Medical Image Analysis 57 (2019) 32–43

Contents lists available at ScienceDirect

Medical Image Analysis

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

Cell mitosis event analysis in phase contrast microscopy images using deep learning

Yunxiang Mao, Liang Han, Zhaozheng Yin*

Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, 65401 United States

ARTICLE INFO

Article history:
Received 7 July 2018
Revised 12 June 2019
Accepted 20 June 2019
Available online 22 June 2019

Keywords:
Cell mitosis event analysis
Low-Rank matrix recovery
Convolutional neural networks
Long short term memory

ABSTRACT

In this paper, we solve the problem of mitosis event localization and its stage localization in time-lapse phase-contrast microscopy images. Our method contains three steps: first, we formulate a Low-Rank Matrix Recovery (LRMR) model to find salient regions from microscopy images and extract candidate patch sequences, which potentially contain mitosis events; second, we classify each candidate patch sequence by our proposed Hierarchical Convolution Neural Network (HCNN) with visual appearance and motion cues; third, for the detected mitosis sequences, we further segment them into four temporal stages by our proposed Two-stream Bidirectional Long-Short Term Memory (TS-BLSTM). In the experiments, we validate our system (LRMR, HCNN, and TS-BLSTM) and evaluate the mitosis event localization and stage localization performance. The proposed method outperforms state-of-the-arts by achieving 99.2% precision and 98.0% recall for mitosis event localization and 0.62 frame error on average for mitosis stage localization in five challenging image sequences.

1. Introduction

Analyzing the proliferative behavior of stem cells in vitro plays an important role in many biomedical applications, such as drug discovery, tissue engineering and stem cell manufacturing. The key to quantitatively analyze the proliferative behavior of stem cells depends on accurate enumeration and localization of the occurrences of *mitosis*, which is the process whereby the genetic material of an eukaryotic cell is equally divided, resulting in daughter cells. Microscopy imaging techniques for measuring the cell proliferation have been developed for many years. Most of the analysis methods use fluorescent, luminescent or colorimetric microscopy images which are acquired by invasive methods, such as staining cells with fluorescent dyes and radiating them with the specific wave-length light. The invasive method damages cells' viability or kills cells, which is not suitable for continuously monitoring the cell proliferation process. Phase-contrast microscopy, as a non-invasive imaging modality, offers the possibility to persistently monitor cells' behavior in the culturing dish without altering them (Li et al., 2008). For small-scale research studies, manual enumeration and localization of mitosis events by biologists may be considered. However, when it comes to the high-throughput experiments, analysis of these images becomes an arduous process which involves many hours of human inspection. Thus, computer-aided *mitosis detection* in phase-contrast microscopy images from high-throughput biological experiments has been one cornerstone of cell behavior analysis.

In fact, the computer vision based mitosis detection contains two tasks: (1) mitosis event localization and (2) mitosis stage localization. Given a microscopy image sequence, the mitosis event localization refers to identify where and when a mitosis event happens in the sequence. For the mitosis stage localization, we are trying to identify different mitotic stages within a mitotic cell image sequence. The quantification of biological metrics will benefit from accurately localizing the time of each stage and biologists are allowed to assess different factors that impact the length of time a cell spends in each stage of the mitosis.

Based on the visual transition of cell appearance, the process of a mitosis event can be segmented into four consecutive stages as shown in Fig. 1: (1) *interphase*, when a mitotic cell reduces its migration speed but its appearance still remains normal, (2) *start of mitosis*, when a mitotic cell increases its brightness, shrinks its size, and rounds up, (3) *formation of daughter cells*, when two daughter cells are visible but they are still attached together and appear like a number "8", and (4) *separation of daughter cells*, when two daughter cells separate completely.

In this paper, we define the mitosis event localization as the classification of a patch sequence containing a mitosis process or not, and the stage localization as the localization of the time at which Stage 2–4 begin in one mitosis sequence.

* Corresponding author.
E-mail addresses: ym8r8@mst.edu (Y. Mao), lh248@mst.edu (L. Han), yinz@mst.edu (Z. Yin).

https://doi.org/10.1016/j.media.2019.06.011
1361-8415/

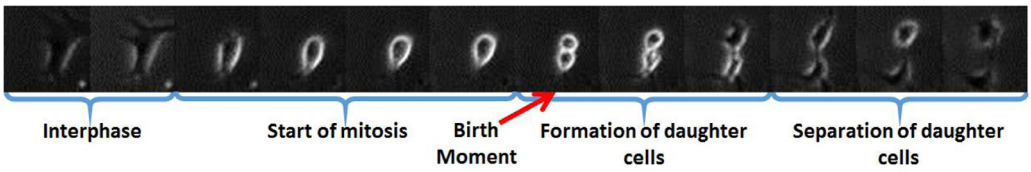

Fig. 1. The process of a mitosis event.

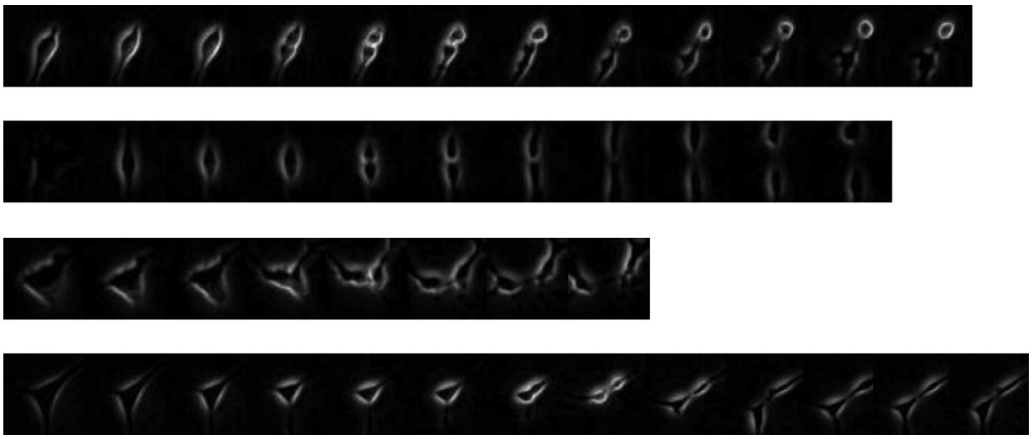

Fig. 2. Different mitosis sequences which show large shape and appearance variations.

The task of mitosis event localization is to find the birth moment of a mitotic cell spatially and temporally in the given time-lapse sequences. As shown in Fig. 2, mitotic cells have irregular shapes and appearances. Some non-mitotic image sequences may have the similar appearance as mitosis sequences, which increases the chance to cause false positives in the mitosis event analysis. Thus, it is not reliable to simply apply some morphological operations on the phase contrast images to search for the shape like the number "8".

1.1. Related work

1.1.1. Traditional mitosis detection

Several cell mitosis detection methods are based on object tracking in phase-contrast microscopy images. Debeir et al. (2005) combine several model-based mean-shift processes to track migrating cells. In Padfield et al. (2009), Padfield et al. segment the nuclei with a shape/size constraint and use an Euclidean distance metric to link different phases of the cell cycle. Al-Kofahi et al. (2006) segment cells in each image of the sequence, and adopt a multiple-object matching method which measures a number of cell attributes such as size, shape and location to track cells. Li et al. (2007) exploit a geometric active contour model to track detected cells. Liang et al. (2008) segment the cell nuclei from the background, link the nuclei into cell sequences and then utilize a conditional random field (CRF) model (Lafferty et al., 2001) with shape and texture features of the segmented nuclei to identify mitosis events. With the goal of tracking cell movements over time, the problem of mitosis detection in these papers is solved based on object tracking algorithms or volumetric image segmentation. The mitoses are identified based on the temporal progression of cell features or the connection between the segmented mother and daughter cells. However, these mitosis event localization approaches heavily depend on the long-term object tracking performance, which itself is a very challenging task.

Considering the drawback of tracking-based mitosis detection, tracking-free approaches have been proposed to detect mitosis events directly in an image sequence. Huang and Lee (2012) propose an algorithm called eXclusive Independent Component Analysis (XICA) which focuses on the components of differences between two classes of training patterns rather than the major components. They classify the given testing pattern by computing the residual of the relative exclusive basis set. Since the mitosis is a dynamic process, the performance of mitosis detection would benefit from taking advantage of the temporal dynamic information in the evolution of visual patterns. These methods usually consist of three steps: candidate detection, feature extraction and classification. In the candidate detection, which aims to produce image patches that potentially contain mitosis events, thresholding and/or morphological operations are typically applied. Regarding to feature extraction and classification, Gallardo et al. (2004) adopt a hidden Markov model (HMM) to classify candidates based on temporal patterns of cell shape and appearance features. Li et al. (2008) apply a cascade classifier framework (Wu et al., 2008) to classify volumetric sliding windows of an image sequence based on 3D haar-like features (Viola and Jones, 2004). The proposed method is efficient since it only needs one sequential scan through the image sequence with a trained classifier. However, the coarse resolution of the sliding temporal window may limit the localization precision.

Liu et al. (2010) propose an approach based on Hidden Conditional Random Fields (HCRF) (Quattoni et al., 2007) in which mitosis candidate patch sequences are extracted through a 3D seeded region growing method, then HCRF is trained to classify each candidate patch sequence. This method does not rely on object tracking algorithms and achieves good performance on C3H10T1/2 stem cell datasets. Since only one label is assigned to a patch sequence, this HCRF-based approach is able to identify if a patch sequence contains mitosis or nor, but it can not accurately localize the birth moment of the mitosis event in the patch sequence.

A few extensions have been made on the HCRF-based approach. Huh et al. (2011) propose an Event-Detection CRF (EDCRF) in which each patch in a candidate sequence is assigned with one label. The birth moment of the mitosis event is determined based on the observation that if there exists a change from "before mitosis" to "after mitosis" label. Liu et al. (2012) utilize a maximum-margin learning framework for training the HCRF and propose a semi-Markovian model to localize mitosis events.

1.1.2. Deep-learning based mitosis detection

The previous approaches rely on handcrafted image features. Deep Convolutional Neural Network (DCNN) which is capable of learning feature representations from big data and modeling the large variation among the data, has shown its effectiveness on object detection and classification. The learned kernels in DCNN are very effective feature extractors compared with handcrafted feature extractors. In order to improve the performance of DCNN on the challenging ImageNet dataset, Yan et al. (2015) propose a hierarchical deep convolutional neural network which classifies the input images with several components within the architecture. The low-level features of an input image are extracted through the lower layers, and then the classification of the input image is done followed by a coarse-to-fine approach. Sun et al. (2016) propose a hybrid deep learning network for the task of face verification. The input to their deep model is multiple pairs of different subregions in the two original images to be compared. The final output is binary which indicates whether the two original images come from the same person. Ciresan et al. (2013) utilize the DCNN as a pixel classifier for mitosis detection in individual breast cancer histology images. During the histology, the histologic specimens are stained and sandwiched between a glass microscope slide and coverslip. So this DCNN method is not suitable for detecting continuous mitosis events in the time-lapse phase-contrast microscopy image sequences. The latest CNN-based methods (Mao and Yin, 2016; Nie et al., 2016) achieve good performance on the task of mitosis event localization. Mao et al. build a hierarchical convolutional neural network in which both appearance and motion temporal information are utilized to detect the birth moment of a mitosis sequence. Compared with traditional CNN which only takes one single image as the input, 3D CNN (Ji et al., 2013; Tran et al., 2015) extracts temporal features though its 3D convolutional kernels. Nie et al. (2016) design several different 3D CNN architectures and demonstrate that 3D CNN outperforms the 2D CNN features and other hand-crafted features.

The previous CNN-based methods only accept a fixed-size vector as the input and produce a fixed-size vector as the output, e.g. probabilities of different classes. However, the length of extracted sequences varies. Furthermore, if the models take fixed-size input and output one label for the patch sequence, they are not able to perform the task of localizing different mitosis stages. Long-term Short Memory (LSTM) (Hochreiter and Urgen Schmidhuber, 1997), which is able to address variant-length inputs, is widely used in the natural language processing. It can be adapted to many-to-many, many-to-one, and one-to-many models according to different tasks (Donahue et al., 2017). For the task of mitosis stage localization, the many-to-many model can be utilized to output one label for each image in the candidate sequence to label its stage.

1.2. Our proposal and system overview

Most of the previous mitosis detection approaches either use handcrafted features or consider a single image as the input of DCNN architectures. In this paper, we try to discover the power of deep learning methods in the task of mitosis analysis.

If we attempt to detect mitosis events by a single image, we may lose the visual appearance change context during the whole process of a mitosis event. Furthermore, motion information hidden in the continuous image sequence can also aid the detection of mitosis event. In the work of Su et al. (2015), 3D objects recognition is achieved by a multi-view CNN. Each single CNN in the first layer takes an image of the object captured in one aspect as the input. After the first-layer CNNs, a single CNN is adopted to converge the features to predict the label of objects. Inspired by their work, a Hierarchical Convolutional Neural Network (HCNN) (Mao and Yin, 2016) is proposed for the task of mitosis detection, which utilizes the visual appearance change information and motion information in continuous microcopy images. Although the (HCNN) performs well in mitosis event detection, it lacks the ability of mitosis stage localization. The BLSTM (Mao and Yin, 2017) is able to solve these two problems simultaneously but the performance of mitosis event detection is not as good as HCNN and also the performance of mitosis stage localization can be further improved.

In this paper, our contribution is that we propose a new system in order to achieve the best performance of each single task of mitosis event detection and stage localization. The overflow of our framework is shown in Fig. 3. The system consists of three consecutive steps: first, a Low-Rank Matrix Recovery (LRMR) model is formulated to find the salient regions in the input image sequence and extract candidate patch sequences that are likely to contain

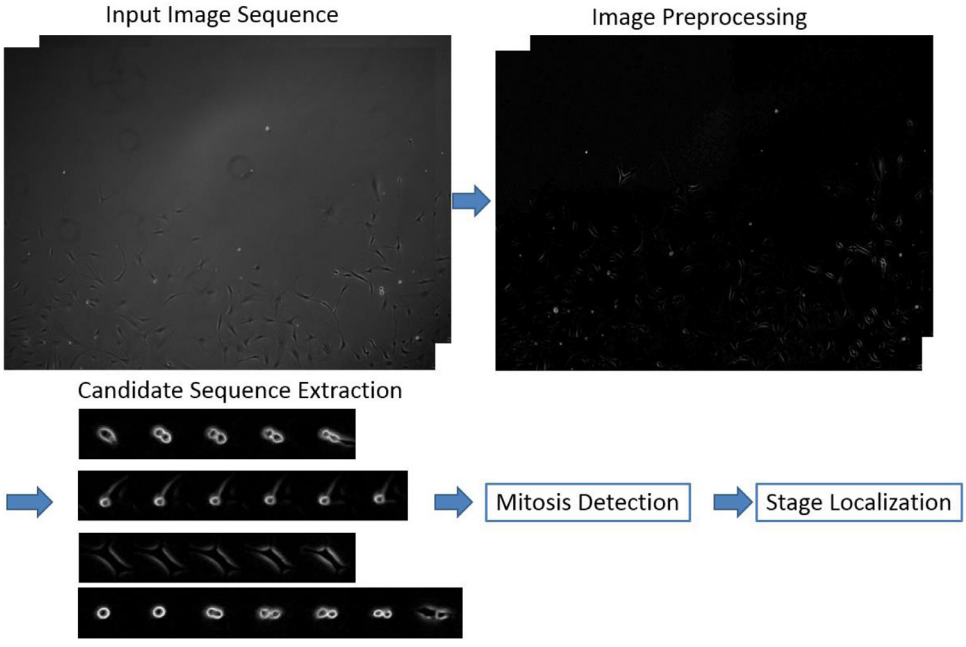

Fig. 3. Overview of our proposed method.

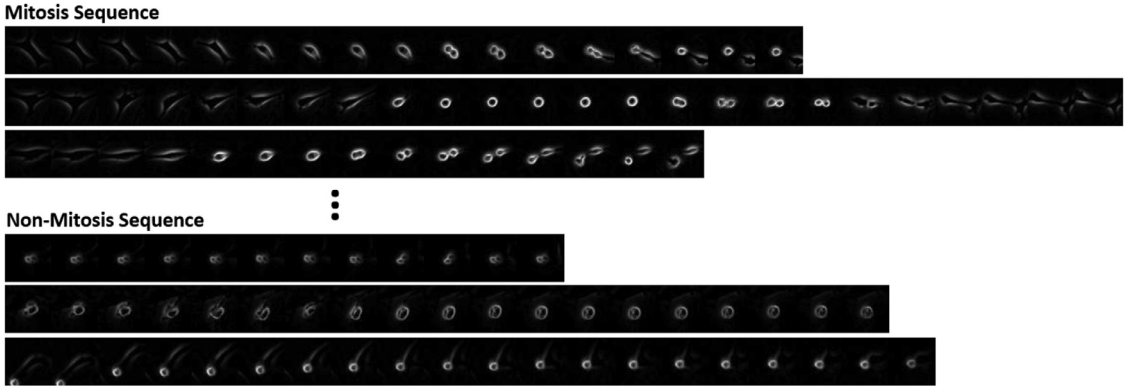

Fig. 4. Samples of extracted candidate sequences.

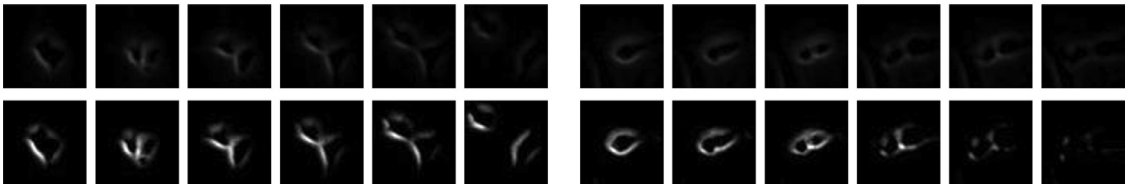

Fig. 5. The comparison of mitosis sequences generated by subtracting mean image and LRMR model. The first row is the mitosis image sequence generated by simply subtracting mean image. The Second row is the mitosis image sequence generated by LRMR model. By LRMR model, the intensity of mitosis cells are kept.

mitosis events; second, each candidate patch sequence is classified by the Hierarchical Convolutional Neural Network (HCNN) based on its temporal appearance and motion information; third, for those sequences which contain mitosis events, they are further segmented into different stages by a new Two-stream Bidirectional Long-Short Term Memory (TS-BLSTM). The next three sections will describe the three steps in details.

2. Mitosis candidate extraction

The first step of mitosis detection is to extract mitosis candidate sequences from the input time-lapse image sequence. The mitosis candidate extraction aims to find region-of-interest (ROI) that potentially contains a mitosis event. This step serves to reduce the search space. As a result, the subsequent steps can be more efficiently conducted, while maintaining mitosis detection accuracy. Fig. 4 shows some examples of candidate patch sequences extracted by our method. Our proposed mitosis candidate extraction consists of two steps: (1) salient region detection and (2) image patch sequence retrieval.

2.1. Salient region detection

Traditional search schemes adopt a sliding window fashion, by which the detectors need to search and classify image patches at all locations. The exhaustive search process increases the search space and potentially has a large possibility of errors. Phase contrast microscopes convert the minute phase shifts, caused by transparent specimens regarding to the illuminating light source, into variations in light amplitudes which can be observed by naked eyes or captured by cameras. Due to the optical principle and the inherent imperfections of the conversion process, phase contrast images contain artifacts such as halos and shade-off. If we are able to only focus on regions where mitotic cells are highly likely to appear, rather than the artifact regions or background regions, we will be able to reduce the search space. In (Huh et al., 2011; Mao and Yin, 2016), the average image of original or illumination-corrected images in the given sequence is computed, which is then subtracted from each image. By this procedure, artifacts from the stationary bright non-cell background are removed. However, this procedure is only able to remove the stationary artifacts. Furthermore, since the intensity values of mitotic cells are decreased when each image is subtracted from the average image, it may potentially harm the performance of later classification Fig. 5. The process of a mitosis event contains large intensity and shape changes in the observed microscopy images, thus in this section we are interested in finding salient regions with large intensity and shape changes while maintaining the intensity values of the original images.

2.1.1. Problem formulation

Given a time-lapse phase-contrast microscopy image sequence, which contains non-mitotic cells, mitotic cells, and artifacts, we are trying to find regions that are most likely to contain mitotic cells from the phase-contrast microscopy images first. The image data can be modeled as:

$$\mathcal{M} = \mathcal{L} + \mathcal{S} + \mathcal{N} \tag{1}$$

where $\mathcal{M} \in \mathbb{R}^{m\times n\times p}$ is the time-lapse phase-contrast microscopy image sequence, $\mathcal{L} \in \mathbb{R}^{m\times n\times p}$ is the image sequence containing stationary artifacts and non-mitotic cells, $\mathcal{N} \in \mathbb{R}^{m\times n\times p}$ is the Gaussian noise image sequence, and $\mathcal{S} \in \mathbb{R}^{m\times n\times p}$ is the image sequence containing mitosis candidates. m and n are the number of rows and columns of the microscopy image, and p is the number of images in the image sequence.

We first transfer the image data matrices $\mathcal{M}$, $\mathcal{L}$, $\mathcal{S}$, and $\mathcal{N}$ to the corresponding Casorati matrices (a matrix each of whose columns is a vectorized image of the image data), $M \in \mathbb{R}^{mn\times p}$, $L \in \mathbb{R}^{mn\times p}$, $S \in \mathbb{R}^{mn\times p}$, and $N \in \mathbb{R}^{mn\times p}$, as shown in Fig. 6. Now, from Eq. (1) we have

$$M = L + S + N \tag{2}$$

2.1.2. Sparse property of mitosis image S

From the phase-contrast microscopy images, we can see that only a small portion of the image contains the mitosis candidates,

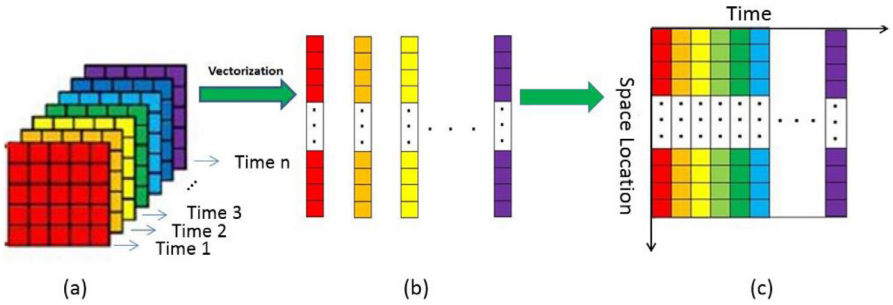

Fig. 6. Formulation of Casorati matrix. (a) an original phase-contrast microscopy image sequence, (b) we vectorize each image to get the corresponding column vectors, (c) these vectors are combined into a matrix.

therefore, the matrix S is sparse, i.e., only a few elements of S are nonzero. In order to get the images which only contain mitosis candidates, we need to estimate the image S from the observed microscopy image M.

2.1.3. Low-rank property of artifact and non-mitotic cell image L

As shown in Fig. 6(c), the Casorati matrix of the phase-contrast microscopy image sequence data has two dimensions, the spatial domain of each image and time domain of all the images. From the phase-contrast microscopy images we can see that the stationary artifacts appear at the same location of all the images. Accordingly, there exists high correlations among the time signatures of the artifact image data (rows of L), and each time signature can be represented by a linear combination of a very small number of pure time endmembers, which is known as the linear mixing model (Golbabaee and Vandergheynst, 2012; Iordache et al., 2011). Suppose the number of pure spectral endmembers for the artifact image data L is upper bounded by r, then the rank of L is also bounded by r, i.e., $rank(L) \leq r$. Usually, this upper bound value of the number of endmembers r is significantly smaller than the column number and row number of L, which suggests the low-rank property of the Casorati matrix L.

Based on the low-rank property of matrix L and sparsity of matrix S, the low-rank matrix recovery (LRMR) model can be used to estimate the image L and S from the original phase-contrast microscopy image M.

2.1.4. LRMR model and RPCA problem

The low-rank matrix recovery (LRMR) model is first proposed in Wright et al. (2009) and is considered as an idealized Robust Principal Component Analysis (RPCA) problem. For our problem, the RPCA can be formulated as follows: given the original phase-contrast microscopy image data matrix M, the low-rank artifact and non-mitotic image data matrix L and sparse mitosis candidate matrix S are unknown, and we are trying to estimate L and S. This optimization problem can be formulated as

$$\min_{L,S} rank(L) + \lambda \|S\|_0 \quad s.t. \quad M = L + S \tag{3}$$

where $rank(\cdot)$ denotes the rank of a matrix, and λ is a positive weighting parameter. However, this is a nonconvex optimization problem, and there is no efficient solution available (Wright et al., 2009). A feasible solution is relaxing this problem by replacing the rank with the nuclear norm and the ℓ_0-norm with the ℓ_1-norm to obtain a tractable optimization problem (Candès and Recht, 2009; Peng et al., 2010).

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t. \quad M = L + S \tag{4}$$

The augmented Lagrangian multiplier (ALM) function of problem Eq. (4) is

$$\mathbf{L}(L, S, Y, \mu) = \|L\|_* + \lambda \|S\|_1 + \langle Y, M - L - S \rangle + \frac{\mu}{2} \|M - L - S\|_F^2 \tag{5}$$

where $\langle \cdot, \cdot \rangle$ is the element-wise multiplication, and $\|\cdot\|_F$ is the Frobenius norm, which is defined as $\|X\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2}$ for $X \in R^{m \times n}$. This ALM function can be solved by applying the Alternating Splitting Augmented Lagrangian Method (ASALM) (Tao and Yuan, 2011) and the Iterative Thresholding (IT) approach (Wright et al., 2009). More specifically, the ALM function is decomposed into two smaller subproblems which solve the variables L and S separably in the consecutive order and in an iterative way. Given $(S^{(k)}, L^{(k)}, Y^{(k)})$, the ASALM updates the optimal solution via the following scheme until convergence:

$$\begin{cases} S^{(k+1)} = \arg\min_{S} \mathbf{L}(L^{(k)}, S, Y^{(k)}, \mu) \\ L^{(k+1)} = \arg\min_{L} \mathbf{L}(L, S^{(k+1)}, Y^{(k)}, \mu) \\ Y^{(k+1)} = Y^{(k)} + \lambda (M - L^{(k+1)} - S^{(k+1)}) \end{cases} \tag{6}$$

The first subproblem in (Eq. (6)) can be written into the following form:

$$\begin{aligned} S^{(k+1)} &= \arg\min_{S} \mathbf{L}\left(L^{(k)}, S, Y^{(k)}, \mu\right) \\ &= \arg\min_{S} \Big(\lambda \|S\|_1 + \langle Y^{(k)}, M - L^{(k)} - S \rangle \\ &\quad + \tfrac{\mu}{2} \|M - L^{(k)} - S\|_F^2 \Big) \\ &= \arg\min_{S} \Big(\lambda \|S\|_1 + \langle Y^{(k)}, M - L^{(k)} - S \rangle \\ &\quad + \tfrac{\mu}{2} \|M - L^{(k)} - S\|_F^2 + \tfrac{(Y^{(k)})^2}{2\mu} \Big) \\ &= \arg\min_{S} \Big(\lambda \|S\|_1 + \tfrac{\mu}{2} \|S - (M - L^{(k)} + \tfrac{Y^{(k)}}{\mu})\|_F^2 \Big) \end{aligned} \tag{7}$$

According to *Lemma 2.1* in (Tao and Yuan, 2011), the optimal solution of this function can be given as follows:

$$S^{(k+1)} = \Phi_{\frac{\lambda}{\mu}} \left(M - L^{(k)} + \frac{Y^{(k)}}{\mu} \right) \tag{8}$$

where $\Phi_{\frac{\lambda}{\mu}}$ is defined componentwisely by

$$\left(\Phi_{\frac{\lambda}{\mu}}(X) \right)_{ij} = \max \left\{ abs\left(X_{ij} - \frac{\lambda}{\mu}, 0 \right) \right\} \cdot sign(X_{ij}) \tag{9}$$

in which $abs(\cdot)$ and $sign(\cdot)$ are the absolute value function and sign function, respectively.

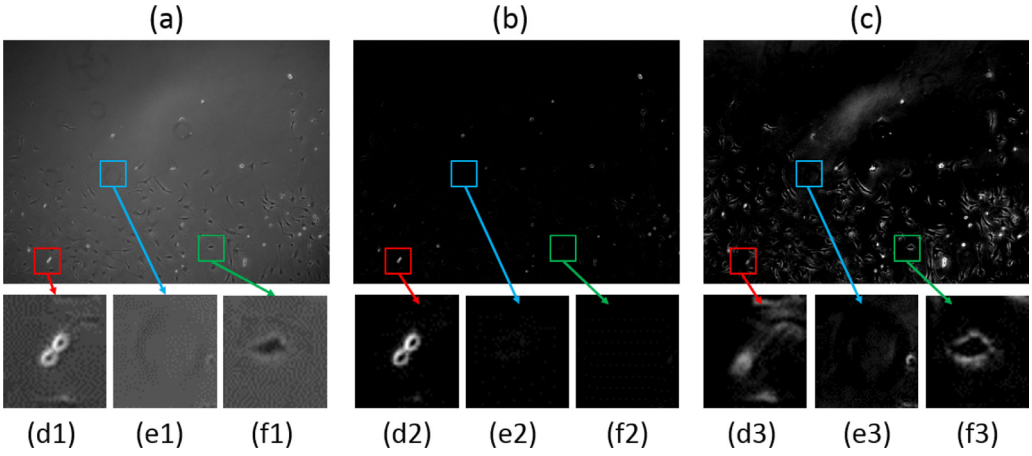

Fig. 7. The result of LRMR model. (a)–(c) are the original image, image containing mitosis candidates and image containing stationary artifacts and non-mitotic cells, respectively. We extract our candidate patch sequences from image (b). As we can see from image (d1)–(d3), the mitosis cell is kept while background in the patch is removed. (e1) only contains background stain and in (e2) the stain is removed. (f1) contains a non-mitotic cell and it is removed in (f2).

The second subproblem in (Eq. (6)) can be written into a more specific form:

$$\begin{aligned} L^{(k+1)} &= \arg\min_{L} \mathbf{L}\left(L, S^{(k+1)}, Y^{(k)}, \mu\right) \\ &= \arg\min_{L} \left(\|L\|_* + \left\langle Y^{(k)}, M - L - S^{(k+1)} \right\rangle \right. \\ &\quad \left. + \frac{\mu}{2} \|M - L - S^{(k+1)}\|_F^2 \right) \\ &= \arg\min_{L} \left(\|L\|_* + \left\langle Y^{(k)}, M - L - S^{(k+1)} \right\rangle \right. \\ &\quad \left. + \frac{\mu}{2} \|M - L - S^{(k+1)}\|_F^2 + \frac{(Y^{(k)})^2}{2\mu} \right) \\ &= \arg\min_{L} \left(\|L\|_* + \frac{\mu}{2} \|L - (M - S^{(k+1)} + \frac{Y^{(k)}}{\mu})\|_F^2 \right) \end{aligned} \tag{10}$$

According to *Lemma 2.2* in (Tao and Yuan, 2011), the optimal solution of this function can be expressed as follows:

$$L^{(k+1)} = \Psi_{\frac{1}{\mu}} \left(M - S^{(k+1)} + \frac{Y^{(k)}}{\mu} \right) \tag{11}$$

The $\Psi_{\frac{1}{\mu}}$ in this equation is defined by

$$\Psi_{\frac{1}{\mu}}(X) = U diag\left(\Phi_{\frac{1}{\mu}}(\Sigma) \right) V^T \tag{12}$$

where U, V and Σ are obtained by the singular value decomposition (SVD) of X:

$$X = U\Sigma V^T, and \quad \Sigma = diag(\sigma_1, \sigma_2, \cdots) \tag{13}$$

We summarize our method in the following form:

The kth iteration of ASALM for problem Eq. (6).
Given $(L^{(k)}, S^{(k)}, Y^{(k)})$, we update them as follows:
1. $S^{(k+1)} = \Phi_{\frac{\lambda}{\mu}}(M - L^{(k)} + \frac{Y^{(k)}}{\mu})$
2. $L^{(k+1)} = \Psi_{\frac{1}{\mu}}(M - S^{(k+1)} + \frac{Y^{(k)}}{\mu})$
3. $Y^{(k+1)} = Y^{(k)} + \lambda(M - L^{(k+1)} - S^{(k+1)})$

The low-rank matrix recovery (LRMR) model is able to remove the stationary artifacts from the phase-contrast microscopy images. Moreover, as most of the cells move slowly in consequent images, they will be regarded as the low-rank component and fall into the matrix L. So a byproduct of the LRMR model is that most of the stationary cells are removed, and only the mitotic and migrating cells are picked out as the mitosis candidates, which are separated into matrix S. This can reduce the number of negative samples greatly, and as a result, the searching space and computational time can be decreased heavily.

We illustrate the result of LRMR model in Fig. 7. By the proposed LRMR model, most background and non-mitotic cells can be removed and the intensity of mitosis cells can be kept, as shown in Fig. 7(b).

2.2. Mitosis image patch retrieval

After the image pre-processing, we are able to obtain images which possibly contain mitosis cells, migrating cells and moving artifacts. We apply a small gaussian filter to smooth the image and threshold it into a binary mask. We only consider those blobs (connected components in the binary mask) whose areas are above a threshold as potential mitotic regions. Finally we track each blob into candidate sequences by formulating the tracking as an association problem (Mao et al., 2014), and each image patch is extracted at the fixed size $d \times d$ around the center of each connected component. colorredWhen one blob divides into two blobs, the tracking continues until the two blobs cannot be captured by one image patch.

The time length of mitosis events may be quite different. However, the most salient images during the mitosis are just a few images around the birth moment, so we choose a fixed short temporal window to extract candidate patch sequences as the input to our HCNN. As for the input of TS-BLSTM, the entire patch sequence is taken as the input.

In our experiments, the Gaussian filter has standard derivation of 3, the threshold in thresholding images is set to be 5% of the maximum intensity value, the patch size is defined as 52×52, and the minimum blob area is required to be 400 pixels in the datasets. We set all the parameters here safely to ensure that the recall of mitosis events is 100% before the classification step. The search space in the video sequence is largely reduced but the precision of mitosis events by the candidate extraction step is low (1.2%), thus we propose the HCNN in the next section to further improve the performance.

3. Hierarchical convolutional neural networks architecture

3.1. Design of HCNN

As shown in Fig. 8, our proposed Hierarchical Convolutional Neural Networks (HCNN) have three layers. Five consecutive patches in the candidate patch sequence are the input to the five convolutional neural networks in the first layer (CNN_1^k, $k \in [1, 5]$),

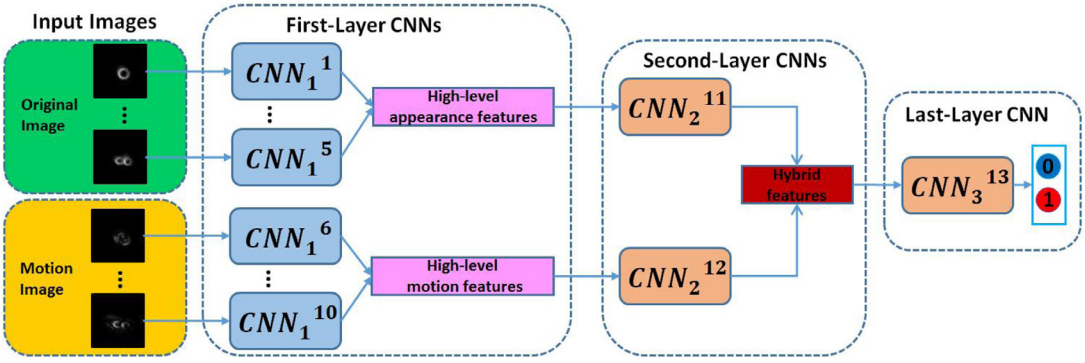

Fig. 8. The architecture of our proposed Hierarchical Convolutional Neural Networks.

and the five corresponding motion images computed by the central finite difference are the input to the five convolutional neural networks (CNN_1^k, $k \in [6, 10]$). Each of the ten convolutional neural networks in the first layer (CNN_1^k, $k \in [1, 10]$) takes a single image as the input. In the second layer of our HCNN, two CNNs (CNN_2^{11} and CNN_2^{12}) are designed to extract joint features from the original patch sequences and their motion patch sequences separately at the patch-sequence level. In the last layer of our HCNN, appearance and motion features are combined to feed into the last CNN (CNN_3^{13}). The binary output indicates if the input candidate patch sequence contains a mitosis or not. In the notation of CNN_i^k, i denotes the layer in our HCNN and k indexes the CNN out of the total 13 CNNs in our HCNN.

Two reasons motivates us to design such an architecture. First, the process of mitosis actually is a continuous event which happens in several frames. Instead of detecting the mitosis events by only targeting at the detection of a single frame (birth moment), leveraging several nearby frames will provide more information to detect the birth moments of mitosis events. Second, the motion pattern of mitotic cells are different from that of migration cells, thus the motion information should be utilized to provide richer feature description.

The first layer of our HCNN contains ten CNNs ($CNN_1^k, k \in [1, 10]$). Each of them classifies a single appearance or motion image at different time instants of a mitosis event. The ten CNNs shares the same architecture as shown in Fig. 9. Note that the ten CNNs share the same architecture design, but they do not share parameters and they are trained separately. In this way, each CNN is trained individually to extract the most accurate features for cell images at different mitotic stages. The underlying motivation is that task-specific feature extraction will perform better than a general feature extractor. There are four convolutional layers. The first two convolutional layers are followed by a 2×2 max pooling layer. One drop-out layer is added in case of over-fitting. The prediction layer outputs the label of the input image, which indicates if the input image is the image at the specific time instant of a mitosis event.

The CNNs in the second and last layer of our HCNN (CNN_2^{11}, CNN_2^{12} and CNN_3^{13}) share the similar architecture as shown in Fig. 10. The only difference is the length of the input in the input feature layer. The input to CNN_2^{11} is the combined features from the Fully-connection Layer 2 of $CNN_1^k, k \in [1, 5]$. The input to CNN_2^{12} is the combined features from the Fully-connection Layer 2 of $CNN_1^k, k \in [6, 10]$. Thus the input for the CNNs in the second layer is a 5120 vector. The input to CNN_3^{13} is the combined features from the Fully-connection Layer 3 of CNN_2^{11} and CNN_2^{12}, leading to a 4096 vector.

3.2. HCNN training

The overall HCNN has 13 CNNs, which leads to a large number of parameters. The training complexity will increase if we train the whole HCNN at once. Given the limited amount of training data, training the whole HCNN together will also increase the risk of over-fitting. Therefore, we propose a two-step training process as below.

3.2.1. Pretraining each CNN independently

First, each CNN in three layers is trained independently. The input to the first-layer CNNs ($CNN_1^k, k \in [1, 10]$) is the five original images and their corresponding five motion images. For each input modality, we use the trained weights of the first CNN (e.g., CNN_1^1) as the initialization for the rest four CNNs (e.g., $CNN_1^k, k \in [2, 5]$) to achieve faster convergence. After the training on CNN_1^k ($k \in [1, 10]$) is completed, we retrieve the features from Fully-connection Layer 2 of each CNN and concatenate them as the input to the seconde-layer CNNs (CNN_2^{11} and CNN_2^{12}). The input to the third-layer CNN (CNN_3^{13}) is the concatenated features from Fully-connection Layer

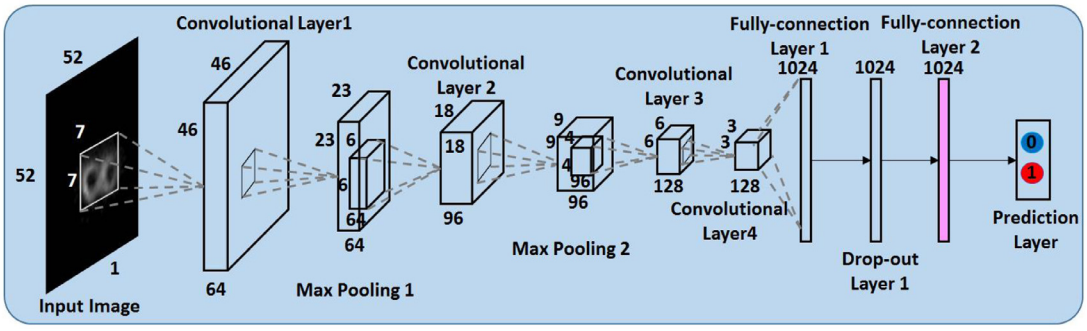

Fig. 9. The architecture of CNNs in the first layers of our HCNN.

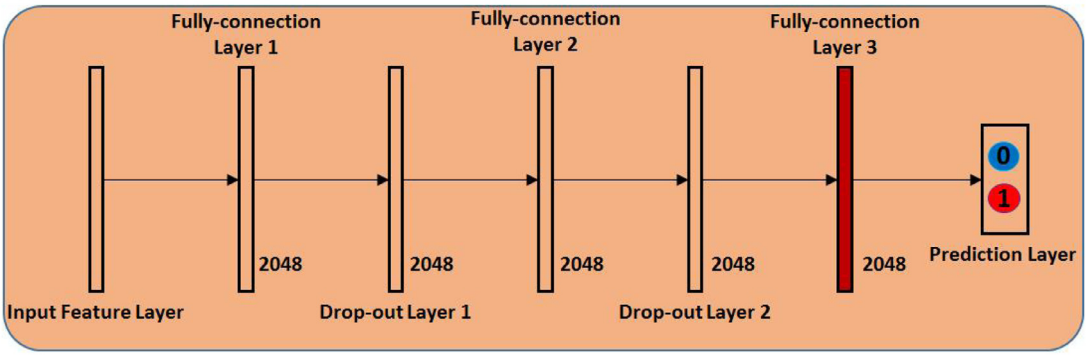

Fig. 10. The architecture of the second and last layer of our HCNN.

3 in the second-layer CNNs. When training the 13 CNNs, we set the batch size as 100 and the number of epochs as 20 with the learning rate gradually decreasing from 10^{-2} to 10^{-4}. We set the drop-out rate to be 0.5 for all drop-out layers.

3.2.2. Fine-tuning HCNN

After each CNN is properly pretrained, we fine-tune the entire HCNN. The Prediction Layers of CNNs in the first and second layers of our HCNN are bypassed and the error from the third-layer CNN (CNN_3^{13}) is back-propagated to all the CNNs to updates their weights.

4. Two-stream bidirectional long-Short term memory

We classify each candidate patch sequence by our HCNN. If the candidate patch sequence contains a mitosis event, we further localize the four stages within it by a new TS-BLSTM, as shown in Fig. 11.

Our proposed TS-BLSTM is shown in Fig. 11. Similar to our HCNN, the first set of input is the N appearance patches X_i, $i \in [1, N]$ in the mitosis patch sequence, and the second set of input is their corresponding motion images M_i, $i \in [1, N]$. The CNNs used to extract the feature representation share the same architecture as the CNN (Fig. 9) in the first layer of HCNN. The feature is extracted from the last fully-connected layer. Therefore, feature vector f_i^x and f_i^m are extracted from appearance image X_i and motion image M_i, respectively. Then the features of appearance images and motion images are fed into Bidirectional Long-Short Term Memory(BLSTM) and generate the label l_i^x and l_i^m, respectively.

For each image, its label l_i^x predicted by the appearance BLSTM and label l_i^m predicted by the motion BLSTM are concatenated to make the final prediction L_i which indicates which stage each image belongs to in the sequence, i.e. solving the mitosis stage localization problem.

It is critical to locate the transition frame between two consecutive stages in order to precisely label each stage in the input sequence. When the human experts annotated the ground truth of different stages, it is hard to determine which frame is exactly the transition frame between two stages without looking back and forth. This suggests that the stage labeling should consider two directions. In our architecture, we use the Bidirectional LSTM which has the ability to unify information in both directions to label one image in the sequence.

Similar to the design of HCNN, we utilize not only the appearance information, but also the motion information over time in the design of our TS-BLSTM due to the fact that the movement pattern of mitotic cells during different stages are different from that of migration cells. Therefore, rich features to describe the data can

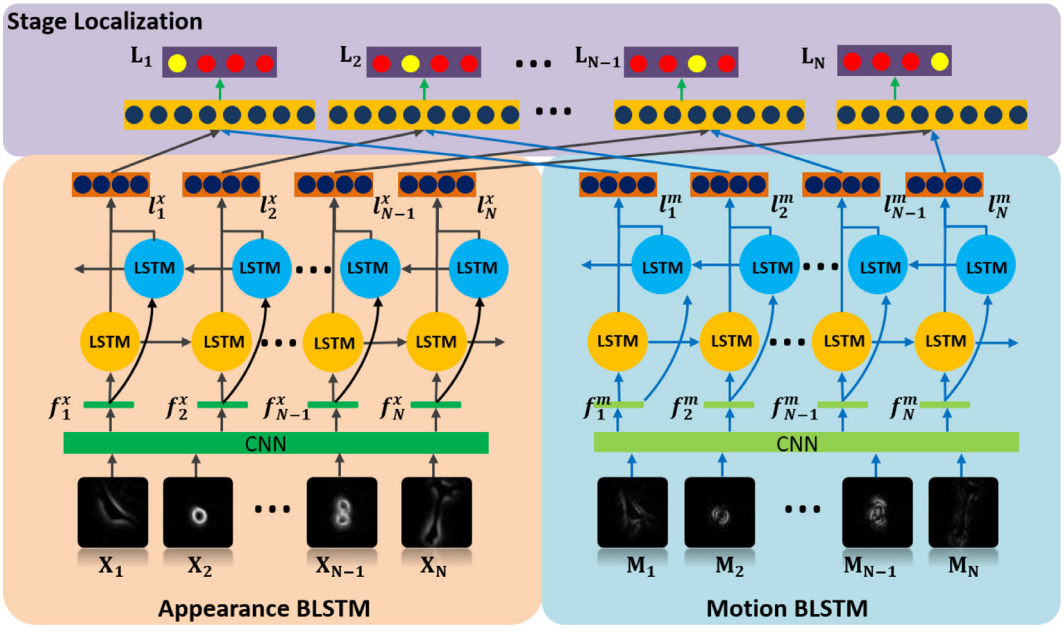

Fig. 11. The overview of our proposed TS-BLSTM.

be provided by unifying both appearance and motion cues, which should boost the stage classification performance.

When training the CNN, only the starting frame of Stage 3 is considered positive, and others are labeled as negative. Two individual CNNs are trained for the appearance input and motion input, respectively. For the training of CNN, we set the patch size as 100 and the number of epoch as 20 with the learning rate gradually decreasing from 10^{-2} to 10^{-4}. The drop-out rate is set to be 0.5. When training the TS-BLSTM with Keras (Chollet et al., 2015), we pad each training sequence to be the length of 50. The number of epoch is set as 10, and the learning rate is 10^{-3} with the weight decay rate as 10^{-6}.

5. Experiments

In this section, we evaluate the performance of our proposed methods on mitosis event localization and stage localization.

5.1. Dataset

Five phase contrast image sequences of C3H10T1/2 mouse mesenchymal stem cell populations (American Type Culture Collection, Manassas, VA) were acquired, each containing 1436 images. Each image sequence contains 79, 94, 85, 120 and 41 mitosis cells, respectively. The multipotent C3H10T1/2 stem cells serve as a model for the adult human mesenchymal stem cells that can differentiate into a variety of cell types. The growing environment consists of Dulbeccos Modified Eagles Media (DMEM; Invitrogen, Carlsbad, CA), 10% fetal bovine serum (Invitrogen, Carlsbad, CA) and 1% penicillin streptomycin (PS; Invitrogen, Carlsbad, CA). The cells were observed during growth under a Zeiss Axiovert 135TV inverted microscope with the 0.15 N.A. objective and phase contrast optics. Time-lapse image acquisition was performed every 5 min using a 12-bit Qimaging Retiga EXi Fast CCD camera at 500 ms exposure with a gain of 1.01. Each image consists of 1392×1040 pixels with a resolution of 19 um/pixel.

After the image acquisition, manual annotation of mitosis events was performed on the image sequences by an expert biologist using a labeling tool with a graphical user interface. For each mitosis, the annotator first locates the center of the boundary between the two daughter cells when the boundary is clearly observed, which marks the start of Stage 3 (formation of daughter cells) as defined in Fig. 1. Then the annotator locates each stage of the mitosis cell in nearby consecutive frames. The location and time of each stage in the video sequences are provided as the ground truth.

In order to train our HCNN and TS-BLSTM, data expansion is performed to generate more positive training data. For each positive mitosis sequence, we rotate the images every 45° (8 variations), slightly translate the images (e.g., by 5 pixels) horizontally and/or vertically (9 variations), which generates 72 times of the original positive training data. We retrieve negative training sequences by our proposed candidate patch sequence extraction method. At last, the training data are balanced by randomly duplicating some positive data so that the numbers of positive samples and negative samples are even. In total, 69,334 training samples are used in the training. During the testing, we apply the HCNN model on sequences with variant lengths in a sliding-window fashion with the stride equal to 1. For example, the HCNN is applied on the first 5 images (1st to 5th image) in the sequence, then it is applied on the 2nd to 6th image, so on so forth.

5.2. Evaluation metrics

To evaluate the accuracy of the proposed method, we adopt the leave-one-out policy in the experiment, i.e., using four image sequences for training and the rest one for testing. We define the True Positive (TP) as a HCNN-detected patch sequence that contains a mitosis event, False Positive (FP) as it does not contain a mitosis event, and False Negative (FN) as the mitosis event that is mistakenly classified as a negative.

To evaluate the accuracy of mitosis detection, we use two evaluation metrics in our experiments. First, we evaluate the performance of mitosis event localization in terms of the mean and standard deviation of precision, recall and *F*-score on the five leave-one-out tests, without examining the timing of birth events.

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F-\text{score} &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \tag{14}$$

Second, the performance of mitosis detection is strictly evaluated in terms of the timing error of birth moments. The **timing error** is measured as the frame difference between the detection result and the ground truth.

Table 1
Validation of the effectiveness of LRMR.

Model	Precision (%)	Recall (%)	*F*-score (%)
HCNN with LRMR	99.2 ± 0.7	98.0 ± 2.1	98.7 ± 1.7
HCNN without LRMR	99.1 ± 0.8	97.2 ± 2.4	98.6 ± 1.3

5.3. Validation of the LRMR and HCNN

First, we evaluate the effect of LRMR on the performance of mitosis detection. Our designed HCNN model without LRMR achieves the precision of 99.1% and recall of 97.2%, further improving the performance is very challenging and the errors are from hard samples. As shown in the Table 1, when we train the HCNN on the microscopy images preprocessed by the proposed LRMR, the precision and recall are 0.1% and 0.8% better than the HCNN trained on images preprocessed by the previous method (Mao and Yin, 2016). The LRMR model allows the network to correctly classify 28.6% mitosis events out of the False Negatives that are incorrectly classified by HCNN without LRMR. In (Mao and Yin, 2016), each image is subtracted by the average image, which results in decreasing the intensity value of mitotic cells and its contrast with the background. With the LRMR model, the intensity value of mitotic cells is kept, thus the performance is better.

The small improvement counts as contribution for two reasons. First, accurate mitosis detection is important for cell tracking that uses data association (e.g., linear programming) to track multiple cells. Over time, it is important to detect all the time instants when the cells divide, so we can obtain accurate lineage trees for cells. To have the rigid biomedical discovery, the small improvement of mitosis detection performance that aims for perfect accuracy will result in big impact on the cell lineage tracking. Second, the challenge here is that there is a little room to improve the mitosis event detection and the rest few percentage points are related to very hard cases. When considering to achieve 99.99% accuracy on the mitosis detection, every step closing to the perfect will be very challenging but meaningful.

Then, we show the effectiveness of each module in the proposed architecture design. Several architectures are compared: a simple CNN which takes the 10-channel (a tensor with 10 images (5 consecutive appearance and 5 corresponding motion images)) as input and our proposed HCNN, a multi-appearance HCNN with only 5 original image patches as input (CNN_1^1 to $CNN_1^5 + CNN_2^{11}$),

Table 2
Mitosis event localization accuracy of different designs.

Model	Precision (%)	Recall (%)	*F*-score (%)
Our HCNN	99.2 ± 0.7	98.0 ± 2.1	98.7 ± 1.7
CNN with multi-channel input	97.6 ± 1.2	94.0 ± 1.9	95.8 ± 1.2
Multi-Appearance HCNN	90.9 ± 3.8	85.6 ± 3.3	88.1 ± 1.4
Multi-Motion HCNN	84.7 ± 2.9	85.6 ± 1.7	85.7 ± 2.0
Single Appearance CNN	85.9 ± 4.7	80.5 ± 8.1	82.9 ± 4.7

a multi-motion HCNN with only 5 appearance image patches as input (CNN_1^6 to CNN_1^{10} + CNN_2^{12}) and a single-appearance CNN (CNN_1^3) which aim at detecting the birth moment, as shown in Table 2. As demonstrated in (Karpathy et al., 2014), fusing the temporal information in the feature level is better than in the input pixel level, thus our HCNN performs better than a simple CNN with 10-channel images as the input. The poorer performance of the multi-appearance CNN is due to the lack of motion information as the input. It is not guaranteed that the multi-appearance CNN model can precisely extract the motion information from features of consecutive frames. It would be better if we just feed the motion information to the network as input. The *F*-Score of multi-appearance HCNN is 10% points lower than that of our HCNN which further incorporates the motion information. The motion features are import in this task. There are some certain cases where shape and appearance features cannot classify the mitosis events. For example, when two cells are close to each other, it just looks like the stage 3 of a mitosis event when one cell dividing into two daughter cells. It is hard to tell whether it is one cell divides into two cells or two cells moving together only by appearance features alone. But the motion image at the beginning of stage 3 of a mitotic event is totally different from two cells getting close to each other. The multi-motion HCNN achieves similar recall with multi-appearance HCNN but has 6.2% lower precision. The reason is that the motion of some cells during mitosis event is not clear enough to be distinguishable with the small motion noise in the background when training the model. Thus during testing, the background motion noise may be classified as mitosis event, resulting in lower precision. Since the single-appearance CNN is only based on one single frame, it cannot capture the temporal appearance change. The *F*-Score of single-appearance CNN is 5% points lower than that of the multi-appearance HCNN which utilizes the temporal information in the whole patch sequence.

Furthermore, we evaluate the effect of time-window length in the design of HCNN architecture. In the Table 3, the HCNN-3, HCNN-5, HCNN-7 and HCNN-9 refer to HCNN with 3,5,7 and 9 CNNs in the first layer. As shown in the table, when increasing the length of time window, the performance of HCNN improves. The HCNN-9 works the best with a *F*-score at 98.8%. The performance of HCNN-7 and HCNN-9 is similar since the key frames are usually 3–5 frames near the birth moment and the remaining frames do not provide more useful information to boost the performance. Moreover, with the increasing number of CNNs in the first layer, the computation cost and the risk of over-fitting will increase. Thus we choose a time-window length of 5 as the proper parameter in the design of our HCNN.

Table 3
Validation of the effectiveness of time-window length.

Model	Precision (%)	Recall (%)	*F*-score (%)
HCNN-3	95.6 ± 2.7	96.3 ± 3.4	95.5 ± 2.4
HCNN-5	99.2 ± 0.7	98.0 ± 2.1	98.7 ± 1.7
HCNN-7	99.5 ± 1.1	98.1 ± 1.5	98.7 ± 1.3
HCNN-9	99.6 ± 0.6	98.3 ± 1.7	98.8 ± 1.4

5.4. Comparisons on the mitosis event localization

We compare our method with eight previous works: BLSTM (Mao and Yin, 2017), Max-Margin Hidden Conditional Random Fields+Max-Margin Semi-Markov Model (MM-HCRF + MM-SMM) (Liu et al., 2012), 3D CNN (Nie et al., 2016), EDCRF (Huh et al., 2011), HCRF (Liu et al., 2010), Hidden Markov Models (HMMs) (Rabiner, 1989), and Support Vector Machine (SVM) (Suykens et al., 2000). As shown in Table 4, our HCNN achieves an average precision of 99.2%, recall of 98.0% and *F*-score of 98.7%, which outperforms the state-of-the-arts by a large margin. The BLSTM (Mao and Yin, 2017) ranks as the second. The model proposed in BLSTM takes advantage of the power of LSTM in utilizing the temporal context. However, the mitosis event detection and stage localization are different tasks after all. When the network tries to update the parameters to reach the best performance for stage localization, the noisy loss from mitosis event detection may be introduced into the back-propagation to decrease the performance of stage localization, and vice-versa. The conflict of two tasks eventually prevent the BLSTM from reaching the best performance for both tasks. MM-HCRF+MM-SMM adopts a two-step mechanism: MM-HCRF is used to classify a candidate sequence if it contains mitosis or not, then MM-SMM further identifies the different stages in the candidate sequence 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 performs better than using MM-HCRF alone (by 4.6% of averaged *F*-score). Compared with MM-HCRF, EDCRF has achieved a higher performance (by 1.7% 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. The performance of 3D CNN (Nie et al., 2016) is lower than the MM-HCRF. Even though the 3D kernel is able to extract features across consecutive frames, the kernel itself is more capable of extracting features in real 3D input, such as corners of a table, not the 3D input constructed by 2D images. Another possible reason is that 3D CNN is hard to train. CRF and HMM methods both have much lower performance (only higher than the benchmark 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 event localization. The reason is that the SVM method is not capable of modeling temporal dynamics among frames in the mitosis sequences, thus the mitosis event localization could only be done frame-by-frame without benefiting from the temporal context.

5.5. Comparisons on the mitosis stage localization

To label one mitosis sequence into the four stages, we only need to localize the starting frame of stage 2, 3, and 4. In the previous work, only MM-HCRF+MM-SMM (Liu et al., 2012) and BLSTM (Mao and Yin, 2017) are able to localize different stages while others only focus on the localization of the starting frame of stage 3. We summarize the comparison of each mitosis stage localization method in Table 5. The results in Table 5 demonstrate that our method performs different stage localization with better performance than (Liu et al., 2012; Mao and Yin, 2017) in term of the frame error. We test our TS-BLSTM with only one bidirectional LSTM, i.e. fusing the appearance and motion features as the input to only one LSTM to predict the mitosis stages. The TS-BLSTM with one LSTM has larger error than our TS-BLSTM, which further

Table 4
Comparison of mitosis event localization accuracy.

Model	Precision (%)	Recall (%)	*F*-score (%)
Our HCNN	99.2 ± 0.7	98.0 ± 2.1	98.7 ± 1.7
BLSTM (Mao and Yin, 2017)	98.4 ± 1.0	97.0 ± 1.8	97.7 ± 1.2
MM-HCRF+MM-SMM (Liu et al., 2012)	95.8 ± 1.0	88.1 ± 3.1	91.8 ± 2.0
3D CNN (Nie et al., 2016)	84.9 ± 1.2	83.6 ± 0.8	84.2 ± 0.9
MM-HCRF (Liu et al., 2012)	82.8 ± 2.4	92.2 ± 2.4	87.2 ± 1.6
EDCRF (Huh et al., 2011)	91.3 ± 4.0	87.0 ± 4.8	88.9 ± 0.7
CRF (Lafferty et al., 2001)	90.5 ± 4.7	75.3 ± 9.6	81.5 ± 4.4
HMM (Rabiner, 1989)	83.4 ± 4.9	79.4 ± 8.8	81.0 ± 3.4
SVM (Lowe, 2004)	68.0 ± 3.4	96.0 ± 4.2	79.5 ± 1.7

Table 5
Comparison of each averaged temporal stage localization error.

Model	Stage 2	Stage 3	Stage 4
Our TS-BLSTM	0.72 ± 0.23	0.59 ± 0.52	0.04 ± 0.05
TS-BLSTM with one LSTM	0.76 ± 0.14	0.61 ± 0.17	0.08 ± 0.07
BLSTM (Mao and Yin, 2017)	0.78 ± 0.40	0.62 ± 0.62	0.06 ± 0.06
MM-HCRF+MM-SMM (Liu et al., 2012)	0.82 ± 1.69	0.73 ± 1.29	1.06 ± 1.72
HCNN (Mao and Yin, 2016)	N/A	0.69 ± 0.91	N/A
EDCRF (Huh et al., 2011)	N/A	0.83 ± 1.34	N/A

proves that fusing the features in later stage is better than fusing in early stage. The BLSTM solves the mitosis stage localization problem together with the mitosis event localization in one single model, but we have an individual model for each task, thus we achieve the better performance. The MM-HCRF+MM-SMM adopts hand-crafted features and it achieves worse performance than the LSTM based models.

6. Conclusion and future directions

In this paper, we explore the power of deep learning in the two tasks of mitosis analysis: mitosis event localization and its stage localization. We propose a new system which consists of three models: first, a Low-Rank Matrix Recovery (LRMR) model is formulated to find salient regions that potentially contain mitosis events and extract candidate patch sequences; second, a Hierarchical Convolution Neural Network (HCNN) which utilizes visual appearance and motion cues to classify each candidate patch sequence is used to solve the task of mitosis event localization; third, we design a Two-stream Bidirectional Long-Short Term Memory (TS-BLSTM) which is able to label every image in one mitosis sequence is designed to precisely localize each mitotic stage. In the experiments, we compare our proposed deep learning models with previous traditional approaches on the mitosis analysis problem. The proposed system outperforms state-of-the-arts by achieving 99.2% precision and 98.0% recall for mitosis event localization and 0.62 frame error on average for mitosis stage localization in five challenging image sequences.

Compared with the static medical image data, time-lapse microscopy image sequences provide important temporal clues to boost the performance of tasks in medical image analysis. Deep learning based models have shown great advantage over traditional approaches. However, annotating biomedical images requires the expert knowledge and it is time-consuming. It is necessary to develop approaches such as unsupervised or semi-supervised methods for the training of deep learning in the future.

Declaration of Competing Interest

None.

Acknowledgments

This project was supported by National Science Foundation (NSF) CAREER award IIS-1351049 and NSF EPSCoR grant IIA-1355406. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Al-Kofahi, O., Radke, R.J., Goderie, S.K., Shen, Q., Temple, S., Roysam, B., 2006. Automated cell lineage construction: a rapid method to analyze clonal development established with murine neural progenitor cells. Cell Cycle 5 3, 327–335.

Candès, E.J., Recht, B., 2009. Exact matrix completion via convex optimization. Found. Comput. Math. 9 (6), 717–772 arXiv:0805.4471. doi:10.1007/s10208-009-9045-5.

Chollet, F., et al., 2015. Keras. https://keras.io.

Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2013. Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted intervention: MICCAI...International Conference on Medical Image Computing and Computer-Assisted Intervention, 16 Pt 2, pp. 411–418.

Debeir, O., Ham, P.V., Kiss, R., Decaestecker, C., 2005. Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. IEEE Trans. Med. Imaging 24, 697–711.

Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., Darrell, T., 2017. Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 677–691 arXiv:1411.4389. doi:10.1109/TPAMI.2016.2599174.

Gallardo, G. M., Yang, F., Ianzini, F., Mackey, M., Sonka, M., 2004. Mitotic cell recognition with hidden Markov models. doi:10.1117/12.535778.

Golbabaee, M., Vandergheynst, P., 2012. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 2741–2744 arXiv:1211.5058v1. doi:10.1109/ICASSP.2012.6288484.

Hochreiter, S., Urgen Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780 arXiv:1206.2944. doi:10.1162/neco.1997.9.8.1735.

Huang, C.-H., Lee, H.-K., 2012. Automated mitosis detection based on exclusive independent component analysis. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 1856–1859.

Huh, S., Ker, D.F.E., Bise, R., Chen, M., Kanade, T., 2011. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. IEEE Trans. Med. Imaging 30, 586–596.

Iordache, M.-D., Bioucas-Dias, J.M., Plaza, A.J., 2011. Sparse unmixing of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 49, 2014–2039.

Ji, S., Yang, M., Yu, K., Xu, W., 2013. 3D convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35 (1), 221–231 arXiv:1102.0183. doi:10.1109/TPAMI.2012.59.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Li, F.F., 2014. Large-scale video classification with convolutional neural networks. Proc. IEEE CVPR 1725–1732 arXiv:1412.0767. doi:10.1109/CVPR.2014.223.

Lafferty, J., McCallum, A., Pereira, F.C.N., 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning, vol. 8, pp. 282–289 arXiv:1011.4088v1. doi:10.1038/nprot.2006.61.

Li, K., Chen, M., Kanade, T., 2007. Cell population tracking and lineage construction with spatiotemporal context. Med. Image Anal. 12 5, 546–566.

Li, K., Miller, E.D., Chen, M., Kanade, T., Weiss, L.E., Campbell, P.G., 2008. Computer vision tracking of stemness. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 847–850.

Liang, L., Zhou, X., Li, F., Wong, S.T., Huckins, J., King, R.W., 2008. Mitosis cell identification with conditional random fields. In: 2007 IEEE/NIH Life Science Systems and Applications Workshop, LISA, pp. 9–12. doi:10.1109/LSSA.2007.4400872.

Liu, A., Li, K., Kanade, T., 2012. A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. IEEE Trans. Med. Imaging 31, 359–369.

Liu, A.-A., Li, K., Kanade, T., 2010. Mitosis sequence detection using hidden conditional random fields. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 580–583.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110 arXiv:0112017. doi:10.1023/B:VISI.0000029664.99615.94.

Mao, Y., Li, H., Yin, Z., 2014. Who missed the class? - Unifying multi-face detection, tracking and recognition in videos. In: Proceedings - IEEE International Conference on Multimedia and Expo, 2014-September doi:10.1109/ICME.2014.6890334.

Mao, Y., Yin, Z., 2016. A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9901 LNCS, pp. 685–692. doi:10.1007/978-3-319-46723-8_79.

Mao, Y., Yin, Z., 2017. Two-stream bidirectional long short-term memory for mitosis event detection and stage localization in phase-contrast microscopy images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 10434 LNCS, pp. 56–64. doi:10.1007/978-3-319-66185-8_7.

Nie, W.Z., Li, W.H., Liu, A.A., Hao, T., Su, Y.T., 2016. 3D convolutional networks-based mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1359–1366. doi:10.1109/CVPRW.2016.171.

Padfield, D.R., Rittscher, J., Thomas, N., Roysam, B., 2009. Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. Med. Image Anal. 13 1, 143–155.

Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y., 2010. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 763–770.

Quattoni, A., Wang, S.B., Morency, L.-P., Collins, M., Darrell, T., 2007. Hidden conditional random fields. In: IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 10, pp. 1848–1852.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE, Vol. 77, no. 2, pp. 257–286. doi:10.1109/5.18626.

Su, H., Maji, S., Kalogerakis, E., Learned-miller, E., 2015. Multi-view convolutional neural networks for 3D shape recognition. IEEE ICCV 945–953 arXiv:1505.00880. doi:10.1109/ICCV.2015.114.

Sun, Y., Wang, X., Tang, X., 2016. Hybrid deep learning for face verification. IEEE Trans. Pattern Anal. Mach. Intell. 38 (10), 1997–2009. doi:10.1109/TPAMI.2015.2505293.

Suykens, J.A.K., Lukas, L., Vandewalle, J., 2000. Sparse least squares support vector machine classifiers. ESANN.

Tao, M., Yuan, X., 2011. Recovering low-Rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. 21 (1), 57–81. doi:10.1137/100781894.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2015. Learning spatiotemporal features with 3D convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision. Vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 4489–4497 arXiv:1412.0767. doi:10.1109/ICCV.2015.510.

Viola, P., Jones, M.M.J., 2004. Robust real-time face detection. Int. J. Comput. Vis. 57 (2), 137–154 arXiv:1011.1669v3. doi:10.1023/B:VISI.0000013087.49260.fb.

Wright, J., Peng, Y., Ma, Y., Ganesh, A., Rao, S., 2009. Robust principal component analysis: exact recovery of corrupted low-rank matrices. Adv. Neural Inf. Process. Syst. (NIPS) 2080—-2088 arXiv:0905.0233. doi:10.1002/cpa.20132.

Wu, J., Brubaker, S.C., Mullin, M.D., Rehg, J.M., 2008. Fast asymmetric learning for cascade face detection. IEEE Trans. Pattern Anal. Mach. Intell. 30, 369–382.

Yan, Z., Zhang, H., Piramuthu, R., Jagadeesh, V., DeCoste, D., Di, W., Yu, Y., 2015. HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2740–2748 arXiv:1410.0736. doi:10.1109/ICCV.2015.314.