Mitosis Detection in Phase Contrast Microscopy Image Sequences of Stem Cell Populations: A Critical Review

An-An Liu, *Member, IEEE*, Yao Lu, Mei Chen, and Yu-Ting Su

Abstract—Detecting mitosis from cell population is a fundamental problem in many biological researches and biomedical applications. In modern researches, advanced imaging technologies have been applied to generate large amount of microscopy images of cells. However, detecting all mitotic cells from these images with human eye is tedious and time-consuming. In recent years, several approaches have been proposed to help humans finish this job automatically with high efficiency and accuracy. In this review paper, we first described some commonly used datasets for mitosis detection, and then discussed different kinds of methods for mitosis detection, like tracking based methods, tracking free methods, hybrid methods, and the most recently proposed works based on deep learning architecture. We compared these methods on same datasets, and found that deep learning based approaches have achieved a great improvement in performance. At last, we discussed the future possible approaches on mitosis detection, to combine the success of previous works and the advantage of big data in modern researches. Considering expertise is highly required in biomedical area, we will further discuss the possibility to learn information from biomedical big data with less expert annotation.

Index Terms—Mitosis detection, computer vision, biomedical image, big data, microscopy image, stem cell

1 Introduction

Modern technology has enabled monitoring of large populations of live cells over extended time periods in experimental settings. Live imaging of cell populations growing in a culture vessel has been widely adopted in biomedical experiments. Such experiments create large amount of time-lapse images, and the information embedded in these images holds tremendous promise for scientific discovery. The growing amount of such data calls for automated or semi-automated computer vision tools to detect and analyze the dynamic cellular processes, and to extract the information necessary for biomedical study. Live cells are often of low contrast with little natural pigmentation, therefore they usually need to be stained or fixed in order to be visible under bright field microscopy or florescence microscopy. However, fixing or staining may destroy the cells or introduce artifacts. Label-free imaging, in particular phase contrast microscopy, is highly desirable for live cell imaging because it allows cells to be examined in their natural state, therefore enables live cell imaging over extended time periods.

In computer vision literature, the detection of cell division is often referred to as *mitosis detection*. Mitosis detection is essential for quantitative analysis of cell populations. The detection of mitotic cells can provide valuable information to improve the performance of tasks such as cell tracking and lineage construction. The information provided by mitosis detection can be further utilized in biological and medical applications such as cell culture monitoring, stem cell manufacturing, drug development, tissue engineering and so on.

Currently, the detection of mitosis is typically performed by people with biological expertise. However, detecting mitosis from large number of cell images by human is both tedious and error prone. The fast growing scale of image data in the biomedical field demands for automated detection in order to improve the accuracy and efficiency of mitosis detection while alleviating manual labor. In this paper, we discuss recent advances in the field of mitosis detection on time-lapse phase contrast microscopy images.

1.1 Mitosis Detection

The task of computer vision based mitosis detection can be further categorized into *identification*, *localization* and *segmentation* problems. Given a static cell image or a dynamic cell image sequence, the goal of mitosis identification is to detect whether there is mitosis in the input data. In addition, mitosis localization approaches can further localize the division of cells in mitotic process spatially and temporally from the input sequence. For the segmentation problem, the mitotic image sequences are further classified into different mitotic phases.

- *A.-A. Liu, Y. Lu, and Y.-T. Su are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China. E-mail: {anan0422, hiluyao}@gmail.com, ytsu@tju.edu.cn.*
- *M. Chen is with the Department of Electrical and Computer Engineering, State University of New York at Albany, Albany, NY 12222. E-mail: meichen@albany.edu.*

Manuscript received 15 Mar. 2017; revised 23 May 2017; accepted 23 June 2017. Date of publication 28 June 2017; date of current version 7 Dec. 2017. (Corresponding author: An-An Liu and Yao Lu.)
Recommended for acceptance by J. Zhu, A.-A. Liu, M. Chen, T. Tasdizen, and H. Su.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2017.2721438

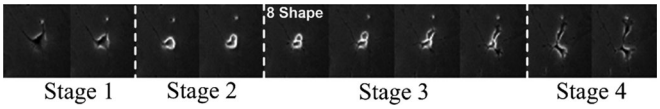


Fig. 1. Four stages of visual appearance transition during mitosis of stem cells [2]. Stage 1: Interphase; Stage 2: Start of mitosis; Stage 3: Formation of daughter cells; Stage 4: Separation of daughter cells.

Based on the changes of visual appearance of stem cells in phase contrast microscopy images, mitosis of stem cells can be segmented into four stages, and each of them is related to biological phases in the cell circle: 1. *Interphase stage*, when the cells still looks normal in appearance; 2. *Start of mitosis*, when the cells begin to shrink, round up and get brighter while going through prophase, metaphase and anaphase; 3. *Formation of daughter cells*, when two daughter cells become visible and attached with each other forming the shape '8' at the period of telophase; 4. *Separation of daughter cells*, when two daughter cells get completely separated after cytokinesis is completed. These four stages in the visual appearance transition are illustrated in Fig. 1.

1.2 Challenges

Unlike fluorescence imaging, phase contrast microscopy is a form of nondestructive imaging method. Under phase contrast microscope, cells can be monitored alive and continuously without changing their structures. Therefore, automated systems can be employed to observe and analyze cell behaviors by recording time-lapse images [1].

The task of mitosis localization is to find the exact time when two daughter cells become visible and shape '8' is formed (stage 3). Different from other pattern recognition tasks, most mitotic cells have irregular shapes and appearance, as shown in Figs. 2a, 2b, and 2c. Samples may vary in orientations (Figs. 2a and 2b), in appearances (Fig. 2b in C3H10 dataset and Fig. 2c in C2C12 dataset), and in temporal lengths (Figs. 2a, 2b, and 2c). Moreover, some non-mitotic sequences may have similar appearance as mitosis sequences (Fig. 2d), which might cause false positives in detection. Therefore it is not sufficient to simply search for the shape '8' through phase contrast images to detect mitosis.

Dense cell population data make mitosis detection challenging, and the level of difficulty only increases as live cells proliferate and further increase the population density where neighboring cells introduce ambiguity. Given their large number in a single microscopy image, it is difficult to find all of them by simply searching for a standard pattern. In recent years, many approaches have been proposed to solve this problem under different applications. Some of them are simply derivatives of cell tracking, and some have utilized the popular deep learning techniques. We will divide these methods into four categories and discuss their strengths and weaknesses respectively.

1.3 Organization

This paper is organized as follows: First, we describe the phase contrast microscopy image sequence datasets we use for evaluating the performance of mitosis detection in Section 2. Then we discuss four categories of methods for mitosis detection in recent years in Section 3, 4, 5, and 6, respectively. In Section. 7 we compare and analyze the performance of different methods. Finally, we conclude and discuss future directions for mitosis detection.

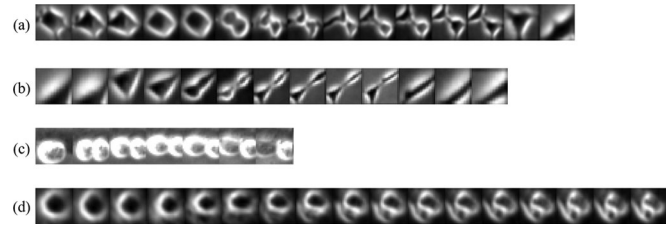


Fig. 2. (a)-(c): The dynamic process of mitosis in phase contrast images. (d) shows a non-mitotic sequence which has similar appearance as mitotic one.

1.3.1 Difference with Previous Surveys

Previous surveys mainly focus on mitosis detection and classification in static cell images such as Indirect Immunofluorescence Image (IIF) [3] and Histology Image [4]. Our survey focuses on the advances of mitosis detection approaches that make use of dynamic information from time-lapse phase contrast microscopy image sequences.

1.3.2 Contributions

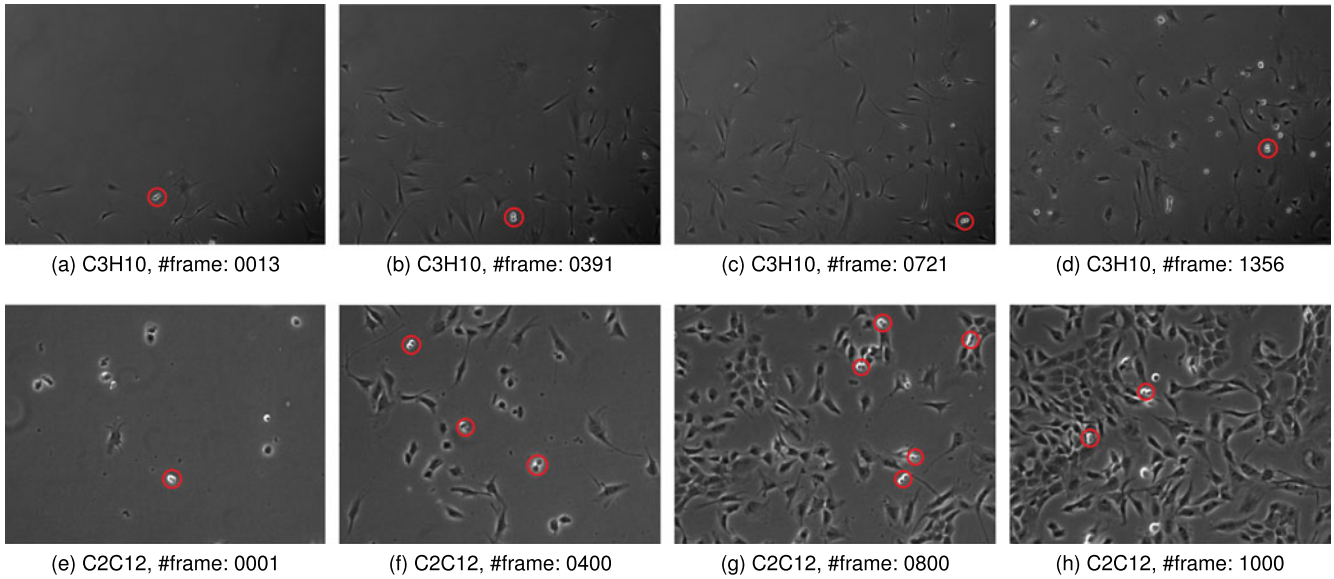
The contributions of this survey are as follows:

- We review the recent advances for mitosis detection in time-lapse phase contrast microscopy image sequences, and group them into four categories. Among these categories, we mainly focus on the last two which take advantage of dynamic information and deep learning features respectively.
- We present a common workflow for mitosis detection and/or localization in phase contrast microscopy image sequences. Based on this workflow, we discuss state-of-the-art techniques for each step in the process.
- We compare the performance of these methods on two popular datasets, and provide in-depth discussion of their respective strengths and weaknesses.

2 Evaluation Datasets

Phase contrast microscope [5] can transform phase variations in light caused by transparent specimen into amplitude changes in light which can be directly observed with human eyes or cameras. Fig. 3 shows phase contrast microscopy images of stem cell populations at different periods of time in C3H10 and C2C12 datasets.

The C3H10 dataset consists of five phase contrast microscopy image sequences of C3H10T1/2 mouse mesenchymal stem cell populations (ATCC, Manassas, VA) [6]. The cells were grown with Dulbecco's Modified Eagle's Media (DMEM; Invitrogen, Carlsbad, CA), 10 percent fetal bovine serum (Invitrogen, Carlsbad, CA) and 1 percent penicillin streptomycin (PS; Invitrogen, Carlsbad, CA). During the culturing, cells were observed under a Zeiss Axiovert 135TV inverted microscope with a $5\times$, 0.15 N.A. objective with phase contrast optics. For every 5 minutes, a time-lapse image was taken by a 12-bit Qimaging Retiga EXi fast 1,394 CCD camera at 500 ms exposure with a gain of 1.01. Each sequence contains 1,436 images with the resolution of



(a) C3H10, #frame: 0013 (b) C3H10, #frame: 0391 (c) C3H10, #frame: 0721 (d) C3H10, #frame: 1356

(e) C2C12, #frame: 0001 (f) C2C12, #frame: 0400 (g) C2C12, #frame: 0800 (h) C2C12, #frame: 1000

Fig. 3. Phase contrast microscopy images in C3H10 and C2C12 dataset at different time of cell culturing. Mitotic cells are marked out with red circles.

1,392 $\times$ 1,040 pixels, 1.3 μm/pixel. After image acquisition, the mitotic cells are annotated by an expert biologist. For the occurrence of each mitosis, the annotator located the center of the boundary between the two daughter cells once they can be observed clearly in the sequence. Each image sequence contains 41-128 mitotic events.

The C2C12 dataset contains 48 sequences of myoblastic stem cell populations, cultured under 4 different conditions [1]: 1. *Control*, 2. *FGF2*, 3. *BMP2*, 4. *FGF2+BMP2*. Cells in the control group were grown under the same condition as that for cells in the C3H10 dataset, and with additional grow factors (FGF2, BMP2 or both) for the other three groups. During cell growth, images were captured every 5 minutes with the same equipment described in the C3H10 dataset for about 84 hours in total. Each sequence contains around 1,000 images at the resolution of 1,392 $\times$ 1,040 pixels, 1.3 μm/pixel. As shown in Figs. 3d and 3h, C2C12 myoblasts were cultured at much higher confluence than C3H10T1/2 mesenchymal stem cells, with each sequence containing significantly more mitotic events than the C3H10 dataset (673 mitotic events in the first sequence of the C2C12 dataset). It was impractical to have expert annotations for every mitosis occurrence in 48 image sequences at high cell confluence. Therefore mitosis annotation was performed on the first sequence of the dataset by an expert biologist.

3 Tracking Based Method

The trajectory generated by cell tracking can provide important information on cell cycle. This information can be further used for mitosis detection. In the tracking based methods, mitosis detection usually serves as part of the tracking system. In [7], Yang et al. extracts Region of Interest (ROI) of mitosis candidate region based on cell trajectories. Some properties of the ROI, like area, perimeter, circularity *etc.*, are computed in adjacent frames to determine whether mitosis occurs in the candidate regions.

Debeir et al. [8] defined a specialized mean-shift kernel to identify cells entering mitotic phase during the cell tracking process. After the cell trajectories are established, the mitotic event can be further detected in the inverse of the time direction when the centroids of two daughter cells progressively become close enough to be merged in the reverse direction.

Thirusittmapalam et al. [9] proposed a cellular tracking framework using segmentation-driven approach. This framework consists of a forward tracking process and a backward tracking process. Forward tracking generates tracking results of cells, and backward tracking constructs a tree structure of cell trajectories, where each branch is associated with a mitosis event. Fig. 4 shows the tree structured trajectories of three cells families.

Liu et al. [10] used a Kalman filter to predict the movement of plant shoot apical meristem (SAM) cells in time-lapse images, and search mitotic events in the neighbourhood of the predicted position. This method uses segmented cell boundary and area size as information to detect cell division events.

Discussion. Mitosis detection plays an important role in the cell tracking task, so it is necessary to integrate mitosis detection into a cell tracking system. On the other hand, considering the movement of cells during mitosis, trajectories can also

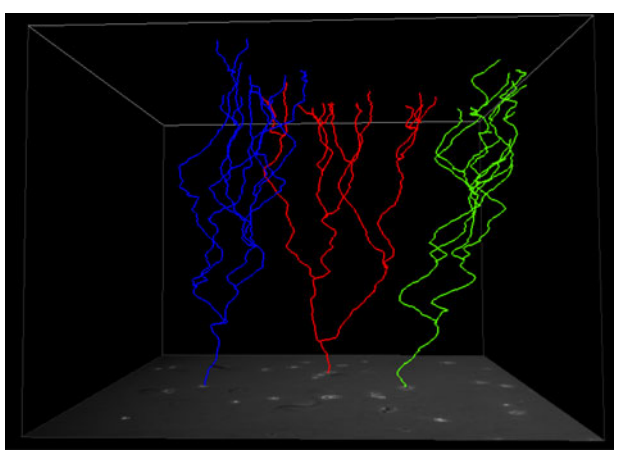


Fig. 4. Trajectories of three cell families. Each branching in the trajectory indicates a mitotic event. The image is generated by the BioCV CTK GUI tool.

be used for the detection of mitotic cells. In this section, we introduced methods for mitosis detection utilizing cell tracking results. However, robust tracking of individual cells in a dense population over image sequences of hundreds to thousands of frames is challenging. Mitosis detection based on cell trajectories typically suffer the same problem of cell tracking. Therefore it is necessary to develop mitosis detection techniques which are independent of cell tracking.

4 Tracking Free Method

Liu et al. [11] proposed a method to detect mitosis events using Support Vector Machine (SVM). The proposed method first generates the preconditioned images with the Nonnegative Mixed-norm Preconditioning method [12], then extracts candidate spatiotemporal volumes using the volumetric region growing method. At last, SVM is utilized to classify each frame of the candidate sequence for mitosis detection. *Area*, *shape* and *intensity* features of candidate regions are used for the training of SVM.

Instead of using segmentation based method to extract candidate regions, Marcuzzo et al. [13] utilized a convergence index filter to find nuclei regions in microscopy image of the Arabidopsis thaliana root. Afterwards, a linear classifier is trained using the shape characteristics of nuclei regions for mitosis detection.

Considering the irregular appearances of cells, most popular visual features have difficulty in discriminatively representing mitotic cells. Liu et al. [14] proposed a nonnegative mix-norm convex optimization method for a sparse representation of mitotic cells to overcome the difficulty in feature formulation for deformable objects. The sparse representation task can be formulated as an optimization problem in [14]

$$(\Phi^*, W^*) = \arg \min_{\Phi \in \mathbb{C}, W \in \mathbb{R}^{M \times N}} \sum_{i=1}^{N} \Big(||X_i - \Phi \cdot w_i||_2^2 + \gamma_1 ||w_i||_1 + \gamma_2 ||w_i||_2^2 \Big), \ s.t.\ w_i \geq 0, \tag{1}$$

where $\Phi = \{\phi_i \in R^{d \times 1}\}_{i=1}^{M}$ is the basis set, $X = \{X_i \in R^{d \times 1}\}_{i=1}^{N}$ represents set of low level features for input images, $w_i \in R^{M \times 1}$ is a sparse vector denoting the correlation between X_i and each basis, $W = \{w_i\}_{i=1}^{N}$, γ_1 and γ_2 control the sparsity and consistence of the sparse representation w_i respectively. In [14], the optimization problem is solved by a tailored online learning algorithm, and then the sparse representation w_i is fed into a SVM classifier for mitosis detection.

Huang et al. [15] proposed an approach to detect mitosis in histopathological images using a sparse representation based method called eXclusive Independent Component Analysis (XICA). Unlike PCA and ICA methods that extract major components from a single set of input signals, XICA is able to extract two sets of exclusive bases for mitotic and non-mitotic cell images respectively. Given a test image, the pattern is evaluated based on both sets of exclusive bases respectively to obtain the relative residuals. The residuals are further used for the classification of the input image.

To take advantage of the dynamic information in cell image sequences, Li et al. [16] proposed a method to locate mitosis in the spatiotemporal domain by training a fast cascade classifier [17] with a set of 3D Haar-like features (Fig. 5). The detection is performed by classifying all $24 \times 24 \times 5$ sub-volumes in a spatial-temporal sliding window. The classification results are convolved with a 3D Gaussian kernel with diagonal covariance and normalized to $[0, 1]$ to represent the probabilities of mitosis.

Fig. 5. A set of volumetric Haar-like filters used to extract features for mitosis detection (from [16]).

Quelhas et al. [18] used optical flow to detect morphology changes in time-lapse images of Arabidopsis thaliana root cells. After global movements removed with image registration, the approach considered the obvious movement in optical flow as mitotic event. This method is segmentation free, which means the detection can be performed on low contrast images. But the detection only relies on optical flow result, making the method unsuitable for images containing moving cells.

Discussion. These tracking free methods are mainly performed on individual images. They usually extract features to represent the visual appearances in static images (e.g., HOG, GIST, and sparse representation) or dynamics within 3D frame volumes (e.g., 3D Haar), and then classify the features with off-the-shelf classifiers (e.g., SVM classifier).

These mitosis detection methods do not rely on movement tracking of cells, which makes them suitable for static images such as Immunofluorescence Image [3] and Histology Image [4]. Since time-lapse phase contrast microscopy image datasets provide dynamic recording of cells, it is essential to take advantage of the temporal information during mitosis for robust detection. Although the 3D Haar feature can extract temporal features between frames, the sliding window scheme is computationally expensive and 3D Haar features cannot extract temporal information of longer durations. Methods that can explore relatively long-term temporal context without cell tracking would be necessary to improve mitosis detection performance.

5 Hybrid Method

Time-lapse phase contrast microscopy images provide the opportunity to monitor cell behaviors continually over time. Since mitosis is a dynamic process, it is necessary to model the evolution of visual patterns between frames given the sequence of mitotic cell images. To overcome the limitations of both tracking based and tracking free methods, whiling taking advantage of visual appearance as well as the temporal dynamic information in mitotic image sequences, a series of hybrid methods have been proposed. These methods employ Graphical Models to discover the conditional dependence structure between random variables. Besides identifying mitotic sequences, some methods can further determine the exact time when the daughter cells are born in mitosis image sequences. These hybrid methods typically consist of three steps: 1. candidate sequence extraction, 2. feature extraction, 3. mitosis detection with graphical models. This workflow is shown in Fig. 6. Based on the workflow, we compared some state-of-the-art methods for

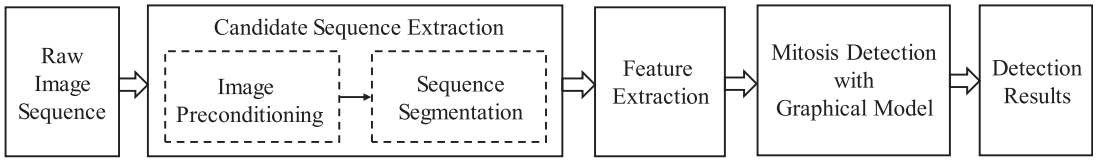


Fig. 6. A common workflow for mitosis detection in hybrid methods. Mitosis candidate sequences are extracted after image preconditioning and segmentation, then visual features are extracted on each frame in sequence. The feature vectors are finally fed into graphical models to explore temporal dynamics for detection.

mitosis detection in Table 1. We will provide a detailed description of the advanced techniques proposed for these steps separately in Sections 5.1, 5.2 and Section. 5.3.

5.1 Candidate Sequence Extraction

The first step of mitosis detection is to extract the candidate sequences from the entire time-lapse image sequence. This step aims to select all the potential mototic regions and filter out regions where mitosis is unlikely to occur. This step can significantly reduce the search space for the subsequent learning-based classification.

In [21], each image is first processed by averaged image subtraction over the sequence to remove stationary bright artifacts and to correct intrinsic illumination variations. Afterwards, the images are convolved with the average filter for noise removal and transformed to a

TABLE 1
Methods for Mitosis Detection in Time-Lapse Phase Contrast Images

method	preconditioning	foreground segmentation	candidate sequence construction	features	detection
SVM [14]	nonnegative mixed-norm algorithm		volumetric region growing	area, shape, intensity features of foreground blob	SVM classification on single frame
3D Haar [16]	calculate local variance map from original image		3D sub-volume generated by temporal sliding window	a set of volumetric Haar-like filters	fast cascade learning framework [17]
HMM [19]	smooth with median filter, binarize with a empirical threshold	immersion watershed	group overlapping regions in neighbouring frames	1.area of the object, 2. change in area, 3. change in position, compactness, 4. average intensity, 5. entropy, 6. contrast, 7. energy, 8. correlation	sequence classification with 3 HMM models
CRF [20]		watershed algorithm, followed by a merging procedure	track segmented area between consectuive frames	Haralick co-occurrence textural features (14), Zernike moment features (47), Gabor features (85), shape features (54)	predict a label sequence corresponding to input sequence with CRF
HCRF [6]	nonnegative mixed-norm algorithm		3D seeded growing	1. Intensity Histogram, 2.Histogram of Oriented Gradients (HOG), 3. Gist	mitosis identification with HCRF model
EDCRF [21]	averaged image substraction, smooth with average filter	binarize with empirical threshold, merge overlapped bounding box	associate spatially overlapped patches in consectuive frames	unique scale gradient histogram	mitosis localization with EDCRF model
TL-HCRF [22]	averaged image substraction, smooth with average filter	detect candidate birth event in single frames with SVM	track candidate birth event backward and forward	unique scale gradient histogram	SVM for candidate birth detection, TL-HCRF for candidate sequence detection
MM-HCRF+ MM-SMM [2]	constrained optimization for artifact-free image restoration		3D seeded growing	SIFT	MM-HCRF for candidate sequence identification, MM-SMM for sequence segmentation
HCNN [23]	flat field correction, background subtraction, smooth with Gaussian filter	binarize with empirical threshold, generage bounding box of connected components	associate spatially overlapped patches in consectuive frames	original cell images + motion images computed by central finite difference	Hierarchical CNN network
3D-CNN [24]			fixed 3D window centered on annotated coordinate	CNN feature extracted by 3D-CNN network	SVM

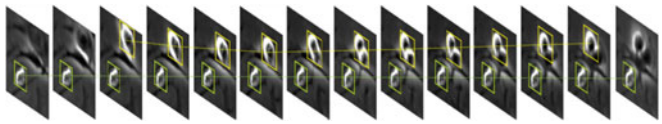


Fig. 7. In [21], candidate sequence is constructed based on spatial overlapped candidate patches in neighbouring frames.

Fig. 8. Restored artifact free image of mitotic cells in phase contrast images. (a) and (c) are the original images, and (b) and (d) are the corresponding restored image.

binary image using an empirical threshold. Bounding boxes are obtained by searching for connected components through the binary image. Overlapped bounding boxes are then merged together. Fixed sized candidate patch is extracted based on the center of the merged bounding boxes. Then spatially overlapped patches in consecutive frames are associated to construct candidate patch sequences, as shown in Fig. 7.

Oftentimes phase contrast microscopy images contain artifacts such as *halo* and *shade-off* due to the optical principle and the inherent imperfections during the conversion process [5]. These artifacts may affect the result of thresholding and cause low performance in the subsequent mitosis detection steps. In [2], artifacts caused by phase contrast microscope can be removed using a constrained optimization method [25] before further processing. According to [25], the phase contrast image can be approximated by a linear model

$$\mathbf{g} \approx \mathbf{P}\mathbf{f} + \mathbf{constant}, \tag{2}$$

where $\mathbf{g}$ is the observed image and $\mathbf{f}$ represent the artifact free image.

$\mathbf{P}$ is the convolutional kernel representing the point spread function (PSF) of the phase contrast microscope

$$PSF(u,v) = \delta(u,v) - airy(\sqrt{(u^2+v^2)}), \tag{3}$$

where $airy(u,v)$ is an obscured Airy pattern ([25], [26]).

With the linear image model, the artifact-free image $\mathbf{f}$ can be restored from observed image $\mathbf{g}$ by optimizing the following constrained objective function with an iteratively reweighted nonnegative multiplicative update algorithm [12]

$$O(\mathbf{f}) = ||\mathbf{P}\mathbf{f} - \mathbf{g}||_2^2 + w_s \mathbf{f}^T \mathbf{L}\mathbf{f} + w_\tau ||\mathbf{D}\mathbf{f}||_1 \quad s.t.\ f_k \geq 0, \forall k, \tag{4}$$

where $\mathbf{L}$ and $\mathbf{D}$ are Laplacian matrix and a diagonal matrix defining local smoothness and sparseness with corresponding weights w_s and w_τ respectively.

After the image being restored, cell regions can be segmented by simple thresholding method. Fig. 8 shows two samples of mitotic cell images before and after artifact removal.

After the segmentation of potential mitotic areas from the background, [2] and [6] applied 3D seeded region growing method [27] to extract spatiotemporal subregions as candidate mitotic sequences. This method relies on two thresholds: a threshold computed by Otsu's method [28] to detect seeds of mitotic regions, and a lower threshold determined by Rosin's unimodal thresholding method [29] used as the stopping criterion for region growing.

5.2 Image Feature Extraction

After the extraction of candidate sequence, one or more types of features need to be extracted from the candidate sequences for the classification. These features attempt to capture a distinct representation of the visual appearances of mitosis.

Histogram of Oriented Gradients (HOG). HOG [30] features are calculated based on the local distribution of intensity gradient to characterize local shapes in the image. In practice, the input image is divided into subregions, and for each subregion a histogram of gradient directions over pixels in this subregion is calculated. The histograms are then combined to form the final feature representation of the image.

GIST. GIST can represent low-level features, intermediate image properties, and high-level information of images at the same time [31]. The advantages of employing GIST for mitosis detection are: 1. GIST can bypass the segmentation of individual objects; 2. GIST is invariant to scale and rotation; 3. GIST is a compact descriptor. Some works used GIST feature [32], or the combination of HOG and GIST features [6] for mitosis detection.

Scale Invariant Feature Transform (SIFT). SIFT [33] can extract scale and orientation invariant representations of the original images. After scale detection and orientation assignment, histograms of gradient orientations are calculated in subregions of image at assigned scale and orientation as the final representation. Each vector is normalized by its L^2 norm to achieve illumination invariance. In [2], SIFT features are computed for each frame of the candidate sequence for mitosis identification and segmentation.

Unique Scale Gradient Histogram. [21], [22] make use of unique scale gradient histogram as the feature for each image patch in the candidate sequence. Each patch is divided into 4×4 subregions. In each subregion, a histogram is calculated by accumulating gradient magnitudes weighted by a Gaussian function into four bins along the orientations. After the calculation, a L_2 normalization is applied to the feature vector. To achieve rotation invariance, features are also extracted with patch images rotated by $90°$, $180°$, $270°$ in [21] for mitosis detection.

Convolutional Neural Network (CNN) feature. CNN is a deep learning method that has demonstrated impressive performance solving certain computer vision problems. After being trained on labeled images, the output of each layer in a CNN network can be used as features to represent the input image. Compared with the hand-crafted features mentioned above, CNN feature can be learned directly from data. This can make the CNN features more relevant to the given image dataset. In addition, different layers in the CNN network have the ability to extract different levels of visual features to describe the image data. The implementations of CNN features for mitosis detection will be described in Section 6.

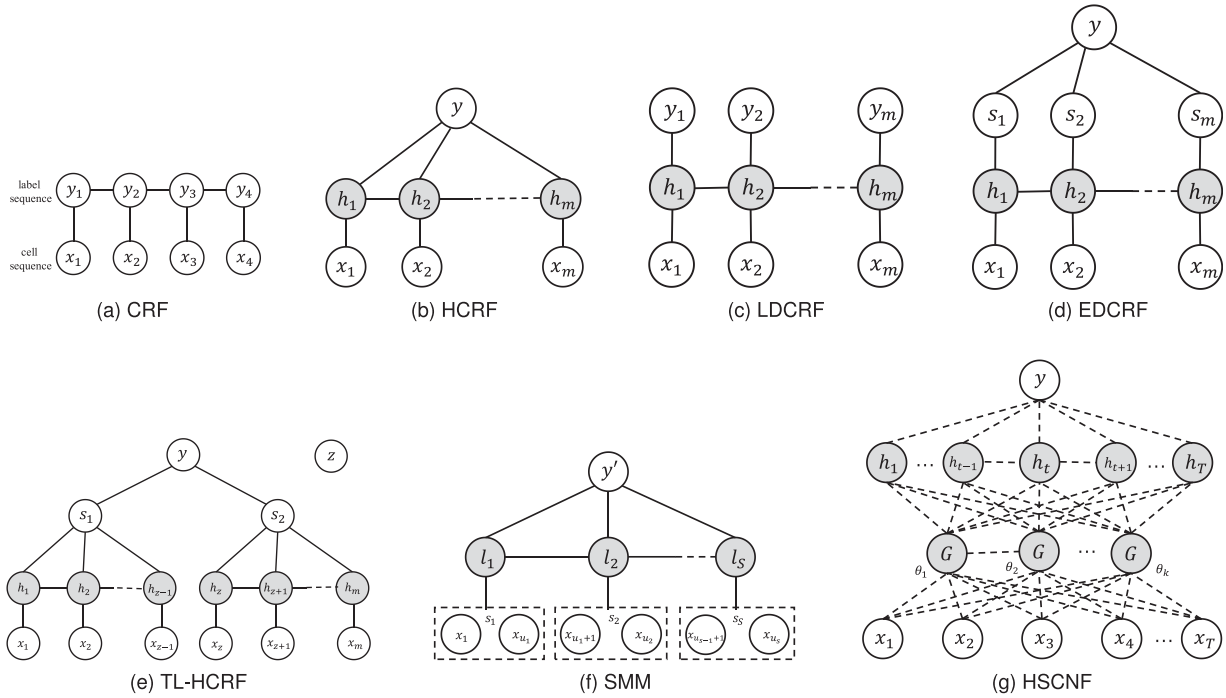


Fig. 9. Structures of different graphical models for mitosis detection.

5.3 Graphical Model

The ability to model temporal dynamics from sequence makes graphical model a suitable solution for mitosis detection in time-lapse phase contrast microscopy image sequences [41]. Different kinds of graphical models have been utilized to detect mitosis in recent years (Fig. 9).

5.3.1 Hidden Markov Model (HMM)

Gallardo et al. [19] proposed a method to detect mitotic cells in time-lapse microscopy images of live cells using the Hidden Markov Model. Before detection, cell images are smoothed using the median filter and binarized using an empirical threshold to find potential mitotic cell areas. These potential areas are then segmented into background and foreground using the immersion watershed algorithm [34]. Overlapped regions in neighboring frames are grouped into candidate sequence. Features are extracted based on several characteristics of candidate regions: area of object, change in area, change in position, compactness, average intensity, entropy, contrast, energy and correlation. The feature vectors are then fed into HMM for classification.

In the classification stage, three HMMs are trained to label and rank the cell candidates. HMM is an extension of Markov chains. First order Markov chain assumes that current state only depends on the preceding one. This process can be described by a probability transition matrix $\mathbf{A} = \{a_{ij}\}$

$$\begin{aligned} a_{ij} &= P(q_t = S_j | q_{t-1} = S_i), 1 \le i, j \le N. \\ a_{ij} &\ge 0, \quad \sum_{j=1}^{N} a_{ij} = 1. \end{aligned} \tag{5}$$

Where q_t is the state at time t, N is the number of states.

The initial state of the process can be express as

$$\pi_j = P(q_1 = S_j), 1 \le j \le N. \tag{6}$$

Unlike Markov chains, the states in HMM cannot be directly observed, instead their relationship with the observations can be represented as conditional probability matrix $\mathbf{B} = \{b_{jk}\}$

$$b_{jk} = P(v_{kt} | q_t = S_j), 1 \le j \le N, 1 \le k \le M. \tag{7}$$

Where v_{kt} represents the observation of symbol v_k at time t, M is the number of observation symbols.

At last the HMM model can be specified by the matrix A, B and vector π: $\lambda(A, B, \pi)$. In [19], three HMMs are trained with different classes of cell candidate sequences for classification, as shown in Fig. 10. During the classification, the feature vector of input sequence are fed into all three HMMs, and the HMM that gives the largest probability will be used to classify the sequence.

5.3.2 Conditional Random Field (CRF)

HMM assumes that observations are independent, which means that the HMM model cannot make use of the context information among cell frames. To address this problem, Liang et al. [20] proposed a method utilizing Conditional

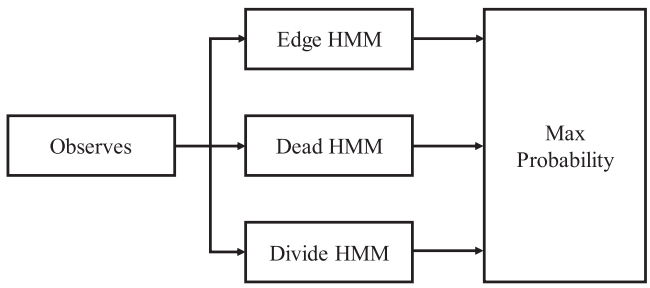


Fig. 10. Classification system built upon three HMMs in [19]. These three HMM models are trained separately on cell edges, dead cells and dividing cells. Given test inputs, the HMM with the largest probability is used to classify the candidates.

Random Field to exploit more contextual information in candidate sequences, and predict the phase label for each frame in the cell image sequence.

As shown in Fig. 9a, given a cell image sequence of length T: $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, the task is to learn a mapping from input $\mathbf{x}$ to a sequence of labels $\mathbf{y}$ indicating the phase of each frame in the input sequence. In the CRF model, given an input sequence $\mathbf{x}$, a conditional probability distribution of $\mathbf{y}$ is defined as

$$p(\mathbf{y}|\mathbf{x}, \theta) = \frac{e^{\Psi(\mathbf{y},\mathbf{x};\theta)}}{\mathbf{Z}(\mathbf{x})}, \tag{8}$$

where θ are the parameters of the model, $\Psi(\mathbf{x}, \mathbf{y}, \theta)$ is the potential function, and $\mathbf{Z}(\mathbf{x}) = \sum_y \Psi(\mathbf{y}, \mathbf{x}; \theta)$ is the partition function.

The key of CRF modeling is to formulate an appropriate potential function. Liang et al. [20] makes an assumption that a particular cell phase only depends on the current cell image, the immediate previous phase, and the immediate following phase. Therefore the potential function can be defined as

$$\Psi(\mathbf{y}, \mathbf{x}; \theta) = \sum_{i=1}^{T} \boldsymbol{\phi}(\mathbf{x_i}) \cdot \theta(y_i) + \sum_{i,j \in E} \theta(y_i, y_j), \tag{9}$$

where $\boldsymbol{\phi}(\mathbf{x_i}) \in \mathbb{R}^d$ is feature vector extracted from cell image $\mathbf{x_i}$, $\theta(k) \in \mathbb{R}^d$ represent the compatibility between image x_i and label y_i, $\theta(i, j) \in \mathbb{R}, i, j \in \{0, 1\}$ represents the compatibility between two labels.

5.3.3 Hidden Conditional Random Field (HCRF)

One drawback of CRF is that it assumes the label sequence is fully observable. Thus, as the training data each frame in the sequence needs to be labeled. This significantly increases the cost for generating the training dataset. HCRF introduces intermediate hidden states to model the latent structure of input sequence, making it possible to train the model with a single label per sequence. In [6] Liu et al. applied the HCRF model for mitosis detection. The structure of a HCRF model is shown in Fig. 9b.

Like Eq. (8), the conditional probability of label y in HCRF model can be defined as

$$\begin{aligned} p(y|\mathbf{X}, \theta) &= \sum_h p(y, \mathbf{h}|\mathbf{X}, \theta) \\ &= \frac{\sum_h e^{\Psi(y,\mathbf{h},\mathbf{X};\theta)}}{\sum_{y' \in \mathbf{Y}, h \in \mathbf{H}} e^{\Psi(y',\mathbf{h},\mathbf{X};\theta)}}, \end{aligned} \tag{10}$$

where $\mathbf{h} = \{h_1, h_2, \ldots, h_T\}$ is a vector of hidden variables, $\Psi(y, \mathbf{h}, \mathbf{X}; \theta) \in R$ is a potential function, and θ are parameters in the HCRF model.

$$\begin{aligned} \Psi(y, \mathbf{h}, \mathbf{X}; \theta) = &\sum_{j=1}^{m} \sum_{l \in L_1} f_{1,l}(j, y, h_j, \mathbf{X}) \theta_{1,l} \\ &+ \sum_{(j,k) \in E} \sum_{l \in L_2} f_{2,l}(j, k, y, h_j, h_k, \mathbf{X}) \theta_{2,l}, \end{aligned} \tag{11}$$

here L_1 is the set of node features, L_2 is the set of edge features, $\theta_{1,l}, \theta_{2,l}$ are the components of θ corresponding to node and edge parameters respectively. $f_{1,l}$ is the feature function depending on a single hidden variable value in the model, while $f_{2,l}$ is the feature function depending on a pair of values.

5.3.4 Event Detection Conditional Random Field (EDCRF)

Considering the structure of the HCRF model, it is not capable of locating the division time of the daughter cells among the image frames of the candidate sequence. Huh et al. [21] proposed an effective approach to explicitly locate birth event from image patch sequences with Event Detection Conditional Random Field. EDCRF is a combination of HCRF and LDCRF [35], as shown in Figs. 9b, 9c and 9d.

In EDCRF, given an input sequence $\mathbf{x}$, the corresponding label of the sequence y is defined as Eq. (12)

$$y = \begin{cases} p, & \text{if } p\text{th patch of } \mathbf{x} \text{ contains a birth event.} \\ 0, & \text{if } \mathbf{x} \text{ is not a mitosis sequence.} \end{cases} \tag{12}$$

And the sub-label $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$ is defined in Eq. (13)

$$s_j = \begin{cases} N(\text{no event}), & \text{if } y = 0 \\ B(\text{before event}), & \text{if } y > 0 \text{ and } j < y \\ A(\text{after event}), & \text{if } y > 0 \text{ and } j \geq y. \end{cases} \tag{13}$$

With the definition of sequence labels and sub-labels, the conditional probability of label y in EDCRF model can be defined in Eq. (14)

$$P(y|\mathbf{x}, \theta) = P(\mathbf{s}|\mathbf{x}, \theta) = \sum_{\mathbf{h}} P(\mathbf{s}|\mathbf{h}, \mathbf{x}, \theta) P(\mathbf{h}|\mathbf{x}, \theta). \tag{14}$$

To make the training of the model tractable, EDCRF restricts that each sub-label s is associated with a disjoint set of hidden states, defined as $\mathcal{H}_s$ in Eq. (15)

$$P(\mathbf{s}|\mathbf{h}, \mathbf{x}, \theta) = \begin{cases} 1, & \text{if } \forall h_j \in \mathcal{H}_{s_j} \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Thus the EDCRF model can be simplified as Eq. (16)

$$P(y|\mathbf{x}, \theta) = \sum_{\mathbf{h}: \forall h_j \in \mathcal{H}_{s_j}} P(\mathbf{h}|\mathbf{x}, \theta). \tag{16}$$

And $P(\mathbf{h}|\mathbf{x}, \theta)$ can be defined as a typical CRF formulation, in Eq. (17)

$$\begin{aligned} P(\mathbf{h}|\mathbf{x}, \theta) = &\frac{1}{Z} \exp\Bigg(\sum_{j=1}^{m} f^{(s)}(h_j, \mathbf{x}, j) \cdot \theta^{(s)}(h_j)) \\ &+ \sum_{j=2}^{m} f^{(t)}(h_{j-1}, h_j, \mathbf{x}, j) \cdot \theta^{(t)(h_{j-1}, h_j)} \Bigg), \end{aligned} \tag{17}$$

where Z is the partition function, $f^{(s)}(h_j, \mathbf{x}, j)$ is a state function defined with feature vector of image patch $\mathbf{x}_j$ in Eq. (18), $f^{(t)}(h_{j-1}, h_j, \mathbf{x}, j)$ is a transition function defined in Eq. (19). $\mathcal{U} = (\mathcal{H}_N \times \mathcal{H}_N) \cup (\mathcal{H}_B \times \mathcal{H}_B) \cup (\mathcal{H}_B \times \mathcal{H}_A) \cup (\mathcal{H}_A \times \mathcal{H}_A)$

represents a set of sub-label transitions: $\{(N \rightarrow N), (B \rightarrow B), (B \rightarrow A), (A \rightarrow A)\}$

$$f^{(s)}(h_j, \mathbf{x}, j) = \phi(x_j) \tag{18}$$

$$f^{(t)}(h_{j-1}, h_j, \mathbf{x}, j) = \begin{cases} 1, & \text{if } (h_{j-1}, h_j) \in \mathcal{U}. \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

5.3.5 Two Labeled Hidden Conditional Random Field (TL-HCRF)

The EDCRF approach has achieved good performance in terms of both cell birth event detection accuracy and computational efficiency. When the cell confluence increases, which is common at later stages of cell culturing, the method has difficulty identifying individual mitotic cells. To address this challenging problem, Huh et al. [22] proposed an approach to detect mitosis in image sequences of high cell confluence in three steps: 1. Pre-detection of cell birth event candidates using visual features extracted from individual image frames; 2. Construction of candidate sequence by tracking candidate birth events over time; 3. Classification of candidate sequence with a specially designed graphical model called Two Labeled Hidden Conditional Random Field (TL-HCRF). This approach has been tested on the last 100 frames of the C2C12 sequence which have very high confluence. The results have outperformed previous methods in terms of detection accuracy. To the best of our knowledge, this is the best result reported on this high confluence data.

Before the pre-detection of candidate birth events, statistical property (mean and standard deviation of pixel values) is first calculated on each image patch to exclude patches that are unlikely to contain birth events (background patches for example). Then Support Vector Machine is trained using unique scale histograms [21] as visual features to detect candidate patches that might contain birth events. After pre-detection, candidate patch sequences are constructed by tracking candidate birth event forward and backward over adjacent image frames.

Unlike EDCRF which assumes that birth event can occur at any image patch in the candidate sequence, TL-HCRF can make use of the information from candidate birth event detection to help determine the true birth event timing. The graphical structure of TL-HCRF is shown in Fig. 9e. TL-HCRF model has two label variables y and z, $y \in \{0, 1\}$ indicating whether the sequence $\mathbf{x}$ contains a birth event, and z represent the time of the birth event in the image sequence.

$$\begin{aligned} P(y|\mathbf{x}, z; \theta) &= P(s_1, s_2|\mathbf{x}, z; \theta) \\ &= \sum_{\mathbf{h}} P(\mathbf{h}, s_1, s_2|\mathbf{x}, z; \theta) \end{aligned} \tag{20}$$

$$P(\mathbf{h}, s_1, s_2, |\mathbf{x}, z, \theta) = \frac{1}{Z} \exp(\Psi(\mathbf{h}, s_1, s_2, \mathbf{x}, z; \theta)). \tag{21}$$

where Z is the partition function, the potential function $\Psi(\mathbf{h}, s_1, s_2, \mathbf{x}, z; \theta)$ can be defined based on the formulations of CRF and HCRF model as

$$\begin{aligned} \Psi(\mathbf{h}, s_1, s_2, \mathbf{x}, z; \theta) &= \sum_{j=1}^{m} f^{(s)}(h_j, \mathbf{x}, j) \cdot \theta^{(s)}(h_j) \\ &+ \sum_{j=2}^{z-1} f^{(t)}(h_{j-1}, h_j, \mathbf{x}, j) \cdot \theta^{(t)}(h_{j-1}, h_j, s_1, s_1) \\ &+ f^{(t)}(h_{z-1}, h_z, \mathbf{x}, k) \cdot \theta^{(t)}(h_{z-1}, h_z, s_1, s_2) \\ &+ \sum_{j=z+1}^{m} f^{(t)}(h_{j-1}, h_j, \mathbf{x}, j) \cdot \theta^{(t)}(h_{j-1}, h_j, s_2, s_2) \\ &+ \sum_{j=1}^{z-1} \theta^{(l)}(h_j, s_1) + \sum_{j=z}^{m} \theta^{(l)}(h_j, s_2), \end{aligned} \tag{22}$$

where $f^{(s)}$ is the state function, and $f^{(t)}$ is the transition function, $\theta^{(s)}$ and $\theta(t)$ are the parameters of state function and transition functions respectively, $\theta(l)$ is the parameter associated with sub-labels s_1, s_2.

5.3.6 Max-Margin Hidden Conditional Random Field and Max-Margin Semi-Markov Model (MM-HCRF+MM-SMM)

Liu et al. [2] proposed an approach to classify and segment candidate sequences with two different models. Max-Margin Hidden Conditional Random Field (MM-HCRF) is first utilized to identify if a candidate sequence contains a mitotic event, then Max-Margin Semi-Markov Model (MM-SMM) is used to segment the sequence into four consecutive stages based on the decision made by MM-HCRF. The four stages are defined as the visual appearance transition of cells during the mitosis: *Interphase, Start of mitosis, Formation of daughter cells, Separation of daughter cells*.

The MM-HCRF model has the same graphical model structure as the HCRF used in [6], but the parameter learning algorithm is based on the idea of latent support vector machine (LSVM) [36] in a max-margin framework.

MM-SMM is a random field model for temporal sequence segmentation by modeling the state transition between consecutive segments as a semi-Markov process. Same as the case with MM-HCRF, the model is trained under a max-margin framework. As shown in Fig. 9f, given an observation sequence $\mathbf{X} = \{x_t\}_{t=1}^{T}$, MM-SMM predicts the segmentation $\mathbf{S} = \{s_i\}_{i=1}^{S}$ of the sequence. Each segment label s_i consists of a pair of integers $s_i = (u_i, l_i)$, here u_i represents the index of the last frame in the segment i, and l_i is the label of the segment. The model also predicts the overall label of the sequence as y'

$$p(y'|\mathbf{X}, \gamma) = p(y'|\mathbf{S}, \gamma) \cdot p(\mathbf{S}|\mathbf{X}, \gamma), \tag{23}$$

here $p(y'|\mathbf{S}, \gamma) = 1$ when $\mathbf{S}$ contains only one continuous stage transition from *stage 1* to *stage 4*. In other cases, $p(y'|\mathbf{S}, \gamma) = 0$. Under this definition, Eq. (23) can be simplified as

$$\begin{aligned} p(y'|\mathbf{X}, \gamma) &= p(\mathbf{S}|\mathbf{X}, \gamma) \\ &= \frac{\exp(\gamma^T \cdot \psi(\mathbf{X}, \mathbf{S}))}{\sum_{\mathbf{S}'} \exp(\gamma^T \cdot \psi(\mathbf{X}, \mathbf{S}'))}. \end{aligned} \tag{24}$$

The potential function $\gamma^T \cdot \psi(\mathbf{X}, \mathbf{S})$ can be expanded as

$$\begin{aligned}\gamma^T \cdot \psi(\mathbf{X}, \mathbf{S}) &= \sum \gamma^T \cdot \psi(l_{i-1}, l_i, \mathbf{X}_{u_{i-1}:u_i}) \\ &= \sum \sum_{a,b \in L} \gamma_{a,b}^T \cdot \begin{bmatrix} \psi_1(\mathbf{X}_{u_{i-1}:u_i}) \\ \psi_2(\mathbf{X}_{u_{i-1}:u_i}) \end{bmatrix} \cdot 1_a(l_{i-1}) \cdot 1_b(l_i),\end{aligned} \tag{25}$$

where $\psi_1(\cdot)$ stands for the mean of feature vectors, $\psi_2(\cdot)$ stands for the standard deviation of feature vectors, $[\cdot]$ denotes vector concatenation.

5.3.7 Hidden State Conditional Neural Field (HSCNF)

For previous graphical model based methods, there are two open issues: 1. Feature representation: The random fields-based methods cannot deal with the non-linearity among features. 2. Model learning: Feature representation and model learning are separately treated. To tackle both problems, Su et al. [37] proposed a Hidden-State Conditional Neural Field model to introduce non-linearity feature transformation. The graphical structure of HSCNF is shown in Fig. 9g, where G is the logistic gate function used for non-linear transformation of the input observation. Given an input sequence $\mathbf{X}$, the conditional probability distribution of label y can be formulated as

$$p(y|\mathbf{x}) = \sum_h \exp(\Lambda(\mathbf{X}, y, h))/Z(\mathbf{X}), \tag{26}$$

where $Z(\mathbf{X}) = \sum_{y',h} \exp(\Lambda(\mathbf{X}, y, h))$ is a partition function. $\Lambda(\mathbf{X}, y, h) = \Phi(\mathbf{X}, y, h) + \Psi(\mathbf{X}, y, h)$. $\Phi(\mathbf{X}, y, h)$, $\Psi(\mathbf{X}, y, h)$ denote the observation function and transition function respectly

$$\Phi(\mathbf{X}, y, h) = \sum_t \sum_g w_{h_t,g} G(\theta_g \phi(\mathbf{X}, y, h, t)) \tag{27}$$

$$\Psi(\mathbf{X}, y, h) = \sum_j u_j \sum_t \psi(\mathbf{X}, y, h, t, t-1). \tag{28}$$

Discussion. In this section, we described an overall workflow for mitosis detection under hybrid framework and introduced advanced approaches for key modules in the workflow. Different from tracking based methods which detect mitosis through long term tracking, the hybrid methods construct candidate sequences in relatively shorter time durations. This could avoid the difficulty for robust long term tracking while making use of dynamic information during the mitosis.

Among these works, different graphical models have been employed to model the temporal dynamics of the mitosis. HMM and CRF are the generative and discriminative models for mitosis sequence labeling respectively. Based on CRF, HCRF uses sequence-wise label and frame-wise hidden states to model the latent dynamics between frames in mitosis sequences. EDCRF combines sequence-wise label and frame-wise label together to improve detection accuracy and locate the exact birth time of the daughter cells in the mitotic sequence. MM-SMM model can further segment the candidate sequence into four stages and get more precise location result by training on manually labeled dataset. Moreover, to overcome the difficulty of candidate sequence extraction under high cell confluence, TL-HCRF used a tracking free method for the pre-detection of the birth event candidate, which has a relatively more distinct visual appearance from regular cells, and two labeled (before or after birth event) HCRF for the subsequent sequence classification. Motivated by deep learning based methods, HSCNF model achieves non-linear feature transformation by introducing a logistic gate function between observations and hidden states in the HCRF model.

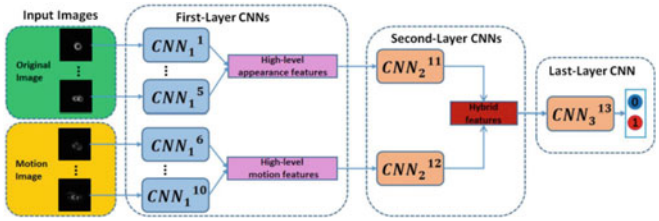


Fig. 11. Overview of HCNN architecture (from [23]). The architecture consists of three parts: First layer extract features from single frames in the sequence and merges them into higher level features (appearance and motion). Second layer takes two feature inputs and merges them into hybrid features. Third layer gets the final detection result based on the hybrid features.

Hybrid methods have been broadly used for mitosis event identification, cell birth event localization and mitotic sequence segmentation problems. We will compare the performance of these methods on the evaluation datasets in Section 7.

6 Deep Learning Based Methods

In recent years, deep learning has demonstrated superior performance on many computer vision problems such as image classification, object detection and segmentation. It is natural to consider employing deep learning methods for mitosis detection. The utilization of deep learning methods makes it possible to learn the intrinsic features of mitotic cell images directly from input. This is potentially more advantageous compared to hand-crafted features considering the deformable nature of cells in microscopy images. Thanks to the growing scale of datasets in phase contrast microscopy image sequences of cells, we have seen several deep learning based methods for mitosis detection, and their performance have achieved considerable improvement over previous approaches.

6.1 Hierarchical Convolutional Neural Network (HCNN)

Deep learning methods have shown strong performance for detecting mitosis from single images [38], [39], [40]. But in applications such as time-lapse phase contrast microscopy imaging, temporal dynamics between frames also provides important information for mitosis detection. Considering mitotic cells have considerable deformation in time-lapse phase contrast images, it is necessary to incorporate dynamic information into classification.

Mao et al. [23] proposed a deep-learning based approach to detect mitosis in phase contrast microscopy images which takes both the original images and the difference images between frames as input. The candidate sequences are first extracted using the method introduced in [21], then candidate sequences are identified using Hierarchical Convolutional Neural Network architecture. An overview of HCNN network is shown in Fig. 11.

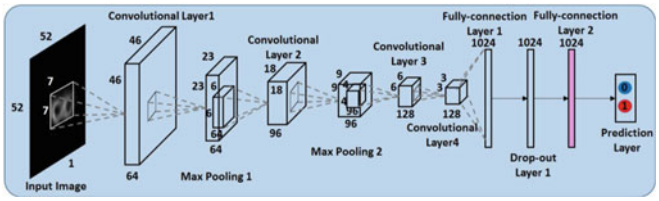


Fig. 12. The CNN architecture in the first part of HCNN (from [23]).

Fig. 13. The architecture of 3D-CNN network (from [24]). Number in each convolutional layer represents the number of kernels for this layer. All convolutional kernels are in size of $3 \times 3 \times 3$ with stride of 1, and pooling kernels are in size of $2 \times 2 \times 2$ with stride of 1.

The HCNN architecture consists of 3 parts: The first part contains 10 separate CNN layers denoted as CNN_1^1, $CNN_1^2, \ldots, CNN_1^{10}$. The structure and parameter of every CNN in the first part are the same, as described in Fig. 12. These 10 layers take two sets of input: 5 consecutive patch images in the candidate sequence, and 5 corresponding motion images computed using central finite difference. In the second part, 2 CNNs (denoted as CNN_2^{11}, CNN_2^{12}) are designed to take appearance features and motion features from the first layer respectively. The joint features are then fed into the third part which contains only one CNN (CNN_3^{13}) to make the final prediction of the candidate sequence.

Since there are 13 CNNs in HCNN, the number of parameters in the model is large relative to the limited amount of training data. Thus the training process is divided into two steps: each CNN in HCNN is trained independently during a pretraining phase in first step, and then the parameters of the entire structure are fine-tuned during an overall training process. In addition, the amount of data is augmented by rotating and slightly translating the original images. The HCNN method has achieved great success on the C3H10 dataset, and the prediction accuracy has exceeded many approaches utilizing traditional features by a large margin. This has once again demonstrated the effectiveness of deep learning methods with big data.

6.2 3D Convolutional Neural Network (3D-CNN)

Nie et al. [24] proposed an approach that applys a 3D-CNN network structure to learn spatiotemporal feature from phase contrast microscopy image sequences. Different from previous methods [23], [38] which only extract features from individual images, 3D-CNN learns the spatiotemporal information directly from volumetric regions of the image sequence. The network design of 3D-CNN architecture is shown in Fig. 13. This network contains five convolutional layers followed by two fully connected layers, which extracts the spatiotemporal feature as a 2,048 dimensional vector. This feature vector is then fed into an SVM classifier for detection.

The paper also discussed the mitosis detection performance of convolutional kernels with different temporal lengths. Experiments show that the performance get best when the temporal length is set to 3. 3D convolutional kernels with too short temporal length cannot extract sufficient dynamic information, and long temporal length makes the network complicated and suffer the overfitting problem when trained on the relatively scale limited image dataset.

Discussion. This section discusses two methods that employ deep learning to capture the spatiotemporal information from time-lapse phase contrast microscopy images of mitotic cells. With the data driven feature learning ability of the CNN network, both methods outperformed the approaches based on traditional hand-crafted feature, In the HCNN approach, the CNN network takes both the visual appearance and the motion information from image sequences as input. With the hierarchical architecture, HCNN merges appearance and motion information together for the final classification. This two-stream CNN network design has the ability to learn from both visual appearance and motion dynamics of image sequences. The 3D-CNN network learns the spatiotemporal information with 3D convolutional kernels directly from input image sequences. 3D-CNN provides more capacity and flexibility to extract spatiotemporal features, but as the temporal depth of the 3D kernel grows, the network becomes complicated quickly, making it harder to train on limited dataset without overfitting problem.

Although deep learning has improved the performance for mitosis detection, it suffers from a common problem in biomedical imaging applications: the network needs to be sufficiently complex to have the capacity to describe the visual and dynamic features of deformable cells; however, the scale of the datasets in biomedical imaging is typically limited, considering the higher reliance on expert knowledge for annotation. Complicated network structures trained on these datasets are prone to overfitting.

7 Performance Comparison

In this section, we compare the performance of some state-of-the-art works on mitosis detection in terms of mitosis identification and localization. Experiments in this section were reported in [2], [6], [21], [23].

7.1 Evaluation Method

To evaluate the accuracy of mitosis identification, three categories of detection results are defined:

1) *True Positive* (TP): The mitotic sample that is identified as positive.
2) *False Positive* (FP): The non-mitotic sample that is mistakenly identified as positive one.
3) *False Negative* (FN): The mitotic sample that is mistakenly identified as negative one.

Fig. 14 gives three sample sequences for each category of detection results.

Based on these definitions, the mitosis detection result can be evaluated by the $Precision$, $Recall$ and $Fscore$

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ Fscore &= \frac{2 \times Precision \times Recall}{Precision + Recall}. \end{aligned} \tag{29}$$

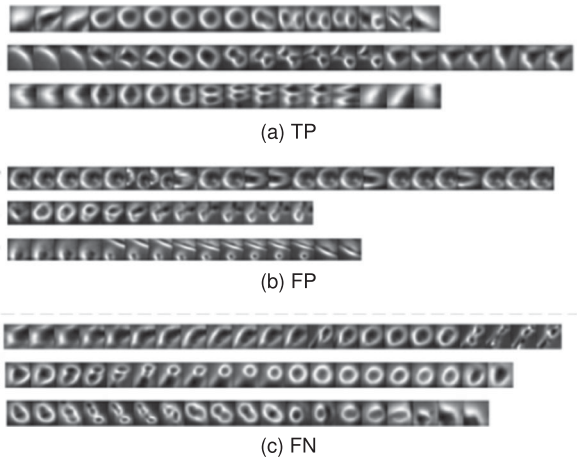


Fig. 14. Samples of true positive, false positive and flase negative detected sequences (from [2]).

To further evaluate temporal localization accuracy for sequence candidates, only positive samples whose frame errors (bias between localized result and ground truth location) are lower than the pre-set threshold are considered as *TP* samples. Performance is evaluated with different pre-set thresholds to compare the temporal localization accuracy between different approaches.

7.2 Analysis of Results

7.2.1 Mitosis Identification

The performance of mitosis identification on the C3H10 dataset is compared among different approaches in Table 2.

Table 2 shows that the HCNN method has achieved the best result by taking advantage of a deep learning framework. MM-HCRF+MM-SMM comes in second with a two-step identification mechanism: MM-HCRF is used to identify candidate sequence as mitosis or not, then MM-SMM further segments the candidate sequence into different stages and in turn helps to improve the identification result from MM-HCRF by reducing false positives. With the temporal stage transition constraint enforced by MM-SMM, the combined method performed notably better than using MM-HCRF alone (by 4.6 percent of averaged F score).

Compared with MM-HCRF, EDCRF has achieved a higher performance (by 1.7 percent of averaged F score). This result indicates that the information of the birth event timing model in EDCRF is helpful to identify the mitosis sequence. In contrast, MM-HCRF could not utilize such information due to its limitation of modeling frame level annotations.

CRF and HMM methods both had much lower performance (only higher than the SVM method). The limitation of the CRF model is that it cannot capture the intermediate structures of the mitosis using hidden states. For the HMM model, it is limited by its assumption that observations are conditionally independent, resulting in the inability to accommodate long range dependencies among frames in sequences.

The SVM method has the worst performance for mitosis identification. The reason is that the SVM method is not capable of modeling temporal dynamics among frames in the mitosis sequences, thus the identification could only be done frame-by-frame without benefiting from the temporal context.

TABLE 2
Comparison of Mitosis Identification Performance of Different Works on C3H10 Dataset

	Precision (%)	Recall (%)	F score (%)
HCNN [23]	$\mathbf{99.1 \pm 0.8}$	$\mathbf{97.2 \pm 2.4}$	$\mathbf{98.6 \pm 1.3}$
MM-HCRF+MM-SMM [2]	95.8 ± 1.0	88.1 ± 3.7	91.8 ± 2.0
MM-HCRF [2]	82.8 ± 2.4	92.2 ± 2.4	87.2 ± 1.6
EDCRF [21]	91.3 ± 4.0	87.0 ± 4.8	88.9 ± 0.7
CRF [20]	90.5 ± 4.7	75.3 ± 9.6	81.5 ± 4.4
HMM [19]	83.4 ± 4.9	79.4 ± 8.8	81.0 ± 3.4
SVM	68.0 ± 3.4	96.0 ± 4.2	79.5 ± 1.7

7.2.2 Mitosis Localization

Table 4 compares averaged frame error between different methods, on both the C3H10 and the C2C12 datasets. Since HCRF is not capable of localizing cell division temporally, SVM classifier or CRF model is used for temporal localization after the sequence being identified by the HCRF model (named as HCRF+SVM and HCRF+CRF methods in the table). In Table 4, the MM-HCRF+MM-SMM method has achieved the most accurate mitosis localization result with an averaged location error of 0.73 frames. Compared with other models, the MM-SMM model makes use of annotations of four stages during mitosis (Section 5.3.6). This indicates that the information of all stages during mitosis is useful for temporal localization.

In addition, the EDCRF model achieved significantly better localization result compared with the HCRF+SVM and HCRF+CRF methods. The cell birth event localization in the HCRF+SVM method is done by frame-by-frame classification using SVM, and no dynamic information is utilized in this process. For the HCRF+CRF method, the localization is done using the CRF model which assumes that the input frames are conditionally independent, which also does not capture the frame-to-frame dynamics of the process.

Table 3 shows the precision, recall and F score with different frame error thresholds (1, 3, 5, 10). Among these results, the deep learning based HCNN method achieves the best performance when the threshold is low.

8 Conclusion and Future Directions

In this paper, we discussed four categories of contemporary approaches for mitosis detection and compared the performances of several of them. Among these methods, tracking based methods detect mitosis based on the result of long-term cell tracking, while tracking free methods classify mitotic candidate regions based on the visual features without tracking. Hybrid methods combine these two categories by applying graphical models on candidate image sequences extracted with short-term tracking. For the last category, we described recent deep learning based methods and discussed their advantages and disadvantages respectively. After comparing these methods, we drawn some conclusions as well as insights on potential future directions for mitosis detection:

- Compared with static image dataset, temporal dynamic information among time-lapse image sequences can provide important information to improve the mitosis detection performance. While long-term tracking is challenging and unnecessary

TABLE 3
Accuracy of Temporal Localization for Cell Birth Event with Different Frame Error Threshold, Compared on C3H10 and C2C12 Dataset

method	threshold	C3H10			C2C12		
		Precision (%)	Recall (%)	F score (%)	Precision (%)	Recall (%)	F score (%)
HCRF+CRF	1	58.3 ± 4.4	56.5 ± 3.6	57.4 ± 3.7	68.7	65.0	66.8
	3	77.1 ± 5.5	74.9 ± 7.1	75.9 ± 6.0	78.8	74.6	76.7
	5	85.3 ± 3.6	82.7 ± 3.9	83.9 ± 2.8	83.8	79.3	81.5
	10	88.4 ± 2.0	85.7 ± 3.4	87.0 ± 1.2	89.2	84.5	86.8
HCRF+SVM	1	60.4 ± 5.2	58.5 ± 5.1	59.4 ± 4.9	55.0	52.0	53.5
	3	79.5 ± 4.5	77.3 ± 6.4	78.3 ± 5.2	76.5	72.4	74.4
	5	87.2 ± 1.0	84.7 ± 4.3	85.9 ± 1.9	87.3	82.6	84.9
	10	88.7 ± 1.9	86.0 ± 3.3	87.3 ± 1.0	90.9	86.1	88.4
EDCRF	1	74.0 ± 6.7	70.3 ± 3.5	72.0 ± 4.1	88.0	82.8	85.3
	3	86.3 ± 4.4	82.2 ± 4.3	84.0 ± 1.5	92.5	87.0	89.7
	5	89.9 ± 3.7	85.7 ± 5.2	87.6 ± 1.4	94.0	88.4	91.1
	10	91.0 ± 4.2	86.7 ± 4.7	88.6 ± 0.9	94.7	89.0	91.8
MM-HCRF+MM-SMM	1	79.8 ± 3.4	73.3 ± 2.4	76.4 ± 2.7	–	–	–
	3	91.1 ± 2.2	83.8 ± 3.7	87.3 ± 2.8	–	–	–
	5	94.7 ± 0.5	87.1 ± 2.8	90.8 ± 1.7	–	–	–
	10	95.8 ± 1.0	88.1 ± 3.1	91.8 ± 2.0	–	–	–
HCNN	1	92.8 ± 1.3	93.1 ± 1.1	93.0 ± 0.4	–	–	–
	3	96.6 ± 1.1	94.9 ± 2.0	95.8 ± 0.8	–	–	–
	5	98.3 ± 1.2	96.9 ± 1.6	97.6 ± 0.9	–	–	–
	10	99.1 ± 0.8	97.2 ± 2.4	98.2 ± 1.3	–	–	–

for mitosis detection, candidate sequences generated with short-term tracking are useful for mitosis detection with a specially designed graphical model.

- The recent proposed deep learning based methods have shown advantage over traditional apporaches on mitosis detection problem. Compared with handcrafted features, deep learning method could learn more discriminative and data relevant features directly from the image data. This allows them to better identify mitotic cells with irregular shapes and deformations.
- Future directions for mitosis detection: Considering the success of deep learning methods and the importance of dynamic information for mitosis detection, it is natural to consider exploring longer temporal context in deep learning based methods. For current approaches, HCNN uses motion images calculated between consecutive frames to capture temporal information, and 3D-CNN kernels only perform well with a temporal depth of 3. It might be useful to combine the CNN network with the Recurrent Neural Network (RNN) or graphical model to explore this potentional.

TABLE 4
Averaged Temporal Localization Error of Daughter Cell Birth Event (Absolute Frame Difference between Localization Result and Ground Truth) on C3H10 and C2C12 Datasets

Method	C3H10	C2C12
HCRF+CRF	1.438 ± 1.826	1.189 ± 2.120
HCRF+SVM	1.273 ± 1.528	1.685 ± 1.769
EDCRF	0.828 ± 1.334	0.482 ± 0.957
MM-HCRF+MM-SMM	$\mathbf{0.73 \pm 1.29}$	–

- Working with limited training data in biomedical imaging: Although deep learning architectures have shown promising potential for biomedical imaging, the limitation of dataset scale often leads to overfitting. Annotating biomedical images requires expert knowledge, which makes the generation of labels on large scale datasets impractical. It is necessary to develop approaches that could generate expert-level annotations from crowdsourcing, or analyze these images using unsupervised or semi-supervised methods.

ACKNOWLEDGMENTS

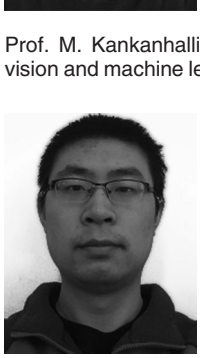
This work was supported in part by the National Natural Science Foundation of China under Grant 61472275, Grant 61572356, and Grant 61100124, in part by the Tianjin Research Program of Application Foundation and Advanced Technology under Grant 15JCYBJC16200, in part by the China Scholarship Council under Grant 201506255073, Grant 201506250046, and in part by the Elite Scholar Program of Tianjin University under Grant 2014XRG-0046.

REFERENCES

[1] T. Kanade, et al., "Cell image analysis: Algorithms, system and applications," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2011, pp. 374–381.
[2] A. Liu, K. Li, and T. Kanade, "A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations," *IEEE Tran. Med. Imag.*, vol. 31, no. 2, pp. 359–369, Feb. 2012.
[3] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 cells classification methods," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1878–1889, Oct. 2013.
[4] M. Veta, et al., "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, 2015.

[5] F. Zernike, "How I discovered phase contrast," *Science*, vol. 121, no. 3141, pp. 345–349, 1955.
[6] A. Liu, K. Li, and T. Kanade, "Mitosis sequence detection using hidden conditional random fields," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano Macro.*, 2010, pp. 580–583.
[7] F. Yang, M. A. Mackey, F. Ianzini, G. Gallardo, and M. Sonka, "Cell segmentation, tracking, and mitosis detection using temporal context," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention.*, 2005, pp. 302–309.
[8] O. Debeir, P. Van Ham, R. Kiss, and C. Decaestecker, "Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes," *IEEE Trans. Med. Imag.*, vol. 24, no. 6, pp. 697–711, Jun. 2005.
[9] K. Thirusittampalam, M. J. Hossain, O. Ghita, and P. F. Whelan, "A novel framework for cellular tracking and mitosis detection in dense phase contrast microscopy images," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 642–653, May. 2013.
[10] M. Liu, Y. He, Y. Wei, and P. Xiang, "Plant cell tracking using kalman filter based local graph matching," *Image Vis. Comput.*, vol. 60, pp. 154–161, 2017.
[11] A. Liu, K. Li, and T. Kanade, "Spatiotemporal mitosis event detection in time-lapse phase contrast microscopy image sequences," in *Proc. IEEE Int. Conf. Multimedia Expo*. 2010, pp. 161–166.
[12] K. Li and T. Kanade, "Nonnegative mixed-norm preconditioning for microscopy image segmentation," in *Proc. Int. Conf. Inform. Process. Med. Imag.*, 2009, vol. 21, pp. 362–373.
[13] M. Marcuzzo, T. Guichard, P. Quelhas, A. M. Mendonça, and A. Campilho, "Cell division detection on the arabidopsis thaliana root," in *Proc. Iberian Conf. Pattern Recognition Image Anal.*, 2009, pp. 168–175.
[14] A. Liu, T. Hao, Z. Gao, Y. Su, and Z. Yang, "Nonnegative mixed-norm convex optimization for mitotic cell detection in phase contrast microscopy," *Comput. Math. Methods Med.*, vol. 2013, Article ID 176272, p. 10, 2013, doi: 10.1155/2013/176272.
[15] C. Huang and H. Lee, "Automated mitosis detection based on exclusive independent component analysis," in *Proc. 21st Int. Conf. Pattern Recognition.*, 2012, pp. 1856–1859.
[16] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Computer vision tracking of stemness," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.: From Nano Macro.*, 2008, pp. 847–850.
[17] J. Wu, S. C. Brubaker, M. D. Mullin, and J. M. Rehg, "Fast asymmetric learning for cascade face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 369–382, Mar. 2008.
[18] P. Quelhas, A. M. Mendonça, and A. Campilho, "Optical flow based arabidopsis thaliana root meristem cell division detection," in *Proc. Int. Conf. Image Anal. Recognition.*, 2010, pp. 217–226.
[19] G. M. Gallardo, F. Yang, F. Ianzini, M. Mackey, and M. Sonka, "Mitotic cell recognition with hidden markov models," in *Med. Imag.*, vol. 5367, pp. 661–668, 2004.
[20] L. Liang, X. Zhou, F. Li, S. T. Wong, J. Huckins, and R. W. King, "Mitosis cell identification with conditional random fields," in *Proc. Life Sci. Syst. Appl. Workshop*, 2007, pp. 9–12.
[21] S. Huh, R. Bise, M. Chen, T. Kanade, D. F. E. Ker, "Automated mitosis detection of stem cell populations in phase-contrast microscopy images," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 586–596, Mar. 2011.
[22] S. Huh and M. Chen, "Detection of mitosis within a stem cell population of high cell confluence in phase-contrast microscopy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition*, 2011, pp. 1033–1040.
[23] Y. Mao and Z. Yin, "A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2016, pp. 685–692.
[24] W. Nie, W. Li, A. Liu, T. Hao, and Y. Su, "3D convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition Workshops*, 2016, pp. 55–62.
[25] Z. Yin, K. Li, T. Kanade, and M. Chen, *Understanding the Optics to Aid Microscopy Image Segmentation*. Berlin, Heidelberg: Springer, pp. 209–217, 2010.
[26] M. Born and E. Wolf, "Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light," *Physics Today*, vol. 53, no. 10, pp. 77–78, 2007.
[27] J. Silvela and J. U. Portillo, "Breadth-first search and its application to image processing problems," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1194–1199, Aug. 2001.
[28] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, and Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[29] P. L. Rosin, "Unimodal thresholding," *Pattern Recognition*, vol. 34, no. 11, pp. 2083–2096, 2001.
[30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition*, vol. 1, pp. 886–893, Jun. 2005.
[31] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[32] A. Liu, K. Li, and T. Hao, "A hierarchical framework for mitosis detection in time-lapse phase contrast microscopy image sequences of stem cell populations," *Medical Imaging*, ch. 17, Dec. 2011.
[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[34] L. M. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
[35] L. P. Morency, A. Quattoni, and T. Darrell, "Latent-dynamic discriminative models for continuous gesture recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition*, 2007, pp. 1–8.
[36] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition*, 2008, pp. 1–8.
[37] Y. Su, J. Yu, A. Liu, and Z. Gao, "Cell type-independent mitosis event detection via hidden-state conditional neural fields," in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, pp. 222–225, 2014.
[38] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2013, pp. 411–418.
[39] H. Chen, Q. Dou, X. Wang, J. Qin, and P.-A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proc. 13th AAAI Conf. Artificial Intell.*, 2016, pp. 1160–1166.
[40] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1313–1321, May 2016.
[41] A. Liu, J. Tang, W. Nie, and Y. Su, "Multi-grained random fields for mitosis identification in time-lapse phase contrast microscopy image sequences," *IEEE Trans. Med. Imag.*, 2017, doi: 10.1109/TMI.2017.2686705.

An-An Liu (M'10) received the BEng and PhD degrees from Tianjin University, Tianjin, China. He is currently an associate professor in the School of Electrical and Information Engineering, Tianjin University. He was a visiting scholar with the Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, from 2008 to 2009, where he researched with Prof. T. Kanade. He was a visiting scholar in the School of Computing, National University of Singapore, Singapore, in 2016, where he researched with Prof. M. Kankanhalli. His current research interests include computer vision and machine learning.



Yao Lu is working toward the PhD degree in the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China. He was a visiting scholar in the Department of Electrical and Computer Engineering, State University of New York at Albany, Albany, New York, where he researched with Prof. Mei Chen. His current research interests include computer vision and machine learning.

Mei Chen received the BS and MS degrees from Tsinghua University, Beijing, China, and the PhD degree in robotics from the School of Computer Science, Carnegie Mellon University. She is an associate professor in the Department of Electrical and Computer Engineering, State University of New York at Albany.

Yu-Ting Su received the BEng and PhD degrees from Tianjin University, Tianjin, China. He is currently a professor in the School of Electrical and Information Engineering, Tianjin University. He was a visiting scholar with Case Western Reserve University, Cleveland, Ohio. His current research interests include multimedia content analysis and security.