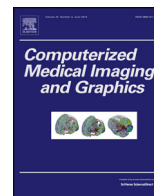




Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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



A two-layer structure prediction framework for microscopy cell detection

Yan Xu^{a,b}, Weiying Wu^a, Eric I-Chao Chang^b, Danny Chen^c, Jian Mu^c,
Peter P. Lee^d, Kim R.M. Blenman^d, Zhuowen Tu^{e,f,*}

^a State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beihang University, China

^b Microsoft Research, Beijing, China

^c Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, USA

^d Cancer Immunotherapeutics and Tumor Immunology (CITI), Duarte, CA, USA

^e Department of Cognitive Science, University of California, San Diego, CA, USA

^f Department of Computer and Engineering, University of California, San Diego, CA, USA

ARTICLE INFO

Article history:

Received 25 January 2014

Received in revised form 29 June 2014

Accepted 4 July 2014

Keywords:

Microscopy cell detection

Layered models

Structural prediction

Computer vision

Machine learning

ABSTRACT

The task of microscopy cell detection is of great biological and clinical importance. However, existing algorithms for microscopy cell detection usually ignore the large variations of cells and only focus on the shape feature/descriptor design. Here we propose a new two-layer model for cell centre detection by a two-layer structure prediction framework, which is respectively built on classification for the cell centres implicitly using rich appearances and contextual information and explicit structural information for the cells. Experimental results demonstrate the efficiency and effectiveness of the proposed method over competing state-of-the-art methods, providing a viable alternative for microscopy cell detection.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

With the development of bioimage informatics, there has been an increasing focus on the cell detection studies, which come over the problem of the traditional microscopy studies that only a few cells are detected, providing a more powerful statistical tool for experiments and medical diagnosis [1,2]. Studying the amount, types, morphological shape, and configuration of the cells in microscopic images has significant biological and clinical significance in e.g. cancer diagnosis and cancer treatment [3–7]. Performing cell segmentation is mostly based on certain families of morphological descriptors, which can also be viewed as a structural prediction problem. For example, the auto-context algorithm was adopted to perform membrane segmentation with supervised

(machine learning) method, see [8]. One could also follow the procedure for object detection as in the recent computer vision literature to perform cell detection by a layout processing step [9]. However, as we will see later in the experiment results, adopting the state-of-the-art vision approaches does not provide satisfactory results in the task of microscopic cell centre detection. Since the Conditional Random Fields (CRFs) [10] model can only detect the structure of cells from one layer, thus may not satisfactorily grasp the full knowledge representation and achieve the requested detection accuracy due to the ignorance of complex information of the cell with much subtle and variable structure. The examples like derivative formulas of the CRFs model, or auto-context, can not achieve the approving performance of the serious demanding for complicated cell detection goal with the ignorance of various size, colour, deformation, and structural dependence more or less. Here, we study the problem from a structured learning perspective and emphasize on the representation issue.

Existing approaches for microscopy cell detection consist of the segmentation and detection steps, which are usually based on geodesic active contour, watershed, gradient vector diffusion or Hessian matrix [5–7,11,12]. In many cases, some of the above-mentioned methods are combined for a better performance of microscopic cell detection. One of the typical examples is using the

* Corresponding author at: State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beihang University, Microsoft Research, Beijing, China.
Tel.: +86 10 82338755; fax: +86 10 82338755.

E-mail addresses: xuyan04@gmail.com (Y. Xu), wuweiying18@gmail.com (W. Wu), echang@microsoft.com (E.I.-C. Chang), dchen@cse.nd.edu (D. Chen), jmu@nd.edu (J. Mu), plee@coh.org (P.P. Lee), kblenman@cinci.rr.com (K.R.M. Blenman), zhuowen.tu@gmail.com (Z. Tu).

ridge measures for cell shape detection, which is widely applied more or less in the field of image segmentation. Ersoy et al. combined the shape feature of cells with the curve evolution model exploiting the halo effect to perform the human cancer cell image detection, which was based on Hessian matrix and directional derivatives [7]. Yeo developed automated image segmentation and classification for separation of clusters of cell nuclei into individual objects, mixing the adaptive two pass threshold-based segmentation, clustered nuclei using segmentation (including operations of the fill hole, noise removal, distance, adaptive h -minima transform incorporated with edge information and watershed transform) and the supervised process with the labelling ground truth for shape descriptors, emphasizing the accuracy of cell nuclei counts from the supervised process [11]. The evaluation of the performance by Smal et al. for the purpose of quantitative analysis of biological image data gave a detailed comparison utilizing seven unsupervised methods and two supervised methods with the same ground truth given by the experts, effectively emphasizing the high accuracy of the machine learning methods [5].

Typically, structured labelling studies the problem of predicting a vector output from an input vector, which is common since structured inputs and outputs are in a wide range of applications. There are many different approaches in which such kind of application with structured labelling has been applied. For example, Markov Random Field (MRF) and Conditional Random Fields (CRFs) [10] have been widely used to model the correlations of the structured labels. However, due to the heavy computational burdens in their training and testing (inference) stages, MRF and CRFs are usually limited to capturing a few neighbourhood interactions only, which will thus limit their modelling capabilities. Structural support vector machine (SVM) [13–15] models the correlation in a way similar to the CRFs but tries to specifically maximize the prediction margin. Nevertheless, the high computational demand just makes it firmly rely on the range of contexts.

Considering of the limitations of the above approaches, we propose a new two-layer structural prediction framework for detecting microscopic image cells, which focuses on the representation issue to propose a model to effectively capture the rich contextual information. The contribution of the approach includes: (1) we tackle the problem of microscopic image cell detection from structure learning perspective; (2) the second layer takes the output of the first layer as knowledge abstraction and propagation.

2. Related work

With the wide application of the biomedical cell detection, various methods have been applied for cell segmentation, which is aiming to segment the objects into their constituent objects and a background [12,6,16–23]. Therefore, the cell segmentation problem can be very difficult, depended by the type of the given specimen. For example, in [18], the aim of segmentation was to distinguish the cell nuclei with a positive staining reaction and other cell nuclei in the field of immunohistochemistry, with a robust classifier and built-in metric corresponding to the colour space. Except the cell nuclei segmentation, there have also been many segmentation algorithms for other cell structures. The paper [24] by Nguyen et al. proposed a rapid cell detection algorithm, providing the Adaptive Boosting baseline as cell candidate rules, optimizing the global parameters from the input image by integration, which had highly accelerated the training process of machine learning for cell detection. Yang et al. [21] utilized the level set methods to acquire the extended-time live cell image, which performed the cell cluster separation and mitotic cell detection steps after identifying the cell trajectories, providing a method to solve the cell analysis problem, including the detection of numbers, locations, borders, areas and

states of cells. In [6], the hierarchical model was combined with three operators of edge detection (Sobel, Prewitt and Laplace) to perform the multiple cell image segmentation, in addition, the false removal algorithm was used to increase the accuracy of the cell detection.

In our task, the first layer takes the auto-context algorithm, basing on the detection of the contextual information from the whole image, which is in widespread use in the object recognition [8,25–28]. Jurrus et al. [8] adopted the idea of structural labelling with a neuro-network implementation to perform membrane segmentation, which meant the ignorance of the complex information of the cells with only one layer of cell. The system developed by Yang et al. [28] aimed to perform multi-class segmentation to define shape priors from the output of object detectors.

There have been various graphical models for image segmentation. The Markov Random Field (MRF) is a typical one with a set of random variables including a Markov property [29]. Following the Markov property, the Conditional Random Fields (CRFs) directly generates the posteriori probability distribution model of the random variables from the given observation image input [30–33], which has been widely used in many image segmentation problems. For example, in [32], Bauer et al. combined the powerful SVM based on the features of multispectral intensities and textures with subsequent hierarchical CRFs regularization to delineate the brain tumour boundaries. The CRFs model was also used for the prostate cancer localization in the multispectral MRI image segmentation when combined with a cost-sensitive framework in [33], which boosted the performance significantly. Our algorithm utilizes a probability model Event Detection Conditional Random Field (EDCRF), which is related to the work of mitosis detection [34] with the use of phase-contrast time-lapse microscopy. However, the input for the second layer they use is a hidden layer in the model whereas our second layer also looks back at the original data.

The jump-diffusion process provides an optional method for our second layer algorithm, which has been an effective method for a range image segmentation [35–37]. The paper [35] focused on the segmentation of the objects of various types and sizes in the complex-real world scenes, which jumped between the parameter subspaces of differing dimensionality, providing a good performance for the object segmentation.

3. Method

This section attempts to give the details of our two-layer model for cell centre detection/segmentation in microscopic images.

We describe cell detection as a structural prediction problem [38–40], which tries to introduce a two-layer structural framework for detecting microscopy cells, with the two probability models to approximate the true cell locations. In the first layer, the auto-context model is performed to capture the structural information of the cells, passing on the results combined with the original cell images as the input for the second layer for segmentation. In addition, some related algorithms are briefly discussed for comparison.

3.1. A Conditional Random Fields approach for the problem

The Conditional Random Fields (CRFs) have been applied a lot in machine learning, with a classifier trained from the given sets X and Y , which are respectively the observed and output variables, to predict a classifier from a given sample for structured prediction [10]. The conditional distribution is an undirected graphical model related to X and Y , which is denoted by $p(Y|X)$ and modelled from them.

We derive the cell centre classification results from the first layer using the seminal CRFs models (such as auto-context [41]), whereas

Table 1
The evaluation of the various methods.

	Precision	Accuracy	F-measure	DisErr1,px	DisErr2,px
Classification model [46]		0.8734			
Direct regression				2.97	
Object detection + layout [41]	0.7693	0.9281	0.8413	2.49	2.21
First-layer (auto-context)	0.8031	0.9457	0.8686	2.37	2.12
Auto-context + NMS [47,9]	0.8912	0.9007	0.8959	2.31	2.01
Two-layer method with belief propagation	0.9279	0.9255	0.9267	2.20	1.91
Two-layer method with jump-diffusion	0.9465	0.9453	0.9459	1.91	1.51

The use of bold values is to highlight the accuracy of our method.

the results imply it with not satisfied performance of dealing with the complexity in the structural cell images (Table 1).

3.2. Formulation

In the two-layer structural model, a set of training images with the corresponding ground truth label maps as $S = \{(\mathbf{x}_k, \mathbf{l}_k), k = 1, \dots, K\}$ are given for training. Considering of denotation simplicity, we consider with a view to one training image: for each image, we rigorously define the data by $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$, where x_i denotes a patch (pixel i is the centre of the patch) with the corresponding ground truth labels $\mathbf{l} = \{l_1, l_2, \dots, l_m\}$, $l_i \in \{+1, -1\}$, among which $l_i = +1$ indicates pixel i being on the centre of a cell and $l_i = -1$ to the contrary. So we can directly get the random variables as $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$, $y_i \in \{+1, -1\}$ (Fig. 1).

3.2.1. First layer

In the first layer, we aim at learning a discriminative function to minimize the training error as

$$e = \sum_i \mathbf{1}(l_i \neq F_i(\mathbf{x})), \quad (1)$$

where $\mathbf{1}(\cdot)$ is an indicator function and $F_i(\mathbf{x})$ is a prediction on the label y_i . A direct patch-based classification throws away the other patches to have $F_i(\mathbf{x}) = \text{sign}\left(\sum_{t=1}^T \alpha_{i,t} \cdot h_{i,t}(\mathbf{x})\right)$ using e.g. as a boosting approach. Considering the rich contextual information, the auto-context algorithm [41] further learns a recursive

function as

$$\hat{p}^{(t)}(y_i|\mathbf{x}) \propto \exp \left\{ -y_i \sum_{t=1}^T \alpha_{i,t} \cdot h_{i,t}(\mathbf{x}, \mathbf{p}^{(t-1)}(\mathbf{y})) \right\}$$

for $F_i(\mathbf{x}) = \text{sign}(\hat{p}^{(t)}(y_i = 1|\mathbf{x}) - 0.5)$, (2)

where $\mathbf{p}^{(t-1)}$ is the classification map. The auto-context algorithm implicitly fuses a large amount of structural information but reports unsatisfactory results (shown in Fig. 2(d)), since the cell centre detection problem is not only a segmentation task and the auto-context almost only uses the implicit information.

3.2.2. Second layer

Therefore, we propose a hierarchical representation in this paper by propagating the cell centre detection information from the first layer into a further process. In the second layer, we use $\hat{p}(y_i|\mathbf{x}) = \mathbf{p}^{(t)}$ (abbreviated as $\hat{\mathbf{p}}$) learned from the first layer as augmented information and apply an explicit model as to minimize the same error function in Eq. (1):

$$e = \sum_i \mathbf{1}(l_i \neq y_i^*), \quad (3)$$

where $\mathbf{y}^* = \{y_1^*, y_2^*, \dots, y_m^*\}$, and can be obtained by:

$$\mathbf{y}^* = \arg \max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}, \mathbf{p}; \theta) = \frac{e^{\psi(\mathbf{y}, \mathbf{x}; \theta_1)} e^{\psi(\mathbf{y}, \hat{\mathbf{p}}; \theta_2)}}{Z(\mathbf{x}, \mathbf{p})}, \quad (4)$$

in which $Z(\mathbf{x}, \mathbf{p})$ is the partition (normalization) function as

$$Z(\mathbf{x}, \mathbf{p}) = \sum_{\mathbf{y}'} [e^{\psi(\mathbf{y}', \mathbf{x}; \theta_1)} e^{\psi(\mathbf{y}', \hat{\mathbf{p}}; \theta_2)}] \quad (5)$$

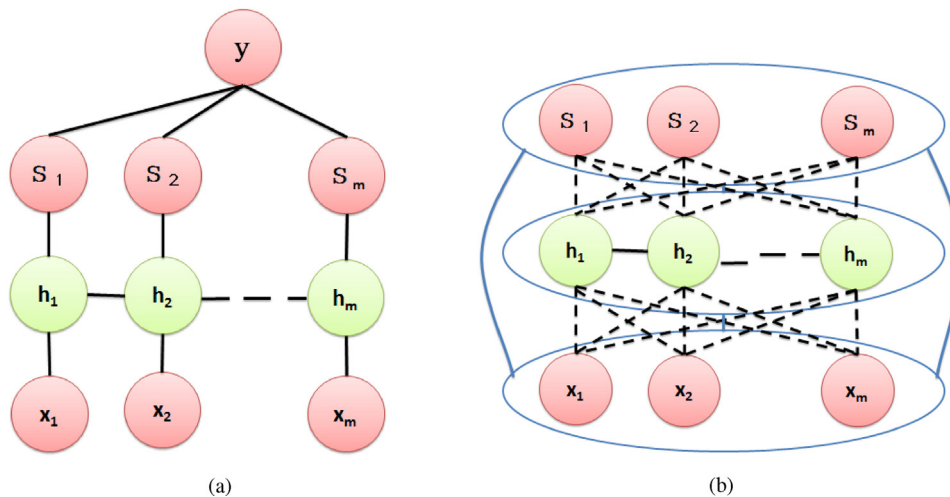


Fig. 1. Graphical representations of an existing hidden CRF approach [42] and our proposed two-layer model. (a) x_i and h_i represent the i th observation and the hidden state assigned on x_i . y represents a class label. (b) Our two-layer model. x_i and $\hat{p}_i$ respectively represent the i th observation and context-based discriminative model. y_i represents the class label for pixel i .

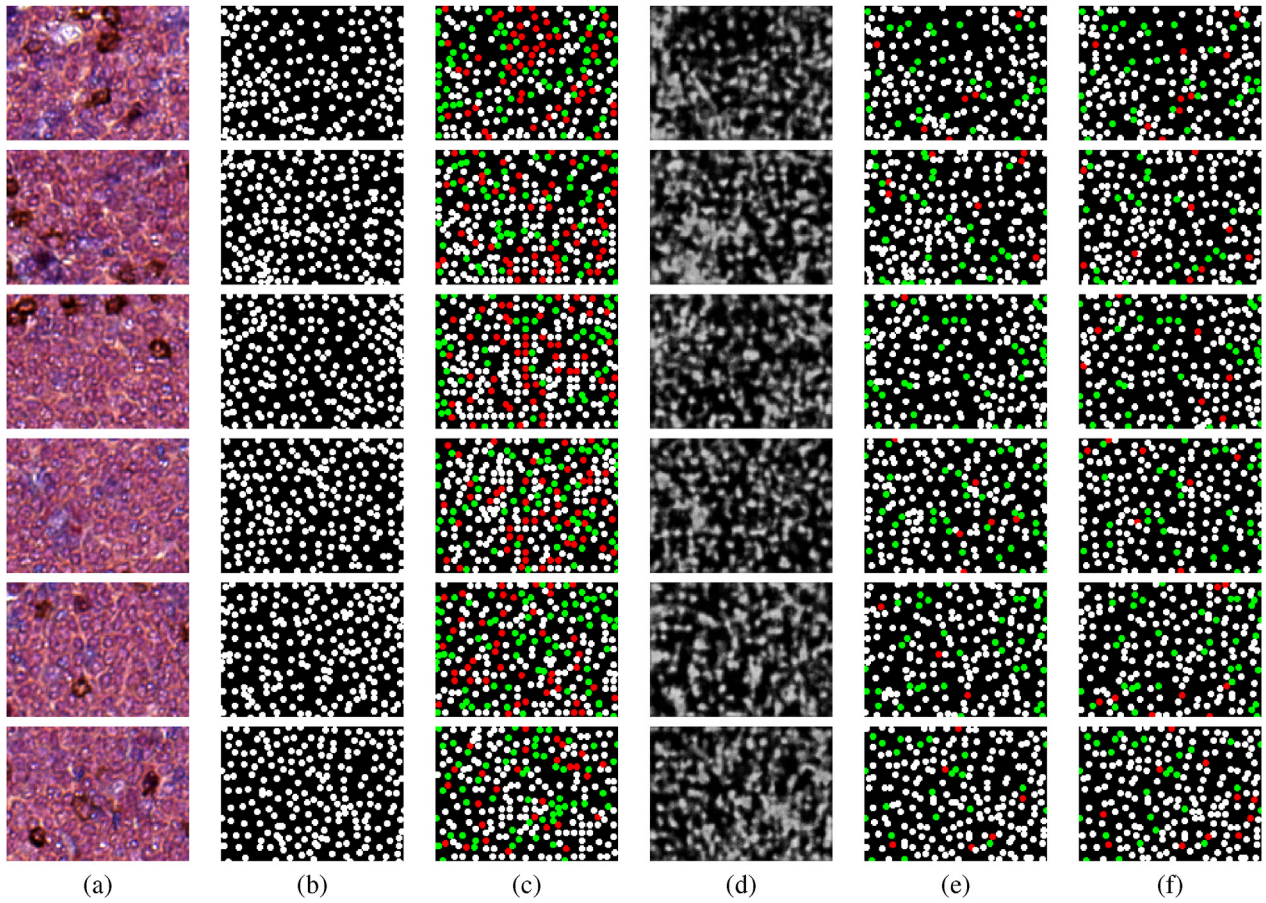


Fig. 2. Comparisons: (a) The original images. (b) The ground truth. (c) Results by object detection + layout. (d) Results by the auto-context. (e) Results of our method with belief propagation. (f) Results of our method with jump-diffusion. Circle dot: cell centre. Red: false detection; green: missed detection; white: true detection. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

In addition, we have

$$\psi(y, \mathbf{x}; \theta_1) = \sum_{j=1}^m \sum_l f_{1l}^1(j, y_j, \mathbf{x}) \theta_{1l}^1 + \sum_{(j,k) \in E} \sum_l f_{1l}^2(j, k, y_j, y_k, \mathbf{x}) \theta_{1l}^2 \quad (6)$$

and the same for $\psi(y, \mathbf{p}; \theta_2)$:

$$\psi(y, \mathbf{p}; \theta_2) = \sum_{j=1}^m \sum_l f_{2l}^1(j, y_j, \mathbf{p}) \theta_{2l}^1 + \sum_{(j,k) \in E} \sum_l f_{2l}^2(j, k, y_j, y_k, \mathbf{p}) \theta_{2l}^2 \quad (7)$$

where $\theta = \{\theta_1, \theta_2\}$ are the parameters of the model, and f_{1l}^1 and f_{2l}^1 are a state function, while f_{1l}^2 and f_{2l}^2 are a transition function. $\theta_{1l}^1, \theta_{1l}^2, \theta_{2l}^1$, and θ_{2l}^2 are the corresponding parameters of state and transition functions, respectively.

3.3. Model learning and parameter estimation

3.3.1. First layer

The training process in the first layer uses the constructed labels, which is similar to that of the auto-context [41], with the goal of minimizing Eq. (1), thus we can get the training set recursively constructed as:

$$S^{(t)} = \{(y_i, (\mathbf{x}, \mathbf{p}^{(t-1)})), \quad i = 1, \dots, m\} \quad (8)$$

where $\mathbf{p}^{(t-1)}$ is the classification map for each round of the auto-context algorithm by minimizing the error as Eq. (1). The parameters of the above Eq. (8) are the coefficients in $\hat{p}^{(t)}(y_i | \mathbf{p}^{(t-1)}, \mathbf{x})$. For example, if the SVM is adopted, then the parameters are the SVM weights; otherwise, they are the weights and thresholds in the weak classifiers for boosting; and the tree splitting nodes for random forest. The classifier we get from the training process is used to compute $\mathbf{p}^{(t-1)}$ for each training image $\mathbf{x}$, which will generate a sequence of classification probabilities as the outputs, see the auto-context algorithm [41].

3.3.2. Second layer

In the second layer, the parameter values are obtained from the regularized log-likelihood features consistent to Eq. (1).

$$L(\theta) = \sum_{i=1}^n \{\log[e^{\psi(y, \mathbf{x}; \theta_1)} e^{\psi(y, \mathbf{p}; \theta_2)}] - \log Z(\mathbf{x}, \mathbf{p})\} - \frac{1}{2\delta^2} \|\theta\|^2 \quad (9)$$

where δ is the variance of a Gaussian prior, and n is the number of cell images. $\theta = \{\theta_1, \theta_2\}$ can be deduced by the partial derivatives:

$$\frac{\partial L(\theta)}{\partial \theta_{1l}^1} = \sum_{j=1}^m f_{1l}^1(j^{(i)}, y_j^{(i)}, \mathbf{x}^{(i)}) - \sum_{l=1}^k \frac{\theta_{1l}^1}{\delta^2} - \sum_{y'} f_{1l}^1(j^{(i)}, y', \mathbf{x}^{(i)}) p(y' | \mathbf{x}^{(i)}) \quad (10)$$

$$\frac{\partial L(\theta)}{\partial^2_{1l}} = \sum_{(j,k) \in E} f_{1l}^2(j^{(i)}, k^{(i)}, y_j^{(i)}, y_k^{(i)}, \mathbf{x}^{(i)}) - \sum_{l=1}^k \frac{\theta_{1l}^2}{\delta^2} - \sum_{(j,k) \in E, y, y'} f_{1l}^2(j^{(i)}, k^{(i)}, y, y', \mathbf{x}^{(i)}) p(y, y' | \mathbf{x}^{(i)}) \quad (11)$$

$$\frac{\partial L(\theta)}{\partial^2_{2l}} = \sum_{j=1}^m f_{2l}^1(j^{(i)}, y_j^{(i)}, p^{(i)}) - \sum_{l=1}^k \frac{\theta_{2l}^1}{\delta^2} - \sum_{y'} f_{2l}^1(j^{(i)}, y', p^{(i)}) p(y' | p^{(i)}) \quad (12)$$

$$\frac{\partial L(\theta)}{\partial^2_{2l}} = \sum_{(j,k) \in E} f_{2l}^2(j^{(i)}, k^{(i)}, y_j^{(i)}, y_k^{(i)}, p^{(i)}) - \sum_{l=1}^k \frac{\theta_{2l}^2}{\delta^2} - \sum_{(j,k) \in E, y, y'} f_{2l}^2(j^{(i)}, k^{(i)}, y, y', p^{(i)}) p(y, y' | p^{(i)}) \quad (13)$$

3.4. Inferences

In the testing process, the probabilities are computed by using the trained classifiers, assigning each pixel i with the label in Eq. (4) given an image $\mathbf{x}$.

First layer: A sequence of trained classifiers is used to generate a probability map p^T , where T is the number of iteration times. The classification result is obtained by applying a series of classifiers, the same for the auto-context algorithm [41].

Second layer: Here we propose two optimization methods for the inference: (1) belief propagation [39] and (2) jump-diffusion [43]. As belief propagation is a common inference method, we will not elaborate much on this topic. Then, we will give the details of the jump-diffusion algorithm which has been less adopted in structured labelling but with excellent performance in our task here as following.

3.4.1. The auto-context using

As is mentioned above, auto-context is firstly used in the first layer of the cell detection, as it is with the ability of CRFs to support context information with flexible sizes, offering a feasible alternative for propagating knowledge and generating a probability mapping image from the original input with high accuracy; and therefore, can well fit the cell detection with no limit of structure and dimensions, which requires especially accurate demanding for microscopic and variable objects detection.

For the subtle structure of the cells we detect, a change from the use of the traditional auto-context algorithm for the detection of natural image (e.g. the horse) that we should point out is that the levels (the circulation times of the optimization for global parameters in the training process to generate a classifier) of the auto-context we utilize for detecting the cells is changed from 4 levels to 10, which means the time for iterations of the training process may become longer, aiming to achieve the higher computational demanding of the cell feature classification, giving a more suitable descriptors for the generation of classifier. More specifically, the classifier generated based on the connection between the ground truth $Y_i = \{y_{i1}, \dots, y_{in}\}$ and the given image x_i with the use of training set Eq. (8) should be iterated for more times. Also, the Decision Random Tree is relatively changed from 3 to 100 for the same purpose, giving more details to the classifier to judge whether the pixels of the input images are positive or negative.

Except attempting to select context information and the appearance of the cell image in a unified framework for the auto-context algorithm in the first layer, the scale-invariant feature transform (SIFT) algorithm [44] is combined with it to detect the subtle shape. The SIFT algorithm insures the invariance property of rotation and

scale expansion, and is partially invariant to illumination changes and robust to local geometric distortion. We rigorously set the parameters (spacing size and grid size) of three different values, which are (30,2), (60,6) and (100,8).

3.4.2. Basics of jump-diffusion

The stochastic jump-diffusion process [43,45] is used in the second layer of the model. In the process, the solution X lies in a heterogeneous space (mixture spaces of different dimensions), which consists of different dimensions $\Omega = \bigcup_{n=0}^N \Omega^{(n)}$; where n denotes the number of cells and N is the maximum amount of possible cells. Each $\Omega^{(n)}$ is the space of the fixed number n cells. Considering that we do not know the number of cells, a union of different spaces is assumed for confirmation. For n cells, the process to optimize Eq. (4) is considered as a diffusion process; also, switching between the different spaces means to give birth to a cell or kill an existing one, which represents the jumping process between different spaces.

We shall rigorously define two types of processes as following: the first process is a diffusion process, which aims to move a cell centre (x_i, y_i) within limits of a domain D_i ; the second process is a jump process, in which we should create or remove a cell centre (x_i, y_i) .

When the sample $N \times N$ images (N is the labelled cell diameter) are given, we firstly perform the Non-maximal suppression (NMS) to extract the initial seeds as the cell centres (the overlap step is $\text{INT}[2/3 * N]$ pixels). Then, the initial cell centres are reflected to the original images as seeds. After the initialization step above, we perform the jump-diffusion inference to optimize Eq. (4).

4. Experiments

4.1. Data

4.1.1. Datasets

The cell images we detect are the images of stained tissues of tumour-draining lymph nodes from patients with breast cancer, which come from a multi-spectral imaging system. The dataset we use is with 200 images (each 102×139 pixels, 100 images for train data and 100 images for test data), which are randomly picked up from tiny images that are cut from the big ones with size of 1020×1390 pixels (generated from the imaging system), for the convenience of the cell segmentation. The final result is the same for the tiny size, and with higher efficiency. With the 200 images, we finally get the average 151.6 ± 22.1 cells per image. As we can see from the above, the large images with 1020×1390 pixels will have more than 10,000 pixels, thus the tiny ones can achieve higher speed for segmentation. Although tissue images can usually see lots of variation, the problem we focus on is whether the location of the image is cell centre or not, while the variation is insignificant if the whole dataset is with the same protocol staining.

4.1.2. Annotations

We labelled all of the cells in each image. To ensure the accuracy of the ground truth, images were carefully labelled by well-trained microscopy experts. Specifically, each image was first annotated by two experts independently. Then, we overlapped the labels given by them in computer for calibration. If the results given by the two experts were different, a third expert would moderate the discussion until they reached the final agreement.

4.2. Comparison

In order to make a comparison with the two-layer structural model, six methods are designed in our experiment.

Classification model [46]: Liu et al. proposed a random forest based classifier with new feature sets to predict the lymph nodes

in multi-spectral microscopy images. Since the source code of this algorithm is unreachable, we use the result provided in the paper. In addition, the same raw data is used in the paper.

Object detection [47] + layout [9]: A two-step method is used combining the Integral Channel Features detection (ICF) [47] and the Object Layout Discriminative Model (DLDM) [9]. Both of them are state-of-the-art methods in the computer vision field, where ICF is used to detect the cell candidates, and OLDLM is used to decide the final cell locations from the results of ICF.

The settings in ICF are with eight channels, including gradient magnitude, grey scale, and six gradient histogram channels with no-pre or post-smoothing. Also, the classifier is the AdaBoost with 1000 depth-2 decision trees, with 5000 first-order features being randomly generated.

The positive instances are the 100 original colour cell images from the training data, while the negative instances are the windows from the positive images with no cell part, which leads to a total number of negative instances as 12,691, while the range of negative instances' size is from 15×15 pixels to 24×24 pixels.

As for the settings in DLDM, the category is replaced by one (the initial value is 21 categories), where the results of ICF are used as the input. It seems that the algorithm does not perform well as the average number of the candidate cells per ICF cell image is 463.5, while the gold standard is 151.6 ± 22.1 .

Auto-context [41] + non-maximum suppression (NMS): A two-step method is used combining the auto-context algorithm and the non-maxima suppression. In auto-context process, the features are scale-invariant feature transform features (SIFT), HAAR-like features (HAAR), histogram of oriented gradients features (HOG) and LUV colour features (LUV). Based on the classifier built from the Decision Random Tree (100 trees), the post-processing is non-maxima suppression to detect accurate cell centres.

Regression based cell counting [48]: Although the ultimate goal of cell detection is to evaluate the number of cells, we also recognize the task as a regression problem, using Random Forest to predict the number of cells.

In this algorithm, the input is the probability mapping images generated from the auto-context algorithm [41], while the output is the number of cells in each image. While counting the number of cells in an image, we define the number of a cell as the area percentage of this cell in this image. That is, 1 is obtained if the cell is completely within the image; and 0.5 is considered to be fifty percentage included in the image. Each image is densely extracted to 48×48 sub-images as instances, while the overlap step is 16 pixels. We finally get the average cell amount per instance as 24.11 ± 4.38 . 100 trees are chosen in our random forests, with the features of grey histogram, mean and variance from the cell probability mapping images.

Two-layer method with belief propagation: In the first layer of the algorithm, the features we use are SIFT, HAAR, HOG and LUV, while the classifier is built from the Decision Random Tree. In total, 100 trees are used, with 10 pixels structural radius and 200 iterations. In the second layer, belief propagation is used as inference.

Two-layer method with jump-diffusion: The setting in the method is the same as the above method in the first layer, while in the second layer, we downsample 5×5 images with an overlap step of 3 pixels.

5. Results

5.1. Computational cost

We performed the training process in a 64-bit operating system, with Intel(R) Core(TM)2 Quad CPU Q9400 @2.66 GHz 2.67 GHz processor and 8.00 GB RAM, as well as Microsoft online IT support.

Finally, 30 h was consumed to the first layer of the training process for 100 cell images, which contributed a lot to the accuracy of our ultimate result whereas 18 min and 20 min were respectively contributed to the jump-diffusion and belief propagation process of the second layer. For the testing process, 50 min was consumed to the first layer for 100 cell images, while 18 min and 20 min were contributed to the jump-diffusion and belief propagation process, respectively.

5.2. Evaluation method

We define TT (true detection) as the number of the predicted cell centres which are within the labelled cells, TF (missed detection) as the number of labelled cells with no predicted cell centre included, and FT (false detection) as the number of the predicted cell centres which are outside the labelled cells. Thus, the *F*-measure is used as the evaluation measurement and formulated as follows:

$$\text{Precision} = \frac{TT}{TT + FT} \quad (14)$$

$$\text{Recall} = \frac{TT}{TT + TF} \quad (15)$$

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

In addition, *DistanceError* is also used for evaluation, which is defined as

$$\text{DistanceError} = \frac{\sum_{k=1}^K \min_{x_g \in X} \text{dist}(x, x_g)}{K} \quad (17)$$

Typically, X denotes the set of gold standard cell centres; x_g denotes a gold standard cell centre; x denotes a predicted cell centre; and *dist* denotes Euclidean distance.

Two kinds of *DistanceError* are computed in our experiment by defining different types of K . One is denoted as the total amount of the predicted cell centres, which determines the *DistanceError* to be *DisErr1*; the other is defined as the total amount of the TT, which results in *DisErr2*.

5.3. F-measure result

Table 1 summaries the performance of five methods, including classification model, auto-context, object detection + layout, auto-context + non-maxima suppression, our method with belief propagation and our method with jump-diffusion, along with the *DistanceError* of the last four methods. As we can see from Table 1, the performance of object detection + layout is not satisfied, which is probably because a mere post-processing on detection relies on the prior configuration only, making the algorithm prone to errors in the detection stage. Our method with belief propagation is with a little worse performance *DistanceError* than the other with the use of jump-diffusion process, due to its capability of getting the global optimal solution. The results demonstrate that our methods are better than other state-of-the-art methods.

In addition, the experiment of regression based cell counting is conducted, where each image is downsampled into 18 sub-images, making the prediction of the number of cells in each image feasible. The results further confirm the advantages of our methods, as shown in Fig. 2 and Table 1.

6. Discussion

In recent years, there has been increasing interest in the cell segmentation, especially the quantitative assessment for the medical image, which may further provide a reference for the choice

of therapy. As the code and dataset from [22,23,46] are not publicly accessible, it is quite difficult to directly compare between different algorithms. Hence, we only list the results of those algorithms as references. In [23], the classification of breast cancer nuclei was conducted by Biopsy Analysis Support System (BASS), which extracted four local features (optical density, chromaticity, and textures) and two global features from the individual nuclei, thus resulting in the performance of 72.1% correct classification on the test dataset. The paper [22] used a robust automated tumour cell segmentation model to apply on both routine histochemical tissue slides and immunohistochemistry (IHC) images, which respectively achieved 80% accuracy, 79% precision and 78% accuracy, 89% precision. Liu et al. [46] developed a learning based model based on a new feature set using the eigenvalues of the Hessian matrix to capture the elliptic shape around each pixel for the identification of different types of cells, which was performed on the same dataset with ours, but only with an average 87.34% accuracy (Table 1) for the cell detection.¹ As we can see from the Table 1, only using the auto-context [41] without the further processing of the second layer is with 80.31% precision and 94.57% accuracy, which gets the result by defining the maximum grey scale point in the neighbourhood as the central point of cell. The ignorance of the second layer seems to get a very low precision and a high accuracy, which means the overlapping rate is very high (Fig. 2(d)), emphasizing the importance of the utilization of the second layer model. Also, the precision of object detection [47] + layout [9] method is with good performance but the accuracy is poor, leading to a bad performance for the output. As the object detection algorithm (ICF and DLDM) does not provide a good identification for the overlap part of cells, the detection amount is very big before the layout process, which may reject the correct part of the overlap cell centre and leave the wrong part. Also, negative instances of a big size for a cell image is hard to find, which is filled with a large number of cell centres waiting to be detected, bringing about a bad provision of negative instances for training process.

The auto-context [41] + non-maximum suppression (NMS) method provides a commonplace result for the microscopic cell detection, although it is with the first process of the method the same to the algorithm we provide. That is, the non-maximum suppression only restrains the non-maximum value of the auto-context result, which gives no contribution to the cell centre classification. As we can observe from the original cell image and the result of auto-context algorithm, the restriction along the gradient direction on the grey level image does not provide an accurate detection for microscopy cell centre.

The two-layer model we propose is with the same result from the first layer auto-context process; however, the second layer process we propose is more effective for cell centre detection: learning the rich information provided from prior step, the belief propagation and the jump-diffusion algorithm are used respectively. The global optimal solution is obtained in the jump-diffusion process, which promotes the optimization for the moving and existence problem of microscopic cell centres, thus gains a better performance than the belief propagation process.

7. Conclusions

In this paper, we cast the problem of cell detection from an angle of structural prediction, and introduce a two-layer model, which can realize knowledge representation and propagation. As we can

see, the advantage of our method is evident with effectiveness and efficiency over the state-of-the-art methods.

For detecting the cells of irregular shape, there are still some defects in our result and we do not classify cell types into detailed groups such as the T-cell, B-cell or others. We will use multi-class prediction and take more explicit structures into account in the future.

Acknowledgements

This work was supported by Microsoft Research (MSR). Thanks to MSR eHealth grant, NSF IIS-1216528 (IIS-1360566), NSF CAREER award IIS-0844566 (IIS-1360568), Grant 61073077 from National Science Foundation of China, Grant SKLSDE-2011ZX-13 from State Key Laboratory of Software Development Environment in Beihang University, and the Fundamental Research Funds for the Central Universities of China and D. Chen's research was supported in part by NSF under Grant CCF-1217906.

References

- [1] Peng H. Bioimage informatics: a new area of engineering biology. *Bioinformatics* 2008;24(17):1827–36.
- [2] Li F, Yin Z, Jin G, Zhao H, Wong ST. Bioimage informatics for systems pharmacology. *PLoS Comput Biol* 2013;9(4):e1003043.
- [3] Lee PP, Yee C, Savage PA, Fong L, Brockstedt D, Weber JS, et al. Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat Med* 1999;5(6):677–85.
- [4] Kohrt HE, Nouri N, Nowels K, Johnson D, Holmes S, Lee PP. Profile of immune cells in axillary lymph nodes predicts disease-free survival in breast cancer. *PLoS Med* 2005;2(9):284.
- [5] Smal I, Loog M, Niessen W, Meijering E. Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Trans Med Imaging* 2010;29(2):282–301.
- [6] Cheng ED, Challa S, Chakravorty R. Microscopic cell detection based on multiple cell image segmentations and fusion algorithms. In: *Biomedical engineering and informatics*, 2009. IEEE; 2009. p. 1–7.
- [7] Ersoy I, Bunyak F, Mackey MA, Palaniappan K. Cell segmentation using Hessian-based detection and contour evolution with directional derivatives. In: *15th IEEE international conference on image processing*, 2008. ICIP 2008. IEEE; 2008. p. 1804–7.
- [8] Jurrus E, Paiva AR, Watanabe S, Anderson JR, Jones BW, Whitaker RT, et al. Detection of neuron membranes in electron microscopy images using a serial neural network architecture. *Med Image Anal* 2010;14(6):770–83.
- [9] Desai C, Ramanan D, Fowlkes C. Discriminative models for multi-class object layout. In: *Proc of CVPR*. 2009. p. 229–36.
- [10] Lafferty J, McCallum A, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of 18th International Conference on Machine Learning* 2001, ACM. 2001. p. 282–9.
- [11] Yeo H, Sheinin V, Sheinin Y. An automated image segmentation and classification algorithm for immunohistochemically stained tumor cell nuclei. In: *SPIE medical imaging*. International Society for Optics and Photonics; 2009. p. 48–51.
- [12] Li G, Liu T, Nie J, Guo L, Malicki J, Mara A, et al. Detection of blob objects in microscopic zebrafish images based on gradient vector diffusion. *Cytometry A* 2007;71(10):835–45.
- [13] Tschantzaris I, Joachims T, Hofmann T, Altun Y. Large margin methods for structured and interdependent output variables. *J Mach Learn Res* 2006;6(2):1453.
- [14] Joachims T, Hofmann T, Yue Y, Yu CN. Predicting structured objects with support vector machines. *Commun ACM* 2009;52(11):97–104.
- [15] Yu CNJ, Joachims T. Learning structural SVMs with latent variables. In: *Proceedings of the 26th annual international conference on machine learning*. ACM; 2009. p. 1169–76.
- [16] Bengtsson E, Wahlby C, Lindblad J. Robust cell image segmentation methods. *Pattern Recogn Image Anal C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii* 2004;14(2):157–67.
- [17] Mao K, Zhao P, Tan PH. Supervised learning-based cell image segmentation for p53 immunohistochemistry. *IEEE Trans Biomed Eng* 2006;53(6):1153–63.
- [18] Ranefall P, Egevad L, Nordin B, Bengtsson E. A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei. *Anal Cell Pathol* 1997;15(3):145–56.
- [19] Shitong W, Min W. A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection. *IEEE Trans Inform Technol Biomed* 2006;10(1):5–10.
- [20] Maslov AY, Barone TA, Plunkett RJ, Pruitt SC. Neural stem cell detection, characterization, and age-related changes in the subventricular zone of mice. *J Neurosci* 2004;24(7):1726–33.
- [21] Yang F, Mackey MA, Ianzini F, Gallardo G, Sonka M. Cell segmentation, tracking, and mitosis detection using temporal context. In: *Medical image*

¹ We have also tried to contact many authors working on medical segmentation related to our topic to validate our method. Unfortunately, they either did not answer our email or cannot share the data with us. Nevertheless, we believe the experimental results shown here demonstrate the effectiveness of our method.

- computing and computer-assisted intervention – MICCAI 2005. Springer; 2005. p. 302–9.
- [22] Wang CW. Robust automated tumour segmentation on histological and immunohistochemical tissue images. *PLoS ONE* 2011;6(2):e15818.
- [23] Schnorrenberg F, Pattichis C, Schizas C, Kyriacou K, Vassiliou M. Computer-aided classification of breast cancer nuclei. *Technol Health Care* 1996;4(2):147–61.
- [24] Nguyen N, Norris E, Clemens M, Shin M. Rapidly adaptive cell detection using transfer learning with a global parameter. *Mach Learn Med Imaging* 2011;209–16.
- [25] Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, et al. Real-time human pose recognition in parts from single depth images. *Commun ACM* 2013;56(1):116–24.
- [26] Choi MJ, Lim JJ, Torralba A, Willsky AS. Exploiting hierarchical context on a large database of object categories. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2010. p. 129–36.
- [27] Lee YJ, Grauman K. Object-graphs for context-aware visual category discovery. *IEEE Trans Pattern Anal Mach Intell* 2012;34(2):346–58.
- [28] Yang Y, Hallman S, Ramanan D, Fowlkes C. Layered object detection for multi-class segmentation. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2010. p. 3113–20.
- [29] Zhang L, Ji Q. Image segmentation with a unified graphical model. *IEEE Trans Pattern Anal Mach Intell* 2010;32(8):1406–25.
- [30] Lee CH, Schmidt M, Murtha A, Bistritz A, Sander J, Greiner R. Segmenting brain tumors with conditional random fields and support vector machines. In: *Computer vision for biomedical image applications*. Springer-Verlag Berlin Heidelberg; 2005. p. 469–78.
- [31] Lee CH, Wang S, Murtha A, Brown MR, Greiner R. Segmenting brain tumors using pseudo-conditional random fields. In: *Medical image computing and computer-assisted intervention – MICCAI 2008*. Springer; 2008. p. 359–66.
- [32] Bauer S, Nolte LP, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: *Medical image computing and computer-assisted intervention – MICCAI 2011*. Springer; 2011. p. 354–61.
- [33] Artan Y, Haider MA, Langer DL, van der Kwast TH, Evans AJ, Yang Y, et al. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Trans Image Process* 2010;19(9):2444–55.
- [34] Huh S, Ker DFE, Bise R, Chen M, Kanade T. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *IEEE Trans Med Imaging* 2011;30(3):586–96.
- [35] Han F, Tu Z, Zhu SC. Range image segmentation by an effective jump-diffusion method. *IEEE Trans Pattern Anal Mach Intell* 2004;26(9):1138–53.
- [36] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995;82(4):711–32.
- [37] Srivastava A, Grenander U, Jensen GR, Miller MI. Jump-diffusion Markov processes on orthogonal groups for object pose estimation. *J Stat Plan Infer* 2002;103(1):15–37.
- [38] Joachims T, Hofmann T, Yue Y, Yu CN. Predicting structured objects with support vector machines. *Commun ACM* 2009;52(11):97–104.
- [39] Torralba A, Murphy KP, Freeman WT. Contextual models for object detection using boosted random fields. In: *NIPS*, vol. 1. 2004. p. 2.
- [40] Yu CNJ, Joachims T. Learning structural SVMs with latent variables. In: *Proceedings of the 26th annual international conference on machine learning*. ACM; 2009. p. 1169–76.
- [41] Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans PAMI* 2010;32(10):1744–57.
- [42] Quattoni A, Collins M, Darrell T. Conditional random fields for object recognition. In: *NIPS*. Citeseer; 2004.
- [43] Grenander U, Miller MI. Representations of knowledge in complex systems. *J R Stat Soc Ser B: Methodol* 1994;549–603.
- [44] Lowe DG. Object recognition from local scale-invariant features. In: *The proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 2. IEEE; 1999. p. 1150–7.
- [45] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087.
- [46] Liu X, Setiadi AF, Alber MS, Lee PP, Chen DZ. Identification and classification of cells in multispectral microscopy images of lymph nodes. In: *SPIE medical imaging*. International Society for Optics and Photonics; 2011. p. 79620J.
- [47] Dollár P, Tu Z, Perona P, Belongie S. Integral channel features. In: *British machine vision conference*. 2009. p. 1–11.
- [48] Liaw A, Wiener M. Classification and regression by randomforest. *R News* 2002;2(3):18–22.