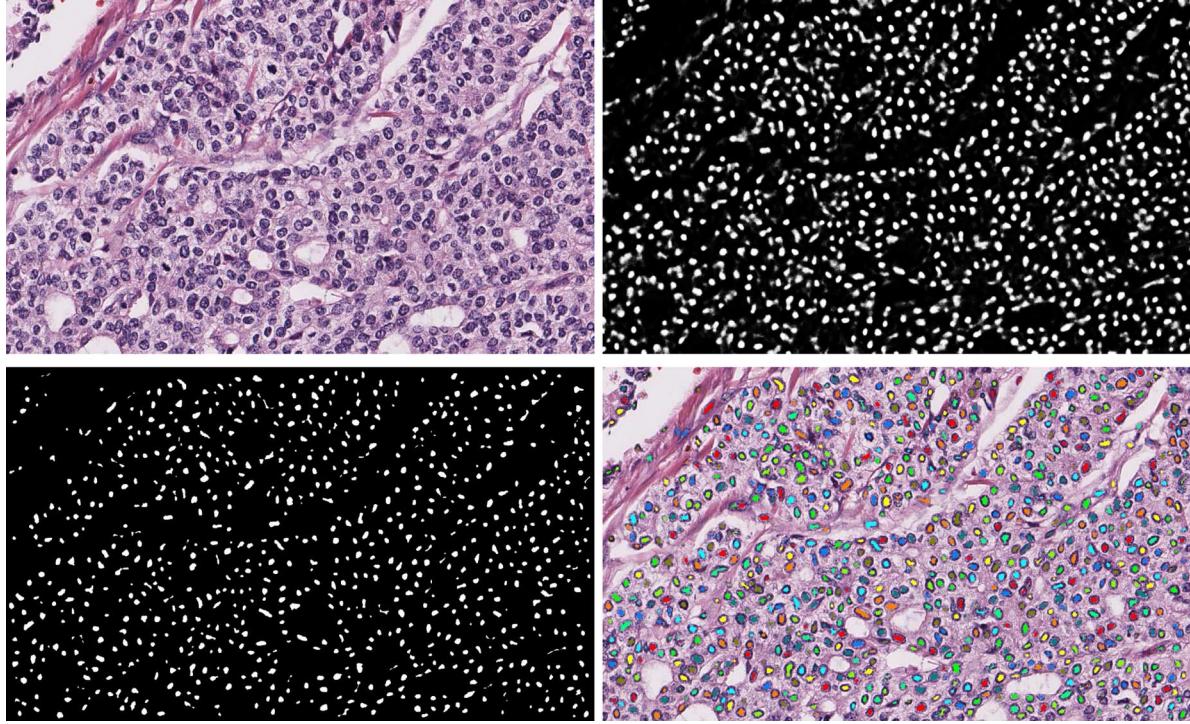


Fig. 6. Shape initialization using the proposed algorithm on one sample image of the breast cancer dataset. Top left: original image; Top right: probability map from CNN; Bottom left: initial nucleus shapes from iterative region merging; Bottom right: original image overlaid with initial nucleus shapes. Note that all the nuclei in the connected regions touching image boundaries are ignored.

combining the driving and repulsive mechanisms thus can be expressed as follows:

$$\begin{aligned} J = & \lambda_1 \sum_{i=1}^N \int_{\Omega_i} (I(\mathbf{x}) - u_i)^2 d\mathbf{x} + \lambda_2 \int_{\Omega_0} (I(\mathbf{x}) - u_0)^2 d\mathbf{x} \\ & + \lambda_3 \sum_{i=1}^N \int_0^1 e(v_i(s)) ds + \omega \sum_{i=1}^N \sum_{j=1, j \neq i}^N \int_{\Omega_i \cap \Omega_j} 1 d\mathbf{x} \\ & + \sum_{i=1}^N \gamma |v_i|, \end{aligned} \quad (10)$$

where  $\Omega_i$  and  $\Omega_0$  represent the region inside  $v_i$  and outside all the contours (background), the third term with  $e(v_i(s))$  is the edge detector and chosen as  $-\|\nabla I(v_i(s))\|^2$  ( $s \in [0, 1]$  is the parameter for contour representation), and the fourth term denotes the repulsion preventing contours from crossing with each other. We remove the area term because its weight  $\eta$  is usually chosen as 0 in our implementation.

For each nucleus, the model in (10) requires calculation of repulsion from all the other nuclei on the image, and this leads to high time cost when  $N$  is large. However, practically each nucleus is often surrounded by a limited number ( $M \ll N$ ) of adjacent nuclei, and only its neighboring nuclei make dominant repulsive contributions to its shape deformation during contour evolution. This suggests that we can deform shape  $v_i$  in its local coordinate system for computational efficiency, and it can be implemented by simply using  $v_i$ 's  $M$  nearest neighbors  $V_i$ . We thus replace the repulsion term in (10) with  $REP$ :

$$REP = \omega \sum_{i=1}^N \sum_{j \in V_i} \int_{\Omega_i \cap \Omega_j} 1 d\mathbf{x}, \quad (11)$$

The repulsion in (11) reduces the computation complexity significantly from  $\mathcal{O}(N^2)$  to  $\mathcal{O}(N \cdot M)$  due to  $M \ll N$ . In our segmentation framework, since the  $V_i$  can be pre-determined using the proposed deep CNN-based iterative region merging algorithm, minimizing the energy function with (11) is much faster compared with using the repulsion from all the other nuclei on the image.

Substituting (11) into (10) and using the Euler-Lagrange equations associated to the minimization of (10), we can get the following evolution equation

$$\frac{\partial v_i}{\partial t} = \left| \frac{\partial v_i}{\partial s} \right| \mathbf{n}_i (-\lambda_1 (I - u_i)^2 + \lambda_2 (I - u_0)^2 - \lambda_3 \nabla e(v_i) - \omega \sum_{j \in V_i} z_j(v_i) + \gamma \rho(v_i)), \quad (12)$$

where  $\mathbf{n}_i$  is the normal unit vector of  $v_i$ , and  $z_j(\mathbf{x})$  represents the indicator function:  $z_j(\mathbf{x}) = 1$  if  $\mathbf{x} \in \Omega_j$ , otherwise 0.  $\rho(\cdot)$  denotes the curvature. Given the initial shapes, we can iteratively evolve the contours toward desired nucleus boundaries.

Following [30], we solve the deformable model with parametric active contours, and this facilitates the sparse shape modeling in the proposed framework. This is different from the models presented in [23], [45], where the level set formulation that represents contours implicitly is used for energy minimization. The proposed model can take advantage of known object topology such that evolving contours are not allowed to split or merge, while generally level set models are designed to handle topology changes and thus it might generate undesired contours due to intensity heterogeneity. Compared with the shape prior based level set in [40], which only

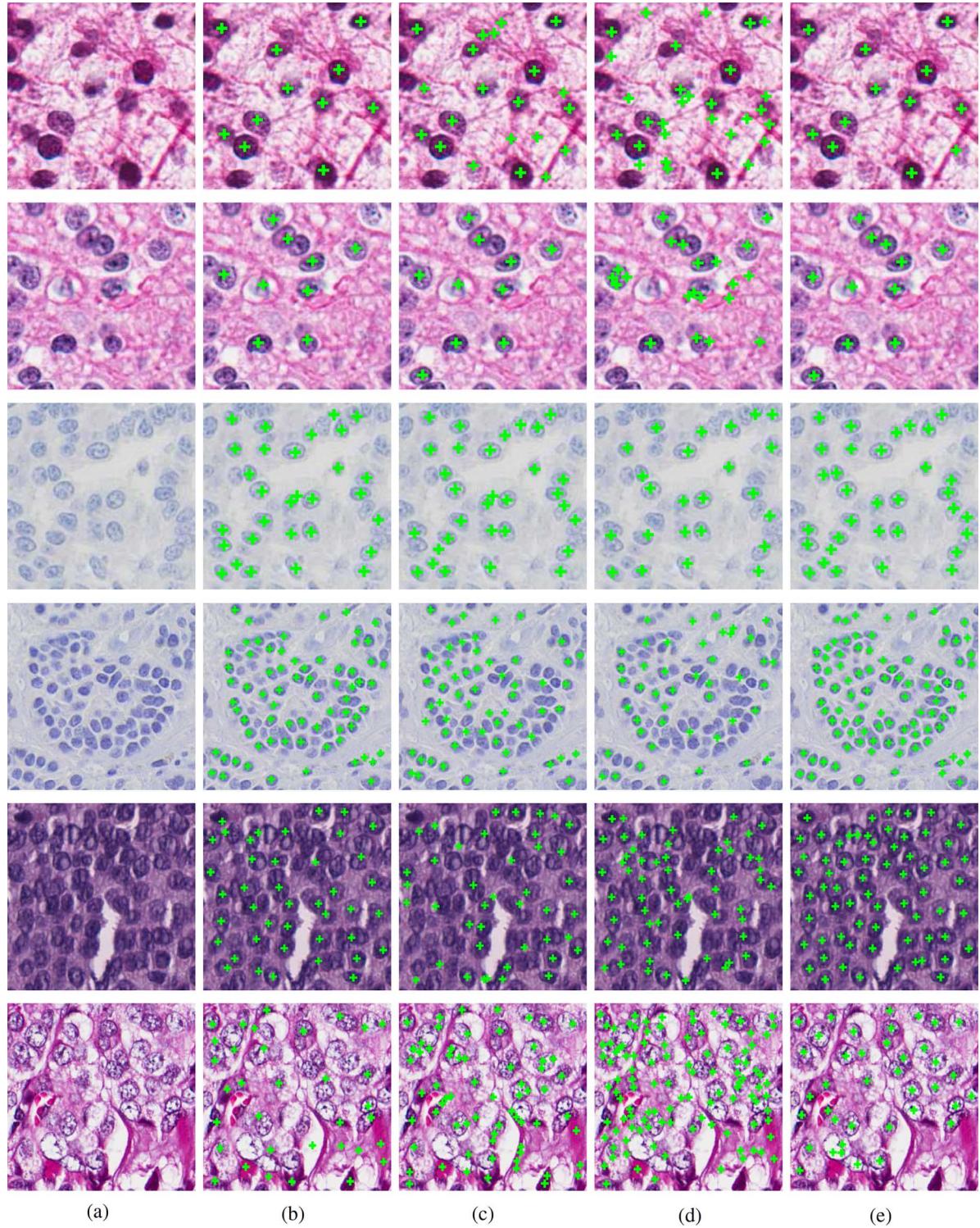


Fig. 7. Comparative nucleus detection using different methods on several sample patches of brain tumor (rows 1 ~ 2), NET (rows 3 ~ 4), and breast cancer (rows 5 ~ 6). (a) Original images, (b) SVM, (c) RF, (d) DBN, and (e) Proposed. Note that the nuclei touching image boundaries are ignored.

use a single shape template, we exploit a shape dictionary that allows much stronger representation power to constrain the contour deformation.

Given initial contours, the proposed segmentation framework alternately performs shape deformation with the repulsive active contour model and shape inference with sparse shape prior. The shapes always expand from inside nuclei, one per

nucleus, and evolve towards nucleus boundaries. In the active contour model, contours move based on image appearance information until (12) reaches a stable state, where the associated energy function achieves a minimum value; in the shape inference stage, contours evolve based on high level shape prior to constrain the shapes. This alternative operation scheme of combining bottom-up and top-down information has been

TABLE II  
EVALUATION OF NUCLEUS DETECTION ON THREE DATASETS IN TERMS OF MEAN AND STANDARD DEVIATION (STD) OF PRECISION, RECALL AND  $F_1$  SCORE

	Brain tumor (Mean/STD)			NET (Mean/STD)			Breast cancer (Mean/STD)		
	P	R	$F_1$	P	R	$F_1$	P	R	$F_1$
SVM	0.72/0.21	0.82/0.17	0.74/0.16	0.60/0.15	0.94/0.04	0.72/0.12	0.56/0.18	0.65/0.14	0.58/0.12
RF	0.53/0.20	0.94/0.12	0.66/0.18	0.71/0.10	0.88/0.10	0.78/0.08	0.36/0.14	0.74/0.15	0.46/0.14
DBN	0.52/0.19	0.79/0.20	0.60/0.17	0.81/0.09	0.77/0.09	0.78/0.05	0.39/0.14	0.67/0.12	0.47/0.11
proposed	0.72/0.20	0.88/0.15	0.77/0.16	0.84/0.08	0.93/0.05	0.88/0.04	0.71/0.14	0.88/0.09	0.78/0.08

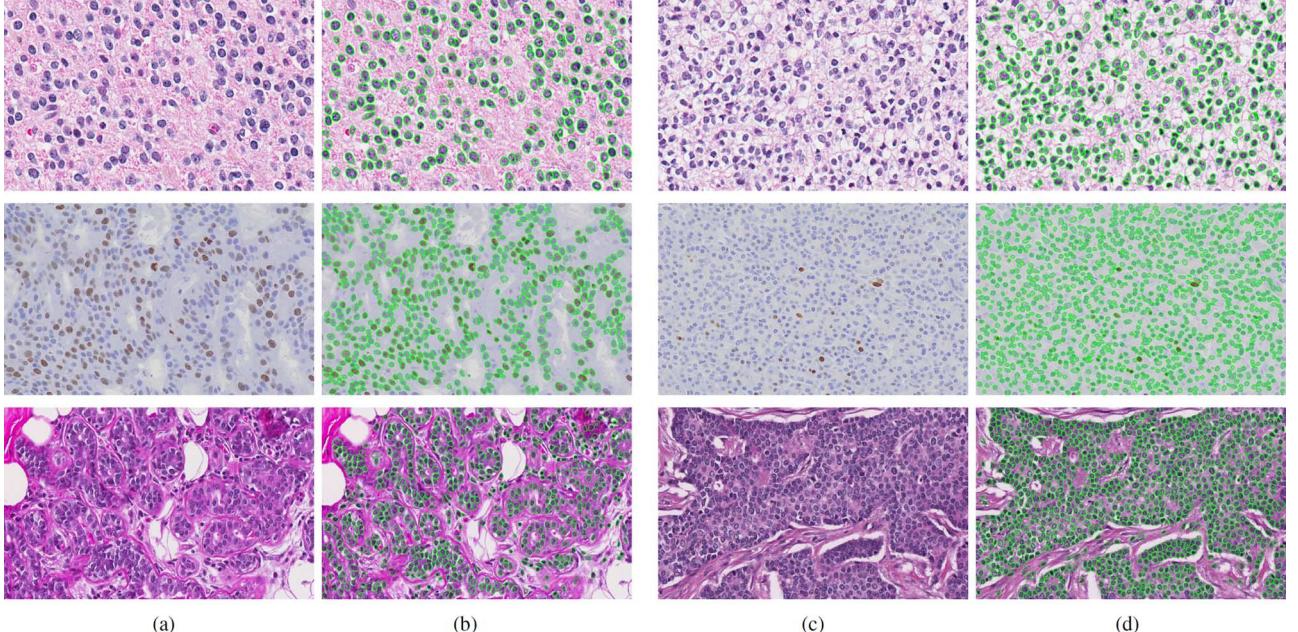


Fig. 8. Segmentation results using the proposed algorithm on six sample images of the brain tumor (top), NET (middle), and breast cancer (bottom) datasets. The original images are shown in (a) and (c), with the corresponding segmentation results in (b) and (d), respectively. Note that all the nuclei in the connected regions touching image boundaries are ignored.

successfully applied to biomedical image segmentation [49], [50].

#### IV. EXPERIMENTAL RESULTS

The proposed framework has been extensively tested on three types of histopathology images: brain tumor, NET, and breast cancer. In each dataset, each image corresponds to one specimen slice. Roughly half of images are randomly selected for training on brain tumor and breast cancer. The NET dataset is not as challenging as the other two, and thus we randomly choose one third for training such that we can have more testing data for algorithm evaluation. In total we have 31, 22, and 35 brain tumor, NET, and breast cancer images, respectively, and 614400 positive and 921600 negative image patches for each dataset are used for CNN model training. For testing data, we have 1469, 1765, and 1514 nuclei on 30 brain tumor, 51 NET, and 35 breast cancer images, respectively. The gold standards of nucleus contours are manually labeled and confirmed by multiple pathologists at University of Kentucky. The CNN model is implemented with the CUDA parallel computing platform [60], and Matlab is used for the iterative region merging algorithm, sparse shape prior model and local repulsive deformable model on a machine with 3.5 GHz Intel Xeon CPU and 128 GB memory.

The initial depth parameter for H-minima transform is set as  $dt_0 = 1$ . The  $\lambda$  in (3), (4) is a regularization parameter, controlling the trade-off between the error and the locality penalty, and it is set as  $5 \times 10^{-3}$ . For the repulsive deformable model in (12),  $\lambda_1$  and  $\lambda_2$  correspond to the first two data fitting terms,  $\lambda_3$  and  $\omega$  weight the contributions of the edge detector and the repulsive term, respectively, and  $\gamma$  controls the intrinsic properties of the contours. The parameter  $\omega$  and the neighborhood size  $M$  are important in nucleus segmentation, and we have experimentally shown their effects on the segmentation performance in Section IV-C. For the other parameters, we set  $\lambda_1 = \lambda_2 = 1$ ,  $\lambda_3 = 0.2$ , and  $\gamma = 1$  by following the suggestions in [72], [23]. The number of shape clusters depends on the shape variations of the nuclei, and in our experiments we found that three clusters are sufficient to model nucleus shape prior on each dataset. Finally, each nucleus contour is represented by 60 control points. The parameters are fixed for all three datasets, which demonstrates the generality of the proposed method.

##### A. Nucleus Detection With Shape Initialization

Shape initialization is actually nucleus detection in our segmentation framework, which provides nucleus positions for the subsequent active contour model. We conduct both qualitative and quantitative analyses to evaluate the nucleus detection. The gold standard of each nucleus center is calculated as the mean of

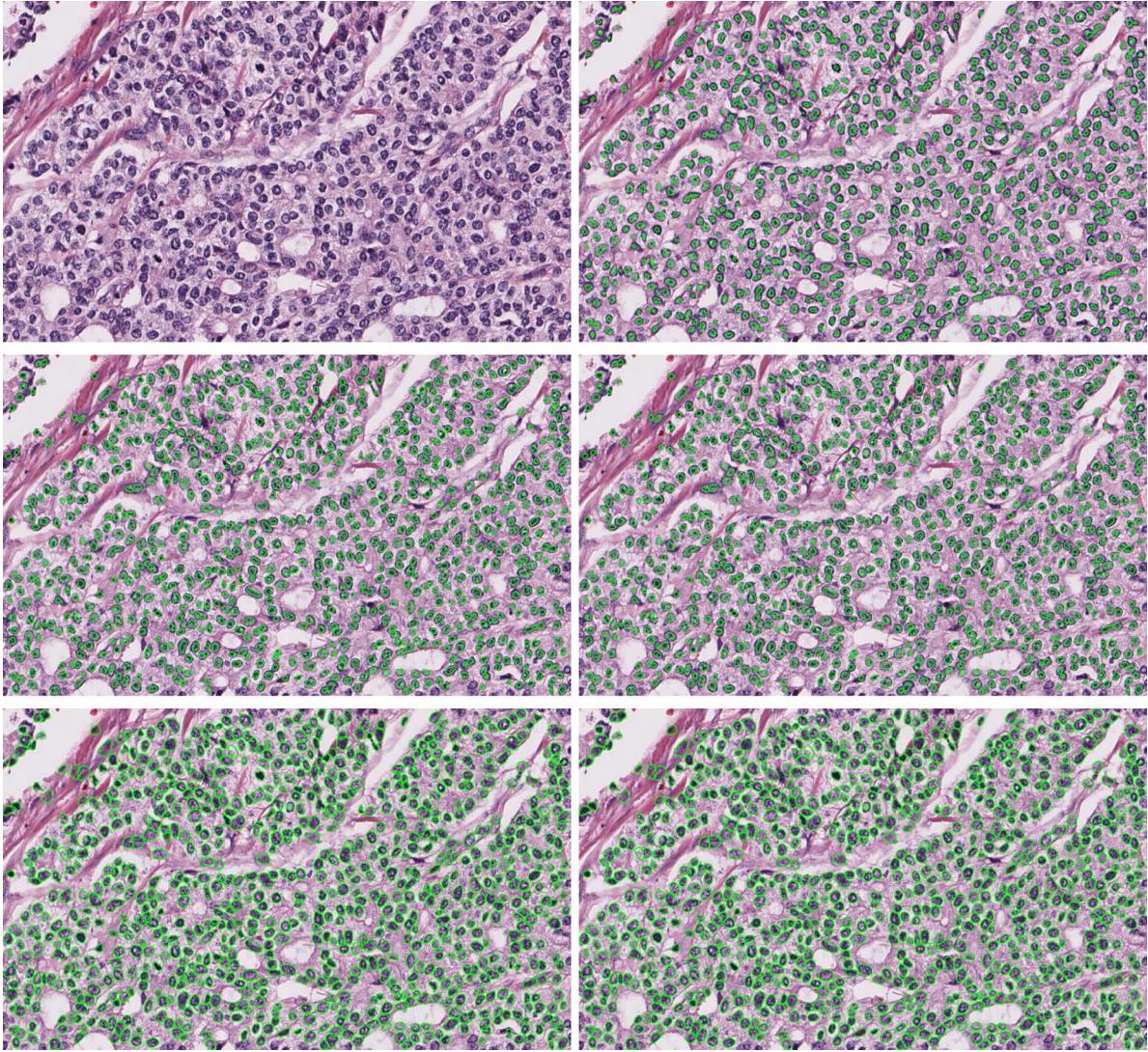


Fig. 9. The procedure of shape evolution on one sample breast cancer image. From top to bottom and from left to right: original image, initial shapes, intermediate shape evolution before using shape prior model for inference, intermediate shape evolution after using shape prior model for inference, shape deformation and inference in last iteration. Note that all the nuclei in the connected regions touching image boundaries are ignored.

the annotated contour, and the detected nucleus centers are the local maxima on the probability map. Fig. 6 shows the shape initialization using the deep CNN model and the iterative region merging algorithm on one sample slide of the breast cancer dataset. We can see that many nuclei are correctly detected. Fig. 7 shows the comparative detection results between the proposed CNN model and scalable SVM [74], random forest (RF) [75], and deep belief network (DBN) [76] on three datasets. As one can tell, the proposed method produces much better performance than the others, especially on the more challenging breast cancer dataset.

We apply the metric in [80] for quantitative analysis, which defines the gold standard regions as circular regions within 15 pixels for every nucleus center. A detected nucleus centroid is considered to be a true positive (TP) only if it lies within the gold standard region; otherwise, it is considered as a false positive (FP). Each TP is matched with the nearest gold standard nucleus center. The gold standard nucleus centers that are not matched by any detected results are considered to be

false negatives (FN). Based on these definitions, we can compute the precision (P), recall(R), and  $F_1$  score as

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F_1 = \frac{2PR}{P + R}. \quad (13)$$

Table II lists the evaluation of nucleus detection for different methods. It is clear that the proposed CNN model produces the best results, and the lowest standard deviations in almost all metrics demonstrates the strong reliability of our method. Specifically, SVM might need sophisticated feature design to provide desired results. RF and DBN produce fair performance on the NET images but much lower detection accuracy on the other two, which are more challenging. The proposed approach outperforms the traditional classification methods (SVM and RF) on all the datasets due to the learned feature representations, which are robust for object classification. The reason why the DBN model provides

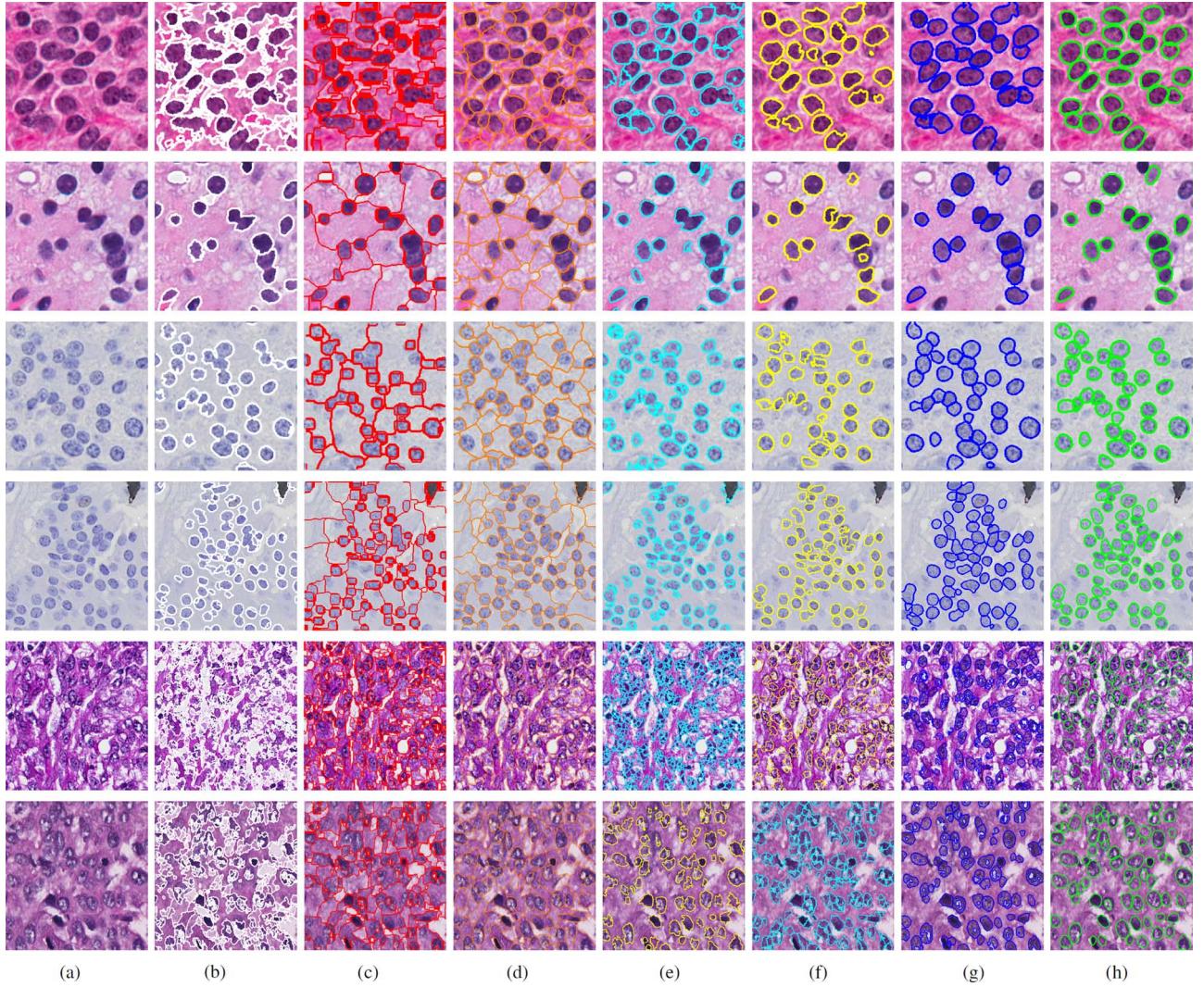


Fig. 10. Comparative segmentation using different methods on several sample patches of brain tumor (rows 1 ~ 2), NET (rows 3 ~ 4), and breast cancer (rows 5 ~ 6). (a) Original images, (b) MS, (c) ISO, (d) SUP, (e) GCC, (f) MWS, (g) RLS, and (h) Proposed. MWS, RLS, and the proposed use the same initialization. Note that the nuclei touching image boundaries are ignored.

low accuracy might be that the current DBN network is not very deep; a deeper DBN might be able to provide better performance, but this will require much higher training and testing time cost.

### B. Nucleus Segmentation

The segmentation results using our method on six sample slide digitized images are shown in Fig. 8. For better illustration, the procedure of shape evolution is shown in Fig. 9. Many nuclei exhibit weak boundaries on the NET dataset, and the brain tumor and breast cancer images have heavy background clutter and dense nucleus clustering. The proposed algorithm can accurately detect and segment thousands of nuclei with shape preservation. Fig. 10 gives the comparative segmentation results on six representative image patches including mean shift (MS) [77], isoperimetric graph partition (ISO) [78], superpixel (SUP) [79], marker-based watershed (MWS), graph-cut and coloring (GCC) [10], repulsive level set (RLS) [24], and the

proposed framework. It is clear that MS, ISO, and SUP are general segmentation algorithms that require further processing for touching object segmentation. The contours created by MWS does not well describe nucleus boundaries. GCC is unable to handle weak boundaries on nucleus overlapping regions, and therefore it is prone to under-segmentation. RLS provides more accurate nucleus boundaries than MWS and GCC; however, it does not preserve topology and therefore may generate some small holes (false splitting contours) inside/outside individual nuclei due to intensity inhomogeneity. In addition, it does not take advantage of the nucleus shape priors either. On the contrary, ours can effectively handle nucleus overlapping because of the constraints of object topology preserving and learned shape prior, and therefore provides more accurate segmentation results (last column of Fig. 10).

In order to quantitatively analyze the pixel-wise segmentation accuracy, we apply multiple metrics including Dice similarity coefficient ( $DSC$ ), Hausdorff distance ( $HD$ ), and mean absolute distance ( $MAD$ ), to the evaluation of the algorithm.

TABLE III

COMPARATIVE PIXEL-WISE SEGMENTATION ACCURACY ON BRAIN TUMOR, NET, AND BREAST CANCER. FOR EACH METRIC (DSC, HD, MAD), THE MEAN, MEDIAN, STANDARD DEVIATION (STD), AND THE SORTED 80% HIGHEST ACCURACY AMONG ALL THE RESULTS ARE LISTED

brain	DSC				HD				MAD			
	Mean	Median	STD	80%	Mean	Median	STD	80%	Mean	Median	STD	80%
MS	0.74	0.84	0.24	0.90	9.98	4.84	11.95	16.29	5.23	3.23	5.29	6.79
ISO	0.70	0.75	0.19	0.84	10.15	5.67	12.41	13.62	6.63	4.97	6.49	7.83
SUP	0.75	0.84	0.20	0.89	10.89	5.00	11.07	19.03	5.32	3.16	5.01	8.04
GCC	0.81	0.86	0.13	0.90	7.07	4.12	6.47	10.63	3.85	2.92	2.95	4.85
MWS	0.81	0.86	0.15	0.92	7.21	4.56	6.30	11.06	3.57	2.50	3.04	5.06
RLS	0.80	0.83	0.11	0.90	8.51	5.00	8.52	14.14	4.59	2.46	6.53	5.58
proposed	0.85	0.89	0.11	0.93	5.06	2.91	5.26	6.37	3.26	2.18	2.89	4.41
NET	DSC				HD				MAD			
	Mean	Median	STD	80%	Mean	Median	STD	80%	Mean	Median	STD	80%
MS	0.66	0.72	0.20	0.82	7.01	4.74	5.88	10.50	4.35	3.43	2.57	6.12
ISO	0.48	0.50	0.20	0.67	10.02	5.94	13.65	10.58	8.14	6.81	6.93	8.89
SUP	0.75	0.81	0.17	0.87	6.62	3.73	6.85	9.62	3.77	2.53	3.29	4.78
GCC	0.61	0.69	0.21	0.77	6.37	4.29	4.23	10.10	5.00	3.87	2.58	6.78
MWS	0.82	0.89	0.17	0.93	4.12	2.75	3.51	6.41	2.33	1.38	5.53	3.18
RLS	0.84	0.86	0.09	0.91	2.71	1.46	3.11	4.01	2.26	1.69	1.93	2.79
proposed	0.92	0.95	0.09	0.98	2.41	1.35	2.99	3.14	1.58	0.98	1.75	2.10
breast	DSC				HD				MAD			
	Mean	Median	STD	80%	Mean	Median	STD	80%	Mean	Median	STD	80%
MS	0.49	0.51	0.25	0.74	20.83	17.77	15.59	28.70	13.17	11.74	8.27	18.63
ISO	0.56	0.58	0.19	0.75	17.19	15.40	10.54	24.36	11.59	10.38	5.82	15.89
SUP	0.68	0.72	0.18	0.84	17.14	13.89	13.10	24.72	9.32	7.74	6.02	13.29
GCC	0.59	0.62	0.23	0.81	16.84	15.59	10.40	23.91	10.75	9.43	6.57	15.64
MWS	0.73	0.79	0.20	0.88	11.12	9.66	7.25	14.80	6.66	5.08	5.49	9.24
RLS	0.77	0.81	0.15	0.88	10.50	8.12	7.46	14.76	6.30	4.80	4.95	8.25
proposed	0.80	0.86	0.15	0.92	8.60	6.37	6.77	13.08	6.24	4.53	4.90	9.61

Let  $\Omega_{sr}$  and  $\Omega_{gt}$  represent the regions inside the automatic segmentation contour  $v_{sr}$  and the gold standard contour  $v_{gt}$ , respectively, the metrics are defined:

$$\begin{aligned} DSC &= \frac{2|\Omega_{sr} \cap \Omega_{gt}|}{|\Omega_{sr}| + |\Omega_{gt}|}, \\ HD &= \max\{\sup_s d(v_{sr}(s), v_{gt}), \sup_s d(v_{gt}(s), v_{sr})\}, \\ MAD &= \frac{\int_0^1 d(v_{sr}(s), v_{gt}) |v'_{sr}(s)| ds}{2|v_{sr}|} \\ &\quad + \frac{\int_0^1 d(v_{gt}(s), v_{sr}) |v'_{gt}(s)| ds}{2|v_{gt}|}, \end{aligned} \quad (14)$$

where  $d(v_{sr}(s), v_{gt})$  denotes the minimum distance from point  $s$  to the contour  $v_{gt}$ ,  $\sup$  means the supremum, and  $|v_{sr}|$  represents the length of  $v_{sr}$ . A large  $DSC$ , or a small  $HD/MAD$  indicates high segmentation accuracy. Table III displays the  $DSC$ ,  $HD$  and  $MAD$  values using MS [77], ISO [78], (SUP) [79], MWS, GCC [10], RLS [24], and the proposed. The 80% column represents the sorted 80% highest accuracy among all the results. It is clear that the proposed method provides the best performance, especially in terms of  $HD$  that calculates the largest error for each segmentation. This is attributed to the fact that compared with the other methods without shape constraints, our approach can correct the corrupted nucleus boundaries. Meanwhile, the proposed method produces the smallest standard deviation of the metrics, which indicates its stronger reliability. We also perform a two-sample t-test [81] between the proposed approach and each of the comparative

methods. Except that our method is statistically equal to RLS [24] in terms of  $MAD$  on the breast cancer dataset ( $p$ -value is approximately 0.76), it is significantly better than all the comparative methods in all the other cases ( $p$ -value is smaller than 0.05). The average running time for the proposed method is approximately 54 seconds on an image of about  $220 \times 230$  pixels, where shape initialization accounts for 49 seconds, and 5 seconds for shape deformation and inference. We have applied a fast scanning technique [82] to the CNN prediction in the testing stage, and significantly improved the time cost, with about 6 seconds for pixel-wise classification.

### C. Parameter Analysis

1) *Repulsion Weight*: The parameter  $\omega$  in (12) controls the penalty of the repulsion between contours of touching or overlapping nuclei. Fig. 11 shows the  $DSC$ ,  $HD$ , and  $MAD$  segmentation accuracy in the brain tumor, NET, and breast cancer datasets. When  $\omega = 0$ , the algorithm reduces to a traditional parametric region-based model, and each contour evolves independently such that adjacent contours might cross each other. In this way, touching or overlapping nuclei would not be correctly separated such that the segmentation accuracy will be very low. With the increase of  $\omega$ , the accuracy grows gradually. However, when  $\omega$  is sufficiently large, the repulsion receives much larger penalty such that the non-overlapping rule would be strictly obeyed and the segmentation for overlapping nuclei might not preserve correct shapes and sizes, which are important for subsequent nucleus-level morphology feature extraction. This is more obvious in the breast cancer dataset, which exhibits more touching or overlapping nuclei. Therefore, it is

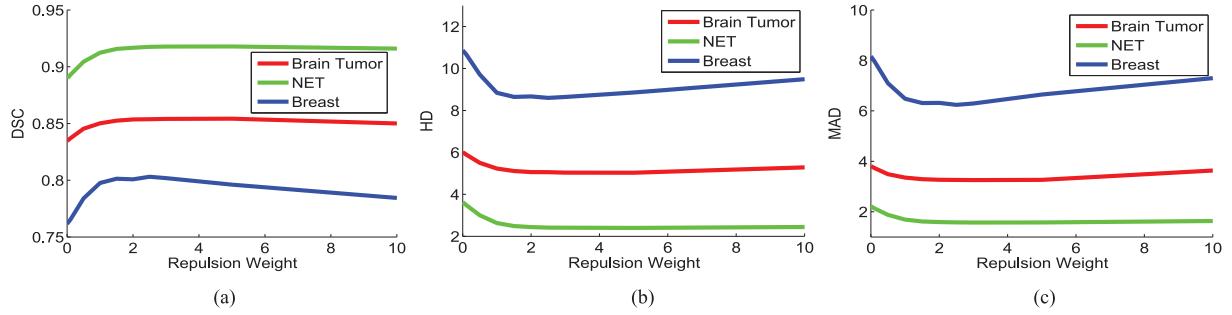


Fig. 11. The performance of the proposed segmentation algorithm with respect to the repulsion weight  $\omega$  in (12).

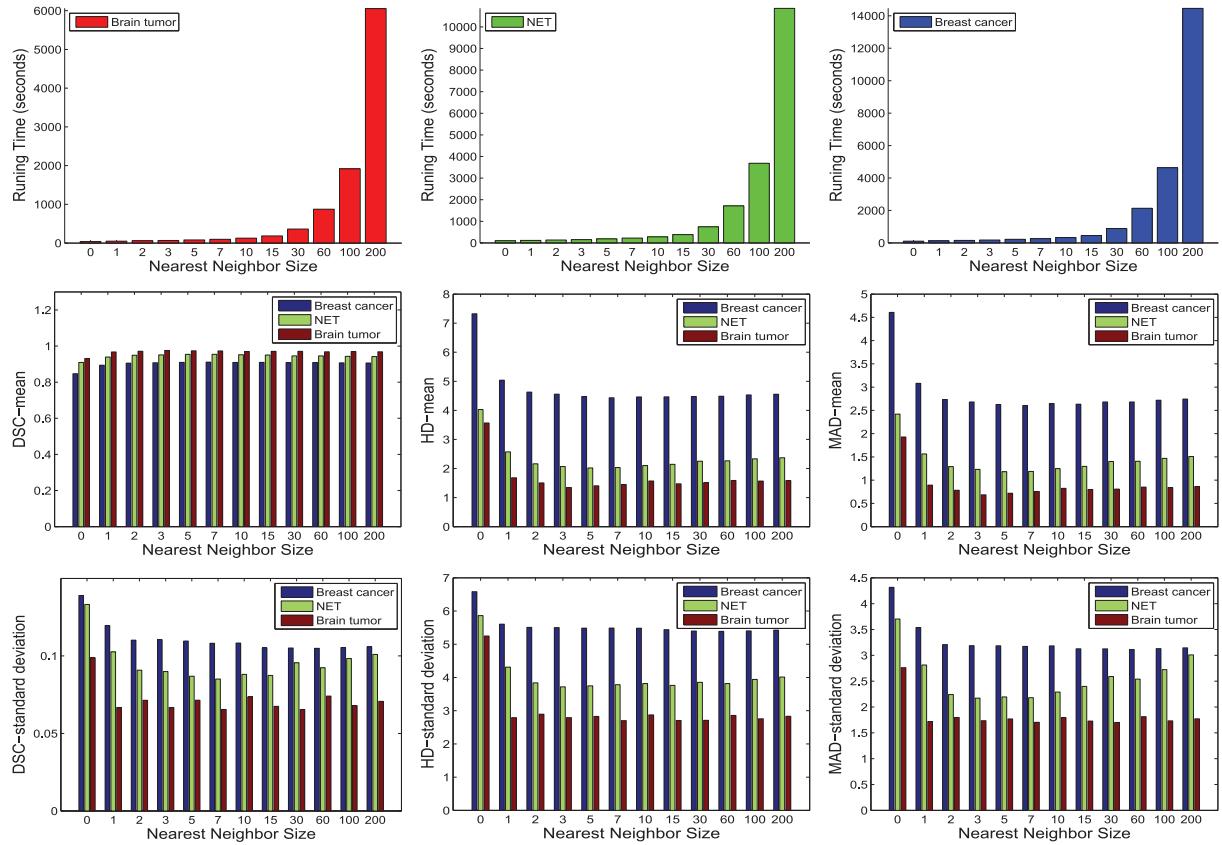


Fig. 12. The performance of the proposed segmentation algorithm with respect to the number of nearest neighbors  $M$  in (12). From top to bottom: running time, mean of the DSC, HD, and MAD, and standard deviation of the DSC, HD, and MAD on brain tumor, NET, breast cancer datasets.

necessary to choose a proper value for  $\omega$ , such as 2.5 (used in our experiments), to achieve desired segmentation accuracy.

2) *Nearest Neighbor Size*: The nearest neighbor size  $M$  in (12) plays a significant role in our algorithm. In Fig. 12, We calculate the segmentation accuracy and running time with different nearest neighbor size  $\{M = 0, 1, 2, 4, 6, 8, 10, 15, 30, 60, 100, 200\}$  on three images, one per dataset. The image size for brain tumor, NET, and breast cancer is  $1121 \times 601$ ,  $1324 \times 817$ , and  $1712 \times 952$  with  $N = 221, 572, 565$  nuclei, respectively. The  $M = 0$  corresponds to no repulsive term considered in (12). The accuracy is low due to the fact that touching or overlapping nuclei are not correctly segmented. The execution time is extremely high when  $M = N - 1$  (6055, 10855, and 14460 seconds on brain tumor, NET, and breast cancer, respectively),

which represents the case that the repulsion from all other nuclei in the image are calculated. When  $M = 5$ , also used in our experiments, the execution time is about 81, 187 and 221 seconds, respectively, which is more than 60 times faster than the one with globally repulsive deformable model. The proposed method in (12) is a regional-based model, so more image background that might contain noises will contribute to contour evolution when a large  $M$  (which usually corresponds to a large region) is used. By contrast, a small local region can preserve intensity smoothness. Therefore, the segmentation accuracy slightly decreases when  $M$  is larger than a threshold, as shown in Fig. 12. Practically, one can select a proper  $M$  by providing a rough estimation to improve both segmentation performance and computational efficiency instead of using  $M = N - 1$ .

TABLE IV  
EFFECTS OF DICTIONARY SIZE ON BRAIN TUMOR, NET, AND BREAST CANCER. FOR EACH METRIC (DSC, HD, MAD), THE MEAN AND STANDARD DEVIATION (STD) ARE LISTED

	Brain tumor (Mean/STD)			NET (Mean/STD)			Breast cancer (Mean/STD)		
	DSC	HD	MAD	DSC	HD	MAD	DSC	HD	MAD
5%	0.85/0.11	5.15/5.40	3.28/2.91	0.92/0.09	2.47/3.09	1.60/1.79	0.80/0.15	8.67/6.85	6.24/4.88
10%	0.85/0.11	5.15/5.42	3.29/2.94	0.92/0.09	2.46/3.08	1.59/1.78	0.80/0.15	8.67/6.87	6.27/4.92
20%	0.85/0.11	5.02/5.27	3.33/2.89	0.92/0.09	2.44/3.00	1.59/1.77	0.80/0.15	8.62/6.76	6.26/4.93
30%	0.85/0.11	5.14/5.39	3.21/2.81	0.92/0.10	2.48/3.09	1.60/1.87	0.80/0.15	8.70/6.83	6.28/4.94
50%	0.85/0.11	5.12/5.39	3.21/2.81	0.92/0.09	2.46/3.06	1.59/1.84	0.80/0.15	8.69/6.82	6.28/4.95
100%	0.85/0.11	4.99/5.06	3.28/2.81	0.92/0.09	2.38/2.89	1.59/1.74	0.80/0.15	8.59/6.72	6.30/4.93

3) *Dictionary Size:* To evaluate the proposed dictionary learning algorithm, we learn multiple dictionaries with different sizes as 5%, 10%, 20%, 30%, 50%, and 100% of the number of all training shapes. The 100% means that the dictionary is equal to the data matrix which collects all the training samples. The evaluation based on *DSC*, *HD*, and *MAD* is listed in Table IV. As we can see, a learned compact dictionary can achieve approximately equal segmentation accuracy as all the data. We also conduct a two sample t-test [81] between different level learned dictionaries and the full dictionary, and it shows that a compact dictionary can consistently provide statistically equal performance as all the data. This might be because the proposed dictionary learning algorithm is sufficient to create a reference repository with strong expressive power. Due to the fast locality-constrained linear coding, the time cost at run-time does not exhibit large variation with respect to the dictionary size.

## V. DISCUSSION AND CONCLUSION

In this work, we propose a novel nucleus segmentation framework using deep convolutional neural network and selection-based sparse shape model. The approach starts with a deep learning-based iterative region merging algorithm to initialize the contours, and thereafter alternately performs efficient bottom-up shape deformation and robust top-down shape inference to achieve correct nucleus segmentation. The comparative experiments demonstrate its superior performance. For fair comparison, all the source (or binary) codes of the other state of the arts, except the marker-controlled watershed which is implemented by us, are obtained from the corresponding authors.

The proposed shape initialization method is robust to image noise and inhomogeneous intensity. Due to the learned feature representation from the CNN model, we can obtain reliable probability maps which can facilitate the subsequent shape inference. Another major contribution in this paper is a novel repulsive deformable model with sparsity-based shape constraint. On one hand, compared with the nonparametric model, level set, which is generally designed to handle topology changes, the proposed repulsive model is a parametric model that can strictly preserve object topology. On the other hand, the learned shape dictionaries constrain the deformable model with shape priors such that it can effectively handle weak or missing nucleus boundaries. The proposed framework is a general method that can be extended to other applications.

The sparse representation-based shape prior model has been experimentally proved to be superior to the conventional PCA-

based shape prior on medical image segmentation [49], [50]. A compact learned dictionary whose size is much smaller than the number of available training samples not only can provide equal performance in nucleus segmentation but also might improve the computational complexity due to the reduced data size.

We also observe that for repulsive deformable models including level set-based implementation [23], [24] or parametric representation [51], it is reasonable and necessary to perform shape deformation in its local system for each individual contour when calculating the repulsion, because each object is only surrounded by a limited number of nuclei, which is much smaller than the total number of nuclei in the entire image. In this way, significant improvement of the computational time can be achieved if we choose a smaller and proper size of the neighbors.

Since currently the proposed approach is mainly implemented with Matlab, it might be not well scalable to large-scale images, and a C/C++ implementation will improve the computational cost. Furthermore, considering that whole-slide scanned histopathological images are usually with very large sizes (e.g.,  $10000 \times 10000$ ), we plan to further reduce the running time of our method using cloud computing techniques. By dividing the whole image into multiple partially-overlapped tiles and distributing them onto different workers, concurrent cell segmentation can be achieved using a master-worker manner in the Spark cloud computing platform [83]. Our future work is to implement the proposed method with cloud computing techniques so that it can be adaptive to large-scale images.

## 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] M. N. Gurcan *et al.*, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [3] G. Lin *et al.*, "A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks," *Cytometry A*, vol. 56, no. 1, pp. 23–36, 2003.
- [4] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst.*, vol. 53, no. 11, pp. 2405–2414, 2006.
- [5] X. Zhou, K.-Y. Liu, P. Bradley, N. Perrimon, and S. T. C. Wong, "Towards automated cellular image segmentation for RNAi genome-wide screening," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2005, vol. 3749, pp. 885–892.
- [6] 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.
- [7] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer-Verlag, 1999.

- [8] O. Schmitt and M. Hasse, "Radial symmetries based decomposition of cell clusters in binary and gray level images," *Pattern Recognit.*, vol. 41, no. 6, pp. 1905–1923, 2008.
- [9] K. Z. Mao, P. Zhao, and P. H. Tan, "Supervised learning-based cell image segmentation for P53 immunohistochemistry," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1153–1163, Jun. 2006.
- [10] 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.
- [11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [12] E. Bernardis and S. X. Yu, "Finding dots: Segmentation as popping out regions from boundaries," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 199–206.
- [13] L. Yang, O. Tuzel, P. Meer, and D. J. Foran, "Automatic image analysis of histopathology specimens using concave vertex graph," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2008, vol. 5241, pp. 833–841.
- [14] Q. Wen, H. Chang, and B. Parvin, "A Delaunay triangulation approach for segmenting clumps of nuclei," in *IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2009, pp. 9–12.
- [15] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1661–1677, Sep. 2011.
- [16] Z. Wu, D. Gurari, J. Y. Wong, and M. Betke, "Hierarchical partial matching and segmentation of interacting cells," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2012, vol. 7510, pp. 389–396.
- [17] T. R. Jones, A. Carpenter, and P. Golland, "Voronoi-based segmentation of cell on image manifolds," *Comput. Vis. Biomed. Image Appl.*, vol. 3765, pp. 535–543, 2005.
- [18] H. Fatakdawala *et al.*, "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.
- [19] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.
- [20] H. Chang, J. Han, P. T. Spellman, and B. Parvin, "Multireference level set for the characterization of nuclear morphology in glioblastoma multiforme," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3460–3467, Dec. 2012.
- [21] D. Padfield, J. Rittscher, N. Thomas, and B. Roysam, "Spatio-temporal cell cycle phase analysis using level sets and fast marching methods," *Med. Image Anal.*, vol. 13, no. 1, pp. 143–155, 2009.
- [22] D. P. Mukherjee, N. Ray, and S. T. Acton, "Level set analysis for leukocyte detection and tracking," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 562–572, Apr. 2004.
- [23] 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, 2008.
- [24] X. Qi, F. Xing, D. J. Foran, and L. Yang, "Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 754–765, Mar. 2012.
- [25] K. Mosaliganti, A. Gelas, A. Gouaillard, R. Noche, N. Obholzer, and S. Megason, "Detection of spatially correlated objects in 3D images using appearance models and coupled active contours," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2009, vol. 5762, pp. 641–648.
- [26] A. Dufour *et al.*, "Segmenting and tracking fluorescent cells in dynamic 3D microscopy with coupled active surfaces," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1396–1410, 2005.
- [27] S. K. Nath, K. Rpalaniappan, and F. Bunyak, "Cell segmentation using coupled level sets and graph-vertex coloring," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2006, vol. 1, pp. 101–108.
- [28] J. P. Bergeest and K. Rohr, "Fast globally optimal segmentation of cells in fluorescence microscopy images," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2011, vol. 6891, pp. 645–652.
- [29] O. Dzyubachyk, W. A. Van Cappellen, J. Essers, W. J. Niessen, and E. Meijering, "Advanced level-set-based cell tracking in time-lapse fluorescence microscopy," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 852–867, Mar. 2010.
- [30] C. Zimmer and J.-C. Olivo-Marin, "Coupled parametric active contours," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1838–1842, Nov. 2005.
- [31] X. Han, C. Xu, and J. L. Prince, "A topology preserving level set method for geometric deformable models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 755–768, Jun. 2003.
- [32] C. L. Guyader and L. A. Vese, "Self-repelling snakes for topology-preserving segmentation models," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 767–779, May 2008.
- [33] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, 1998.
- [34] L. Yang, P. Meer, and D. J. Foran, "Unsupervised segmentation based on robust estimation and color active contour models," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 3, pp. 475–486, 2005.
- [35] H. Cai *et al.*, "Repulsive force based snake model to segment and track neuronal axons in 3D microscopy image stacks," *NeuroImage*, vol. 32, no. 4, pp. 1608–1620, 2006.
- [36] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, pp. 321–331, 1988.
- [37] H. Chang *et al.*, "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670–682, Apr. 2013.
- [38] S. Kothari *et al.*, "Automatic batch-invariant color segmentation of histological cancer images," in *IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2011, pp. 657–660.
- [39] M. Rousson and N. Paragios, "Shape priors for level set representations," in *Europ. Conf. Comput. Vis. (ECCV)*, 2002, pp. 78–92.
- [40] T. Chan and W. Zhu, "Level set based shape prior segmentation," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, vol. 2, pp. 1164–1170.
- [41] M. Leveton, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2000, vol. 1, pp. 316–323.
- [42] X. Bresson, P. Vandergheynst, and J. Thiran, "A priori information in image segmentation: Energy functional based on shape statistical model and image information," in *Int. Conf. Image Process. (ICIP)*, 2003, vol. 3, p. III-425-8.
- [43] D. Cremers, N. Sochen, and C. Schnorr, "Towards recognition-based variational segmentation using shape priors and dynamic labeling," in *Int. Conf. Scale Space Theories Comput. Vis.*, 2003, vol. 2695, pp. 388–400.
- [44] N. Paragios, M. Rousson, and V. Ramesh, "Matching distance functions: A shape-to-area variational approach for global-local registration," in *Europ. Conf. Comput. Vis. (ECCV)*, 2002, pp. 775–789.
- [45] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1448–1460, Jul. 2012.
- [46] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [47] H. Cai *et al.*, "Shape-constrained repulsive snake method to segment and track neurons in 3D microscopy images," in *IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2006, pp. 538–541.
- [48] 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.
- [49] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou, "Deformable segmentation via sparse shape representation," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2011, pp. 451–458.
- [50] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse shape representation and dictionary learning," *Med. Image Anal.*, vol. 16, no. 7, pp. 1385–1396, 2012.
- [51] F. Xing and L. Yang, "Robust selection-based sparse shape model for lung cancer image segmentation," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2013, vol. 8151, pp. 404–412.
- [52] H. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2013, pp. 583–590.
- [53] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 2852–2860.

- [54] A. Cruz-Roa, J. A. Ovalle, A. Madabhushi, and F. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2013, vol. 8150, pp. 403–410.
- [55] A. Prasoon *et al.*, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2013, vol. 8150, pp. 246–253.
- [56] B. Hariharan, P. Arbelaez, and R. Girshick, "Hypercolumns for object segmentation and fine-grained localization," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 447–456.
- [57] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2013, vol. 8150, pp. 411–418.
- [58] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *IEEE Int. Sym. Cir. System. (ISCAS)*, 2010, pp. 253–256.
- [59] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inform. Processing Syst. (NIPS)*, 2012, pp. 1106–1114.
- [61] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [62] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
- [63] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv:1207.0580, cs.NE*, pp. 1–18, 2012.
- [64] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 160–167.
- [65] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, pp. 291–299, 2001.
- [66] C. R. J. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 265–270, Feb. 2003.
- [67] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [68] D. W. Scott, "Parametric statistical modeling by minimum integrated squared error," *Technometrics*, vol. 43, pp. 274–285, 2001.
- [69] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 3360–3367.
- [70] Y. Shi and D. Shen, "Hierarchical shape statistical model for segmentation of lung fields in chest radiographs," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2008, vol. 5241, pp. 417–424.
- [71] Y. Zhan, M. Dewan, and X. S. Zhou, "Cross modality deformable segmentation using hierarchical clustering and learning," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2009, pp. 1033–1041.
- [72] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [73] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [74] N. Djuric, L. Lan, S. Vucetic, and Z. Wang, "BudgetedSVM: A toolbox for scalable SVM approximations," *J. Mach. Learn. Res.*, vol. 14, pp. 3813–3817, 2013.
- [75] P. Dollár, Piotr's Computer Vision Matlab Toolbox (PMT) [Online]. Available: <http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html>
- [76] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [77] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [78] L. Grady and E. L. Schwartz, "Isoperimetric graph partitioning for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 469–475, Mar. 2006.
- [79] G. Mori, "Guiding model search using segmentation," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, vol. 2, pp. 1417–1423.
- [80] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect cells using non-overlapping extremal regions," in *Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2012, vol. 7510, pp. 348–356.
- [81] J. H. McDonald, *Handbook of Biological Statistics*. Baltimore, MD: Sparky House, 2014.
- [82] A. Giusti, D. C. Ciresan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *IEEE Int. Conf. Image Process. (ICIP)*, 2013, pp. 4034–4038.
- [83] M. Sapkota, F. Xing, F. Liu, and L. Yang, "Skeletal muscle cell segmentation using distributed convolutional neural network," in *High Performance Comput. (HPC) Workshop Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2015.