Cell mitosis detection using deep neural networks

Yao Zhou, Hua Mao*, Zhang Yi

Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, People's Republic of China

Keywords:
Cell mitosis detection
Deep neural networks
Convolutional neural networks

A B S T R A C T

Quantitative analysis of cell mitosis, the process by which cells regenerate, is important in cell biology. Automatic cell mitosis detection can greatly facilitate the investigation of cell life cycle. However, cell-type diversity, cell non-rigid deformation and high cell density pose difficulties on handcrafting visual features for traditional approaches. Aided by massively captured microscopy image sequences, deep neural networks have recently become available for automatic cell mitosis detection. This paper proposes an end-to-end framework named as F3D-CNN for mitosis detection, and F3D-CNN is directly trained from data without requiring designing domain dependent features. Well-trained F3D-CNN first filters out potential mitosis events based on the static information in each individual image, and further discriminates candidates by incorporating the spatiotemporal information from image sequences. The state-of-the-art performance of F3D-CNN was confirmed in experiments on two public datasets (multipotent C3H10T1/2 mesenchymal stem cells and C2C12 myoblastic stem cells).

1. Introduction

Cell mitosis [1] is a complex process by which mature cells produce next-generation cells. During this process, the ancestor cell's membrane divides to form two new cells, and its genetic material is duplicated and evenly distributed. To measure cell proliferation and analyze the cells' responses to various stimuli, cell biologists usually perform tedious and time-consuming procedures in wet laboratories. In particular, they monitor cells over time to collect informative data, then study the cell dynamics. However, modern microscopy image capture systems can automatically and regularly take images of the monitored cells [2]. Using computer vision based approaches, cell mitosis can be studied from a large volume of collected high-quality biomedical data without intervening with cell processes [3]. Apparently, there is a keen requirement for automatic and robust approaches that can detect the time and location of cell mitosis events from given image sequences [1]. As cells undergo non-rigid deformations, and are generally diverse and densely packed, developing efficient cell mitosis detection approaches remains a challenging problem.

Deep neural networks (DNNs) have achieved state-of-the-art performance in various tasks [4,5], as they can automatically learn representative features from high-dimensional data [6]. With representation learning [7], the performance of data-driven mitosis detection from histology images has been improved [8,9]. Convolutional neural networks (CNNs) [10], which constitute one class of DNNs, differ from traditional multilayer perceptrons (MLPs) by employing local connectivity and shared weights to reduce the number of free parameters, thereby preventing over-fitting problems. In microscopy images, modeling spatiotemporal features are important for mitosis detection rather than only focus on static features [1,11,12]. In 3D convolutional neural networks (3D-CNNs), the extended 3D convolutional kernels can process temporal data, e.g., human actions can be recognized from image sequences [13]. In typical CNN-based applications [14], high-dimensional input images or image sequences are mapped into (usually) simple result labels such as classification tasks. Fully convolutional networks (FCNs) include up-sampling layers that perform image-to-image prediction [15]. The network output of an FCN can be sized identically to the input images. CNN and its variants offer several advantages in cell mitosis detection. First, they automatically learn robust features from raw data, avoiding the need for domain dependent feature designing. Second, 3D-CNNs can efficiently capture both spatial and temporal features simultaneously. Finally, CNNs can be easily parallelized on computing platforms with graphical processing units (GPUs) for efficient computing.

In order to automatically detect cell mitosis events from microscopy image, by combining FCNs and 3D-CNNs, this paper proposes a deep neural network named as F3D-CNN. F3D-CNN comprises two stages: candidate detection and mitosis discrimination. In the candidate detection stage, after learning static features of cell mitosis events in a supervised manner, FCNs retrieve areas,

* Corresponding author.
E-mail addresses: zy3381@gmail.com (Y. Zhou), huamao@scu.edu.cn (H. Mao), zhangyi@scu.edu.cn (Z. Yi).

where contain potential cell mitosis events, from individual microscopy images. As cell mitosis processes usually span several consecutive images, a positive mitosis event can only be concluded after considering both spatial and temporal information from adjacent image frames. In the mitosis discrimination stage, previous detected candidates are further discriminated by 3D-CNNs. The proposed F3D-CNN relaxes the requirement of manual feature designing and selection, as it can automatically learn robust and representative features, including the static, spatial, and temporal ones, directly from captured data. As F3D-CNN is an end-to-end solution, it is applicable given any type of cell and image capturing equipment without tedious feature designing and parameter tuning. After training, the time efficiency of F3D-CNN meets the requirement for real-time microscopy image processing, because feed-forward computation is always efficient. The performance of F3D-CNN, including the precision of position and time of finally detect mitosis events, has been empirically verified on a publicly available dataset of microscopy image sequences [16], and a comparison study with other methods [1,17,18] has also been conducted. Experimental results indicate that F3D-CNN outperforms state-of-the-art approaches.

The rest of this paper is organized as follows. Section 2 reviews and discusses related works, and Section 3 briefly introduces the basic models of the proposed framework. The details of F3D-CNN, including the candidate detection and mitosis discrimination stages and the practical considerations, are presented in Section 4. Section 5 conducts a thorough empirical study and a comparison study on publicly available datasets, and analyzes the results. The paper concludes with Section 6.

2. Related works

Existing cell mitosis detection methods can be grouped into two categories: cell tracking detection and candidate discrimination methods. In the former category, cell segmentation [2] is typically followed by trajectory tracking [19]. Here, the trajectories are constructed by associating the cells in consecutive frames [20]. One cell at the present frame could be associated with any individual at the next frame, and the likelihood of such associations between cells are quantified by cost functions. By optimizing the cost function, the relationship among cells in consecutive image frames can be constructed. One mitosis event could be detected if there one cell at the present frame is associated to two new cells at the next frame. Apparently, the performance of mitosis detection highly depends on the tracking algorithm. Most importantly, tracking of all cells, regardless of whether they will undergo future mitosis or not, is inefficient. Moreover, the segmentation performance is degraded by cell overlaps and the indistinctness between cell membrane and background, and long term tracking is prone to drift [21]. Therefore, the precondition of cell mitosis detection is unreliable.

Candidate discrimination methods first extract the candidate sequences that might contain cell mitosis events, and select their representative features [22]. Mitosis events in the candidate sequences are then identified through supervised learning [23]. As candidate sequences which may contain mitosis events are first filtered out, the search space has been reduced, therefore this type of approaches are more effective than tracking based ones. However, constructing candidate sequences from the images requires carefully designed image processing algorithms, which are highly depend on various conditions, e.g., cell type and population, illumination conditions, and the image acquisition equipments.

Additionally, supervised learning is usually performed based on In particular, brightness characteristic has been extensively used for constructing candidate sequences, and descriptive features (e.g. gradient histogram, local binary pattern) are used to characterize specific cell mitosis events. Based on that, typical machine learning approaches, like conditional random field, sparse Gaussian process, can determine the location and time of mitosis events [1,11,12,24]. As feature designing and selection are largely depend on prior domain knowledge, to the best of our knowledge, there has not been universal solution for mitosis detection in general.

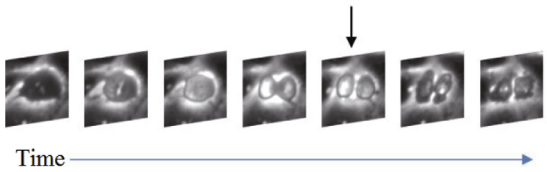

Fig. 1. Example of a cell mitosis event in consecutive microscopy images. Cell mitosis (indicated by the arrow) appears in the fifth image of the sequence.

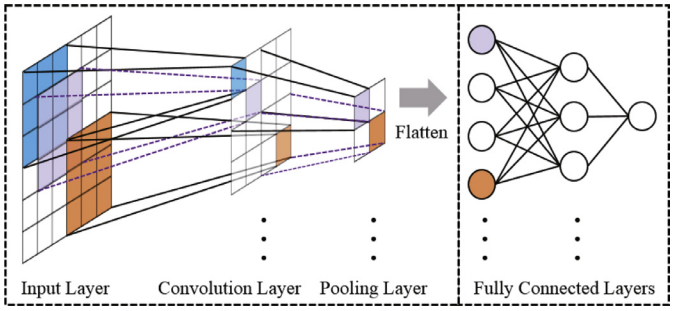

Fig. 2. Architecture of a CNN with one convolutional layer using a (3×3) kernel, one pooling layer using a (2×2) kernel, and two fully connected layers. Flatten transforms the feature matrices to vectors.

3. Preliminaries

In this section, we formally describe the cell mitosis detection problem, then briefly introduce the deep neural networks used in this paper, including CNNs [10], FCNs [15], and 3D-CNNs [13]. More details could be found in the suggested references.

3.1. Cell mitosis detection

A distinctive feature of cell mitosis [1] is the division of the ancestor cell's membrane to form two new cells. A cell mitosis event is defined from the moment that a clear boundary appears between two daughter cells. Given a sequence of microscopy images capturing the cell activities over time, cell mitosis detection aims to determine the timing (i.e., the exact frame in the sequence) and location (in that frame) of a mitosis event. An example of cell mitosis is shown in Fig. 1. The event manifests as a sequence of patches cropped at a fixed position in consecutive original microscopy images. The timing and location of the mitosis event are indicated by the arrow in the fifth frame and the position of the patch in the original images, respectively.

3.2. Convolutional neural networks (CNNs)

CNNs [10] are composed of convolutional layers, pooling layers and fully connected layers (Fig. 2). The weights of the convolutional layers, called kernels, locally connect to the input and are updated by a back-propagation algorithm during training. CNNs with stacked convolutional layers can learn features with hierarchical structure. The pooling layer relieves the computational burden by shrinking the feature maps, and confers transformation invariance. After the convolutional and pooling layers, the neuron activities are mapped to an output vector by the fully connected layers.

Let the indices of an L-layer CNN be l ($1 \le l \le L$), and denote the weight and bias by w and b, respectively. P and Q represent the

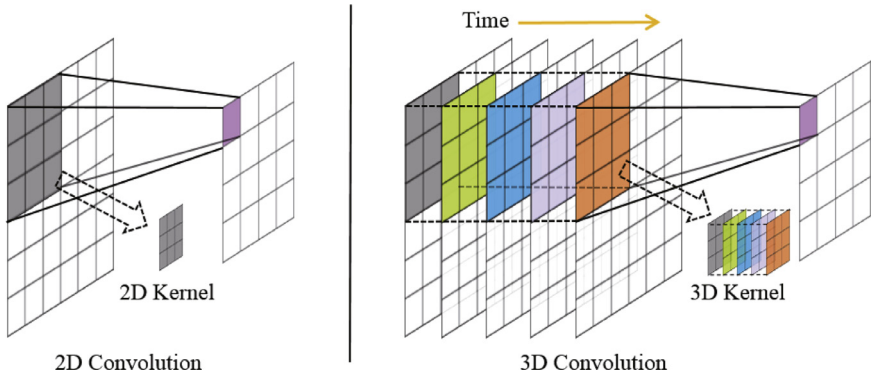

Fig. 3. Comparison between 2D and 3D convolutions with kernel sizes of (3 × 3) and (5 × 3 × 3), respectively. The stride is 1 in both convolutions. In this example, the 2D convolution convolves on 1 frame using a 2D kernel, whereas the 3D convolution convolves on 5 frames using a 3D kernel.

size of the kernels. A non-linear activation function $f(\cdot)$ is applied on each layer; common choices are sigmoid and rectifier functions [25]. The activation of the lth layer at position (c, r) is denoted by $a^l_{c,r}$, and is calculated as follows:

$$a^l_{c,r} = f\left(b^l + \sum_{i=1}^{P}\sum_{j=1}^{Q} w^l_{i,j} \cdot a^{l-1}_{c+i,r+j}\right). \quad (1)$$

3.3. Fully convolutional networks (FCNs)

In MLPs, the dimensions of weight matrices between fully connected layers are directly determined given the number of neurons in each layer. Differently, in CNNs, the sizes of convolutional kernels between layers have no such strong constrain, and could be even independent from the size of adjacent layers. As a variant of CNNs, FCNs focus on learning convolutional kernels for general purpose without imposing constrains regarding the sizes of intermediate representation layers, therefore arbitrarily sized inputs could be processed. In FCNs, each layer is obtained by convolutional operations towards the previous layer, and its size can be determined by the previous layer and convolutional kernels during running time. For example, it is known as semantic segmentation [15] when the size of the output layer is set as the same as the input layer.

To ensure the same size between the output and input, the spatial resolution loss of the input is compensated by up-sampling layers. The up-sampling is commonly achieved through deconvolution [26] and un-pooling [27] layers, which perform the reverse operations of convolution and pooling respectively. Training FCNs from scratch is prone to over-fitting. Practically, an informative CNNs learned from a different but related domain is usually used as the initialization, where its fully connected layers have been converted to 1×1 convolutional layers through net surgery. After appending up-sampling layers, FCNs are later fine tuned based on the dataset.

The original FCN design also included skip connections from the down-sampling to the up-sampling for recovering the fine-grained information.

3.4. 3D Convolutional neural networks (3D-CNNs)

3D-CNNs [13] extend the original CNNs by introducing 3D convolution operations, which incorporate the temporal information over consecutive image frames into the network. Therefore, 3D-CNNs can learn the features in both spatial and temporal dimensions. Whereas a 2D convolution kernel operates on one area of a single image frame, a 3D convolution kernel operates on the same area stacked over multiple consecutive frames, as shown in Fig. 3.

Similarly to Eq. (1), the 3D convolution is computed as follows:

$$a^l_{x,y,z} = f\left(b^l + \sum_{i=1}^{P}\sum_{j=1}^{Q}\sum_{k=1}^{R} w^l_{i,j,k} \cdot a^{l-1}_{x+i,y+j,z+k}\right) \quad (2)$$

where z denotes the third dimension of the input data, and w is the three dimensional weight matrix of the lth layer at position (i, j, k), with size R along the temporal dimension, $a^{l-1}_{x,y,z}$ represents the activation value of the (l-1)th layer.

4. F3D-CNN

The proposed framework, F3D-CNN, can be regarded as a two-stage candidate discrimination based mitosis detection approach. As shown in Fig. 4 (right), the microscopy images are sequentially fed into a FCN, and the output score maps indicate the location of candidate events on each image. Those candidates are further discriminated by incorporating temporal information from adjacent image frames to finally determine positive mitosis events. F3D-CNN is directly learned from data, and the available event annotations contribute to the training procedure of both stages. As manually feature designing and selection are no longer needed, F3D-CNN can be regarded as a universal end-to-end solution for mitosis detection. The details and practical considerations about the architecture of F3D-CNN will be presented in this section.

4.1. Candidate detection

Candidate detection identifies positions of potential of cell mitosis events in each frame. Previous studies filtered out candidates depending on the variation of brightness [1,11,24]. Brightness characteristic is a feature which typically exhibits in the process of mitosis events. However, the magnitude of pixel intensity variation appears to be different among cell mitosis events, and can exhibit vast diversity between different type of cells. Hence, it requires a carefully designed threshold to utilize brightness to detect candidates, and may prone to artifacts. Different from previous methods, we build an FCN to alleviate these problems in candidate detection with following customizations:

- Point annotation conversion. To fully exploit context of cell mitosis events, a Gaussian-like smoothing method is applied to point annotations.
- Training with crops. As mitosis events are sparsely distributed, for training efficiency, original images are cropped into smaller ones.
- Multi-loss objective function. To mitigate multi-scale issue of cell size and gradient vanishing problem, there are multiple loss functions considered.

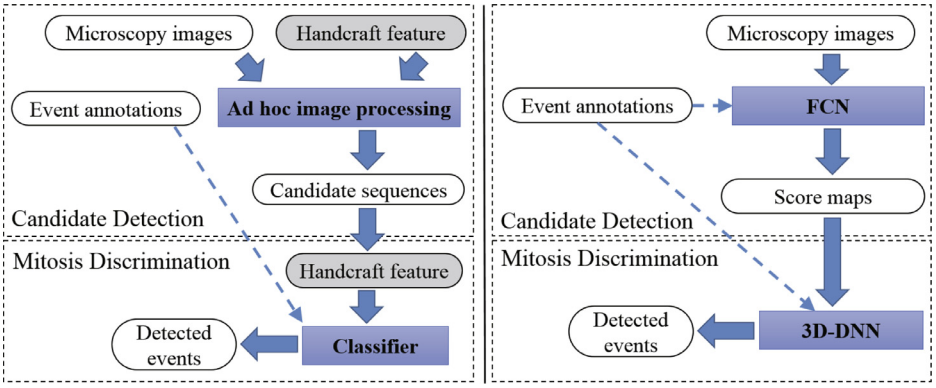

Fig. 4. Left: a conventional two stages cell mitosis detection framework [1]. Right: the proposed F3D-CNN framework. Event annotations are only used during the training stage, as indicated by dash arrows.

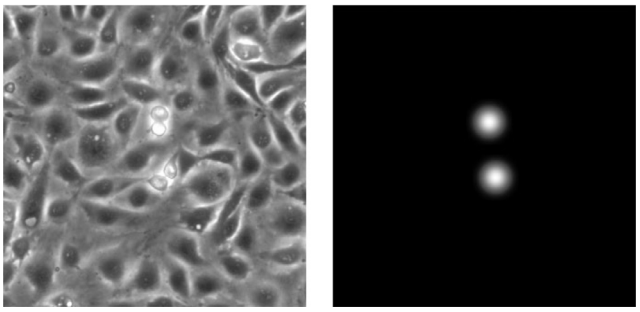

Fig. 5. Example of a cropped microscopy image (left) and its corresponding score map (right).

Point annotation conversion. Public cell image datasets commonly mark only the point annotation, defined as the centroid point of a cell mitosis event [1]. However, this annotation does not fully exploit the context around cell mitosis events. To deal with this problem, the score maps are generated by applying a Gaussian-like smoothing strategy to the point annotations of training set. Formally, the positions of mitosis events in a microscopy image I_t are represented in a score map S_t of the same size as I_t, where t is the time step and the value at each point in S_t represents the probability of a cell mitosis event at the corresponding image position. The FCN maps image I_t to a same-sized score map S_t in an image-to-image prediction manner. A microscopy image and its corresponding score map are displayed in Fig. 5. Specifically, cell mitosis events on a given frame at time step t can be denoted as $\mathbb{G}_t = \{p_{t1}, p_{t2}, \ldots, p_{ti}\}$, where p_{ti} is the location of the ith centroid of the cell mitosis events marked at frame t, or $\mathbb{G}_t = \{\varnothing\}$ for a frame containing no mitosis events. Hence, the point annotations are denoted as $\mathbb{G} = \{G_1, G_2, \ldots, G_T\}$, where T is the number of frames, and the value S_t^p of the score map at time step t and position p is generated by the following process:

$$S_t^p = \begin{cases} e^{-\frac{\|\tilde{p}-p\|}{2\sigma_s^2}} & \tilde{p} \in \mathbb{G}_t \text{ and } \|\tilde{p}-p\| < r_s, \\ 0 & otherwise \end{cases} \tag{3}$$

where σ_s denotes the variance of the mitosis event at the centroid positions and r_s controls the truncated spatial context scope of each cell mitosis event.

Training with crops. Because cell mitosis events are sparsely distributed on each frame, most of the pixels on a frame are labeled as non-mitosis, which imbalances the FCN learning dataset. To cope with this imbalance, we crop the small regions centered on the events.

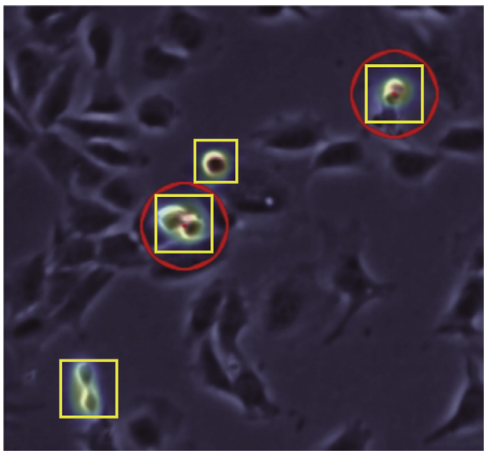

Fig. 6. Input image overlapped with its output score map. The centers of squares and circles represent candidates and annotated cell mitosis events, respectively.

Convolutional networks learn local features with local connections. For a given image, those learnt kernels are applied upon local areas subject to their receptive field regardless the image's size. With an appropriate receptive size, and cropped patches cover mitotic cells, non-mitotic cells, artifacts and backgrounds, the learnt convolutional kernels can effectively discover potential mitosis events. Given a higher resolution input image, kernels are applied in the feed forward computation. Kernels learned from low-resolution images patches can be applied on original high-resolution images, but the size of the final representation layer will be larger, which is the distinctive feature of FCNs from traditional CNNs. In addition, similar strategy was also adopted for training FCNs in the literature in [28]. For a well-trained FCN, candidates can be obtained by suppressing the non-maximal scores on its produced score maps for a given input image. Fig. 6 shows an example of detected cell mitosis candidates and their corresponding point annotations.

Multi-loss objective function. Although cells in microscopy images are generally considered within an unique scale [1], cell mitosis events generally exhibit deformation which results in varying cell sizes. Meanwhile, deep FCN models are vulnerable to gradient vanishing which may leads to training difficulty. Therefore, the FCN architecture design focuses mainly on the receptive field size and the gradient vanishing problem. The receptive field is the area of the lower layers connected by units in the higher layers [15]. This field, which determines the scale of the detectable features [29]. As the receptive field size increases with increasing network depth,

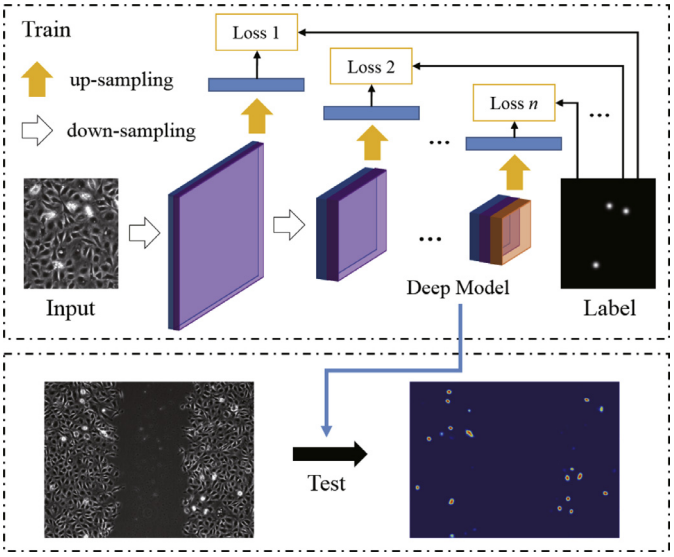

Fig. 7. FCN for mitosis candidate detection. Each hidden layer is followed by an auxiliary loss layer. The training and testing are performed on cropped and whole images, respectively.

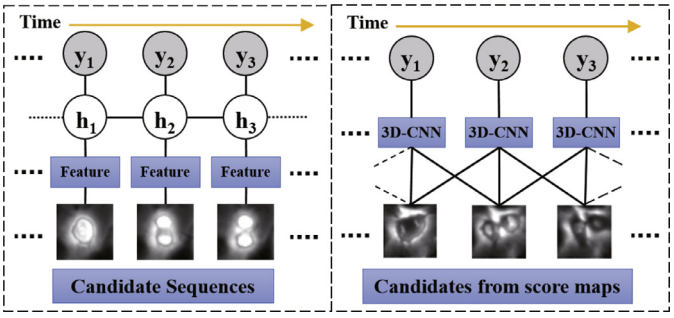

Fig. 8. Left: CRF based model for mitosis event temporal localization. Spatial features designing is firstly employed, then it models the temporal dynamic of mitosis events through hidden state (h_i) transition, and output a probability y_i determines there is a mitosis event at time step i. Right: 3D-CNN for mitosis discrimination. It directly learns spatiotemporal feature from context of candidates obtained from score map, and produce a probability y_i which indicates there is a mitosis event.

hierarchical features at different levels focus on different scaled objects. Low layer captures small size cell mitosis events while large size events are captured by high layers. In FCN, skip connections [15] are typically used for combining features from both high layers and low layers, and it uses a single loss function for the final output. As deep layers commonly give better results than the low layers, the skip connections don't explicitly provide guidance for low layers to capture small size objects. While a FCN with auxiliary output loss on intermediate layers can alleviate this shortage [30], due to it regularizes each output to produce score map at a certain scale. With these considerations, a FCN is constructed as the detailed architecture shown in Fig. 7.

Formally, for an FCN with L layers, the feature map a^{l+1} at (l+1)th layer is computed as follows:

$$a^{l+1} = f(w_c^l \otimes a^l)$$

where $0 < l < L$ and $a^0 = I_t$. The parameter of the lth convolutional layer is denoted as w_c^l, $\otimes$ is the convolution or pooling operation that shrinks the feature maps, and $f(\cdot)$ is an activation function.

The output o^l at the lth layer is computed as follows:

$$o^{l+1} = f(w_d^l * a^l)$$

where w_d^l denotes the parameters of the lth deconvolutional layers, and $*$ is the deconvolution operation that enlarges the feature maps to the size of I_t.

Finally, the network is trained by optimizing the objective function as follows:

$$\mathcal{L}_1(\boldsymbol{\theta}) = \sum_{l=1}^{L} \mu^l \cdot \psi(o^l, S_t)$$

where $\mathcal{L}_1$ is the cost function w.r.t. a set of parameters of the FCN, $\psi(o^l, S_t)$ denotes the cross entropy loss about output o^l and the score map S_t, and μ^l is a parameter for weighting each loss function.

By taking advantage of spatial context of cell mitosis events, the FCN learn to filter candidates without relying on handcraft features, and training FCN with cropped images is effective. Moreover, by using multi-loss function, low layers are explicitly guided to produce multi-scale features. Meanwhile, it prevents the magnitudes of gradient in each layer from vanishing, due to the error signal of intermediate layers always comes from both local output and high layer losses, rather than only from the final output loss.

4.2. Mitosis discrimination

Individual image provides FCNs static features to detect mitosis candidates. It dramatically reduces the search space, but may misidentify two close cells as a mitosis event. To alleviate this problem, the candidate detection stage is followed by a 3D-CNN that discriminates the true mitosis events among the detected candidates. The 3D-CNN accepts a candidate with spatiotemporal context as input, and outputs a probability indicate whether it is a cell mitosis event. On the other side, CRF based models are extensively adopted to capture temporal dynamic of cell mitosis events through their hidden state transitions [1,17], and are able to simultaneously perform both sequence classification and temporal localization. Unfortunately, they heavily depend on the candidate sequences construction and representative spatial feature designing, which may not capable of describing sophisticated characteristics of cell mitosis events. Comparison between CRFs and 3D-CNNs is depicted in Fig. 8.

To train the 3D-CNN, the candidates are labeled as training samples. Specifically, a patch centered on a candidate is selected as a frame of a candidate sequence, and patches at the same position in adjacent frames are then selected in the same way to construct the candidate sequence. In particular, patches centered on annotations are extracted as positive samples, and augmented by rotation and mirroring. Formally, each candidate at position p and time step t is denoted as p_t, and its closest annotation is $\bar{p}_t$. A candidate is labeled as a positive sample if $\|\bar{p}_t - p_t\| < r_s$, where r_s represents its spatial context scope as that used in candidate detection stage. Otherwise, the candidate is labeled as a negative sample. Examples of training samples are shown in Fig. 9. For a well-trained 3D-CNN, the candidate patch sequences are constructed identically to those in the training stage, and each of them is assigned with a probability of containing a mitosis event. Among spatially overlapped detections, the one with the highest probability is decided as a mitosis event occurrence.

As most of the candidates classified by the FCNs are negative samples, the class imbalance presents a difficulty to effective learning by the 3D-CNNs (e.g. the number of negative samples is generally 8 times more than that of positive samples.). To deal with this difficulty, we introduce a weighted cost function. Formally, the set of training samples H comprises a set H^+ of positive samples and a set H^- of negative samples. Let X represent an input sample. The

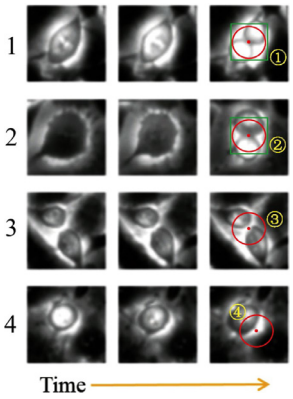

Fig. 9. Each row is a patch sequence. The centroids of the green circles numbered 1 to 4 are candidates detected by FCNs. Green squares are annotations. Sequences 1–2 and 3–4 are positive and negative samples, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

weighted cost function is then given by

$$\mathcal{L}_2(\boldsymbol{\theta}) = -\sum_{X\in H^+} \alpha \log P_X - \sum_{X\in H^-} (1-\alpha)\log(1-P_X) \tag{4}$$

where $\alpha = |H^+|/|H|$, $\boldsymbol{\theta}$ represents the parameters of 3D-CNNs, and P_X is the output for a given sample X, which indicates the probability of that sample is positive.

Cell mitosis discrimination by 3D-CNN should capture the motion information in both spatial and temporal dimensions. Moreover, the 3D-CNNs replace the design of handcrafted volumetric features covering the whole image sequence [31] by automatically learning of the spatiotemporal features from sequence data. As an extreme case, 3D-CNN can be independently used to detect mitosis events in the whole sequence with a sliding-window fashion.

5. Experiments

The proposed F3D-CNN framework was evaluated empirically on publicly available datasets C3H10T1/2 and C2C12 [16]. Following commonly used metrics, like the precision of position and time of finally detect mitosis events, the performance of the proposed method was compared with existing approaches. This section presents the F3D-CNN training and evaluation details, including the datasets descriptions, network architectures and all obtained results.

5.1. Datasets

The CMU cell image analysis group provides two types of microscopy image sequences for cell mitosis detection; C3H10T1/2 and C2C12, representing multipotent C3H10T1/2 mesenchymal stem cells and C2C12 myoblastic stem cells respectively. C3H10T1/2 contains five image sequences (210 frames for each), and C2C12 contains a single 1013-frame image sequence. The resolution of all images is (1392 × 1040) pixels. The time elapse between frames is 5 min. Cell mitosis events (specifically, the center of each critical state of an event) have been manually annotated by cell biologists. The number of the cell mitosis events in each sequence are summarized in Table 1.

Table 1
Mitosis event amount statistics of datasets.

Sequence	Amount
C3H10T1/2-1	465
C3H10T1/2-2	379
C3H10T1/2-3	319
C3H10T1/2-4	324
C3H10T1/2-5	245
C2C12	679

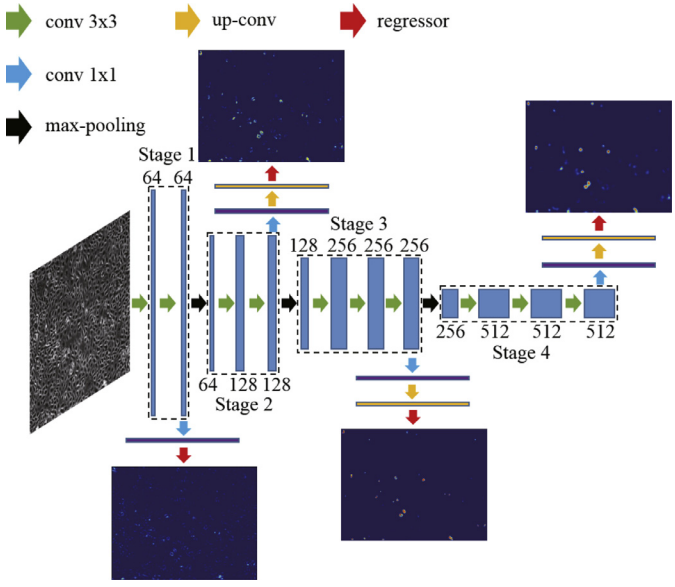

Fig. 10. Architecture of the FCN for candidate detection. Numbers attached to each layer represent the amount of feature maps. Images on left and following the regressors are input and score maps, respectively.

5.2. Training setup

The F3D-CNN is sequentially trained. First, a FCN learns to detect mitosis candidates from the annotations. Then a 3D-CNN is trained to discriminate mitosis events by exploiting candidates' spatiotemporal context. The F3D-CNN can be used for automatically mitosis detection after they are well trained. Training setup of F3D-CNN is described in this section.

5.2.1. FCN

In the candidate detection stage, σ_s was set to 3 to represent the variance of mitosis event centroid positions and r_s was set to 25 to control context scope of mitosis events through cross validation. Approximately 440 cropped smoothed score maps were generated for each sequence, all of which were used for training the FCNs. The threshold for eliminating the background and non-mitosis areas in the non-maximum suppression was set to 0.1.

To deal with the insufficient training data, a CNN with five stage convolutional layers and two fully connected layers [32] was first trained on Imagenet dataset [33]. Although the images are from a different domain, the filters in its lower layers act as corner detectors, which are suitable for cell edge detection. In order to construct a fully convolutional model with suitable receptive field size and low computational cost, all fully connected layers and the last stage of convolutional layers were removed. Besides, each stage was followed by a convolutional layer to produce a single channel feature map, deconvolutional layer was used to up-sample that feature map and *sigmoid* normalization was employed to produce the regression output. Rectifier [25] activation function was applied after each convolutional layer in the four stages. Details of the FCNs architecture was illustrated in Fig. 10. Moreover, co-adaption of intermediate features was reduced by dropout regularization [14], and the invariant transformation and rotation properties of the cell images were enhanced by data augmentation (mirror, rotation and distortion).

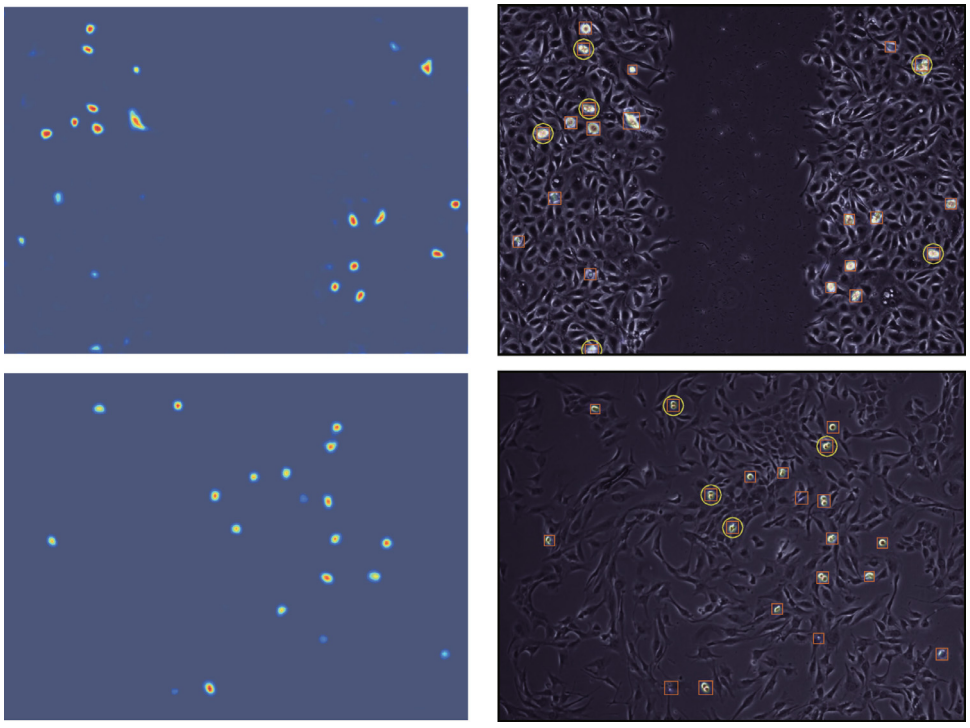

Fig. 11. Examples of candidate detection results on *C3H10T1/2* (top row) and *C2C12* (bottom row) data set, respectively. Generated score maps are shown in the left side, and original images overlapped with score maps are shown in the right side. Squares indicate detected candidate cell mitosis events, and centroid of circles are the given annotations.

Table 2
Architecture of 3D-CNN used mitosis discrimination.

Layers	Type	Filter size	Stride	Filters	Units
1	3D Conv	$3\times3\times3$	$1\times1\times1$	64	–
2	3D Conv	$3\times3\times3$	$1\times1\times1$	64	–
3	3D Conv	$3\times3\times3$	$1\times1\times1$	64	–
4	3D Max-pool	$1\times2\times2$	$1\times2\times2$	–	–
5	Dropout(0.25)	–	–	–	–
6	3D Conv	$3\times3\times3$	$1\times1\times1$	128	–
7	3D Conv	$3\times3\times3$	$1\times1\times1$	128	–
8	3D Conv	$3\times3\times3$	$1\times1\times1$	128	–
9	3D Max-pool	$1\times2\times2$	$1\times2\times2$	–	–
10	Dropout(0.25)	–	–	–	–
11	3D Conv	$3\times3\times3$	$1\times1\times1$	256	–
12	3D Conv	$3\times3\times3$	$1\times1\times1$	256	–
13	3D Conv	$3\times3\times3$	$1\times1\times1$	256	–
14	3D Max-pool	$2\times2\times2$	$2\times2\times2$	–	–
15	Dropout(0.25)	–	–	–	–
16	FullyConnected	–	–	–	1000
17	Dropout(0.2)	–	–	–	–
18	FullyConnected	–	–	–	128
19	Dropout(0.2)	–	–	–	–
20	FullyConnected	–	–	–	2

Table 3
Performance of candidate detection. (*seq* represent sequence).

	Training *seq*	Testing *seq*	Recall	Candidate numbers
C2H10T1/2	1	2,3,4,5	0.986	33,181
	2	1,3,4,5	0.986	28,265
	3	1,2,4,5	0.995	31,921
	4	1,2,3,5	0.996	36,475
	5	1,2,3,4	0.988	25,036
C2C12	part1	part2	0.988	5082
	part2	part1	1	7430

As a result, candidate detection was performed by an FCN. Kernel size was set to (3×3) for convolutional layers in each stage, and a stride of 1 prevents the feature map from shrinking. The convolutional layer for producing single channel output in each stage adopts (1×1) for both its kernel and stride. All max-pooling layers down-sample the feature map with (2×2) for both its kernel and stride. For up-sampling, the kernel size and stride of deconvolutional layers are set to 2^n and 2^{n-1} respectively, where n denote the stage number. As discussed in the previous section, the *Sigmoid* activation function was adopted after each up-sampling layer to produce a normalized regression output, and cross entropy is adopted as loss function for each regressor. With this architecture, scale of cell size are explicitly handled by different stages, and the FCN generates 4 score maps with different receptive field sizes. In particular, the score map generated from the last stage was selected as the final output for a well-trained FCN, which is mainly because the Gaussian-like strategy considers area of cell mitosis events in a unique scale. Loss with other regressors were allocated as regularizers. The weighting terms of μ^l were all set to 1. The hyper-parameters to train the *FCN* include: leaning rate $(1e-06)$, weight decay $(2e-4)$, momentum (0.9), and training epochs (10). Details of the FCNs architecture is illustrated in Fig. 10.

5.2.2. 3D-CNN

In the mitosis discrimination stage, spatiotemporal context is decided by the size of patch sequence, which is set to $(3\times51\times51)$. In order to learn a 3D-CNN for accurate temporal localizing, temporal scope r_t is set to 1, due to each event is annotated within 1 frame. Meanwhile, r_s is set to the same value as that in FCN. Inspired by Tran et al. [34], we modeled the spatiotemporal features of cell mitosis events by a 3D-CNN with small kernels. Specifically, we aggregated the spatial and temporal information using $(3\times3\times3)$ 3D convolutional kernels and $(2\times2\times2)$ max pooling operations. The 3D features are mapped to the last output layer, and softmax normalization was adopted to produce a probability which indicates whether the input patch sequence contains a cell mitosis event. Rectifier activation function was applied after 3D

Table 4
Performance comparisons of mitosis detection.

		Training	Testing	Precision	Recall	F1	AUC
C3H10T1/2	F3D-CNN	seq1	seq2,3,4,5	**0.837**	**0.821**	**0.829**	**0.899**
		seq2	seq1,3,4,5	**0.824**	**0.799**	**0.811**	**0.880**
		seq3	seq1,2,4,5	**0.753**	**0.732**	**0.741**	**0.805**
		seq4	seq1,2,3,5	**0.743**	**0.743**	**0.742**	**0.808**
		seq5	seq1,2,3,4	**0.801**	**0.729**	**0.763**	**0.833**
	EDCRF [1]			0.740	0.703	0.720	0.66
	HCRF+SVM [17]			0.604	0.585	0.594	0.466
	HCRF+CRF [18]			0.583	0.565	0.574	0.463
C2C12	F3D-CNN	part1	part2	**0.889**	**0.829**	**0.858**	**0.944**
		part2	part1	**0.886**	**0.830**	**0.857**	**0.928**
	EDCRF [1]			0.880	0.828	0.853	0.819
	HCRF+SVM [17]			0.550	0.520	0.535	0.270
	HCRF+CRF [18]			0.687	0.650	0.668	0.473

convolutional layers and fully connected layers, and the loss function is described as Eq. (4). The architecture of 3D-CNN is detailed in Table 2.

The training and testing were implemented in Keras on Theano, and all computation was boosted by NVIDIA GeForce GTX TITAN X. Hyper-parameters to train the *3D-CNN* include: leaning rate ($5e-3$), weight decay ($1e-6$), momentum (0.9), and training epochs (25).

5.3. Evaluation

Experimental setting adopted in two most closely related literatures [1,17] is training on one sequence while evaluating on the rest. In particular, each C3H10T1/2 sequence was iteratively evaluated in [17]. In order to conduct a fair comparison with existing mitosis detection approaches, we followed the same experimental setting. For C3H10T1/2, one sequence was used for training, while the remaining ones were reserved for testing. The sequence in the C2C12 dataset was divided into two sub-sequences for training and testing, and each contains approximately half number of mitosis events. To evaluate the performance of FCN, the number of detected candidates was adopted as a metric. Similar to Huh et al. [1], the performance of F3D-CNN was evaluated by four metrics: precision (P), recall (R), f1-measure (F), and area under curve (AUC):

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, F = \frac{2 \cdot P \cdot R}{P+R} \tag{5}$$

where TP, FP and FN denote the total numbers of true-positive, false-positive and false-negative detection results, respectively.

In particular, the undetected mitosis events in candidate detection stage are counted as false negatives. The AUC was obtained by varying the decision probability of each detected mitosis event.

5.4. Results and analysis

In the candidate detection stage, the proposed method can process a (1392 $\times$ 1040)-pixel image in less than 0.3 s. This fast runtime is attributable to the FCNs, which enable efficient localization of mitosis event candidates. Ideally, the method should reduce the search space while maintaining a high sensitivity of true positives. The candidate numbers and recall results of our proposed method are listed in Table 3.

The results confirm that FCNs can effectively detect cell mitosis candidates, and accurately reject the background and non-mitotic cells. In each dataset, the recall approximated 1, meaning that almost none of the cell mitosis events were missed in the candidate detection stage. The candidate numbers in Table 3 denote all candidates detected in the remaining four testing sequences. Note that, there are fewer candidates were found in C2C12 than in C2H10T1/2, as the cell density in C2C12 is lower. Examples of candidates detected by the FCNs are shown in Fig. 11.

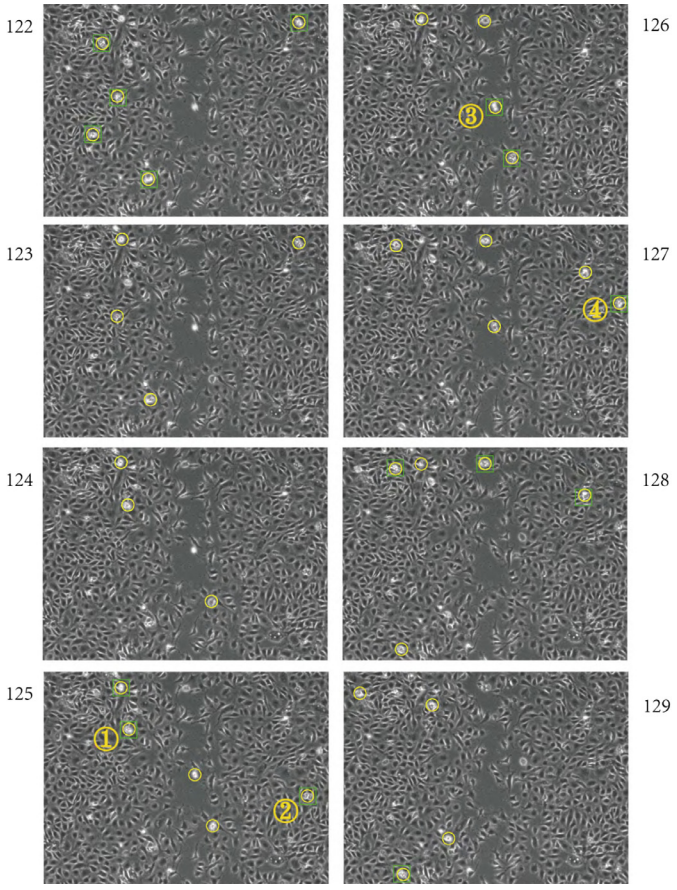

Fig. 12. Examples of cell mitosis detection results on sequence *C3H10T1/2-1*. Squares represent cell mitosis event annotations, and circles represent detections produced by proposed method. There are totally 34 detections, only 4 of them are false positives. More details of those events with numbers enclosed by yellow circles are shown in Fig. 13. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Batch processing in the mitosis discrimination stage identified the candidates in one frame in less than 0.1 s. The total computational time of cell mitosis detection by F3D-CNN was below 0.4 s per sequential frame, corresponding to a cell mitosis detection rate of up to 2.5 frames per second (fps). As this time frame is much less than the acquisition time of a microscopy cell image, our F3D-CNN would be applicable in real-time analysis in advanced microscopic imaging techniques. The training of FCNs and 3D-CNNs required approximately 1 and 8 h respectively on a single GPU.

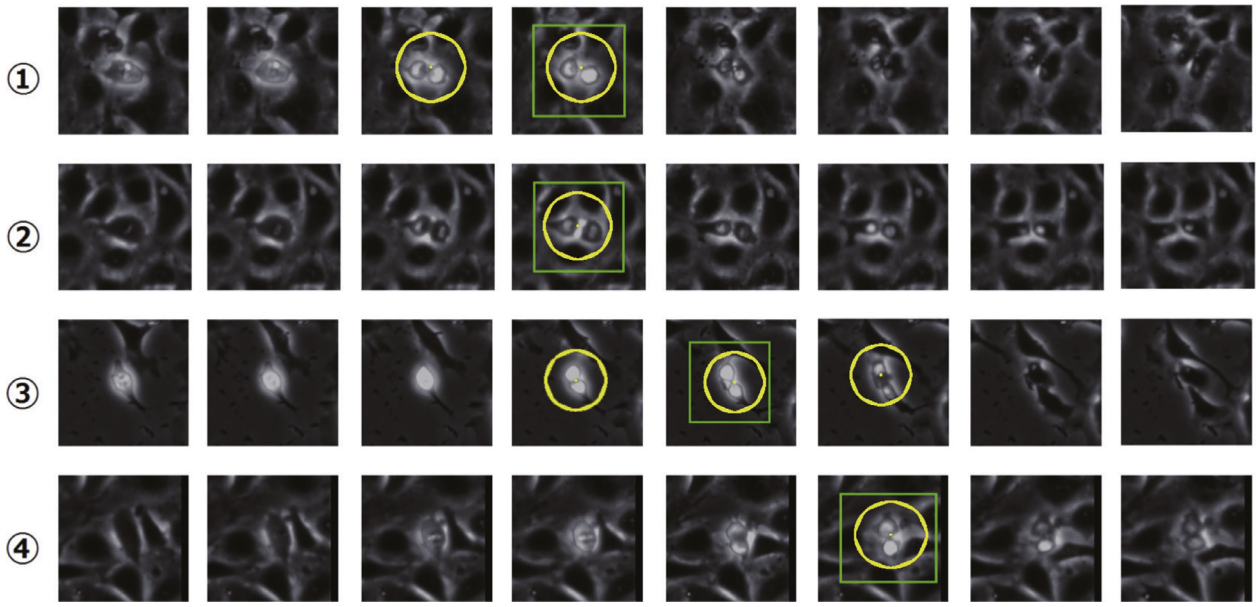

Fig. 13. Details of detections. Frames range from 122 to 129 in sequence *C3H10T1/2-2*. Circles represent the detection results and centers of squares represent the annotations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

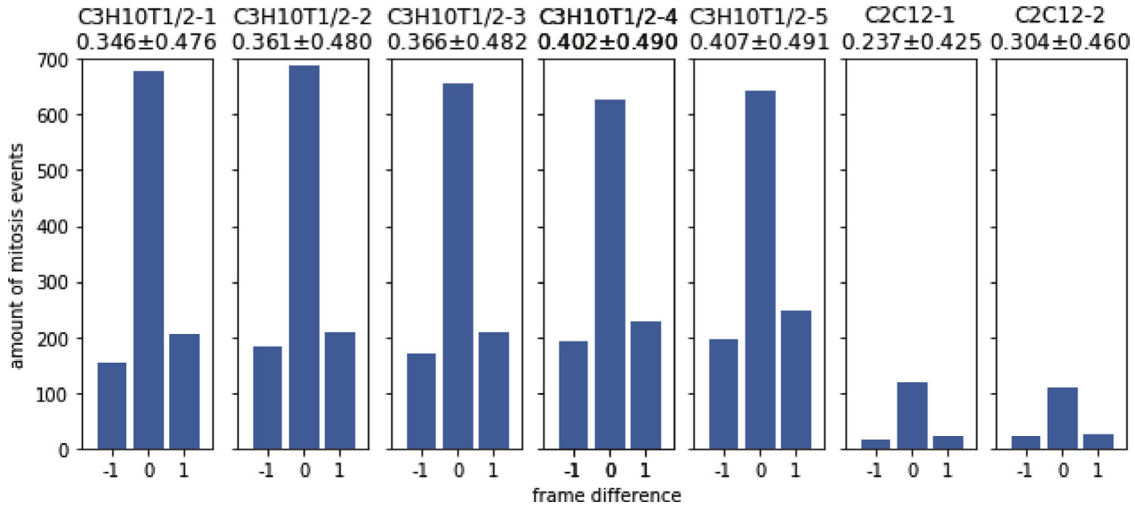

Fig. 14. Temporal localization performance. The dataset name is shown as the first line of each figure's caption. Average and standard deviation of timing error in terms of absolute frame difference are shown as the second line of each figure's caption.

By applying a sliding-window technique, mitosis can also be detected by 3D-CNNs alone. However, the search space of this technique was (1392 × 1040 × 210) patch samples per C3H10T1/2 sequence, at least 4000 times higher than in the F3D-CNN method (see Table 3). The high complexity of the sliding-window method is infeasible in practical applications.

Due to some of the birth event of mitosis cannot precisely observed within only one frame, it is reasonable to evaluate timing of detected events within a small timing error. In particular, the detection result are considered as true positive only when the timing error is not greater than 1, and quantitative results of cell mitosis detection by the proposed method (F3D-CNN) and state-of-the-art methods are compared in Table 4.

F3D-CNN consistently outperformed EDCRF [1], HCRF+SVM [17] and HCRF+CRF [18] on both datasets (C3H10T1/2 and C2C12). The superior performance of F3D-CNN was especially apparent when using the first or second C3H10T1/2 sequence as training data, possibly because more mitosis events are available for training in these sequences. Although the C2C12 dataset contains more mitosis events than any of the C3H10T1/2 sequences, it comprises a single sequence, which must be split into two sub-sequences for the training and testing data. Single-sequence splitting invariably introduces variances because cells deforms over time. Moreover, C2C12 dataset presents a difficulty of identify mitotic cells with high adhesion. With these challenges, F3D-CNN still achieved a comparable performance compare to the other methods, which demonstrated it can deal with different circumstances. In addition, when timing of mitosis event is not considered, the methods of EDCRF [1], HCRF+SVM [17] and HCRF+CRF [18] are detecting sequences which may contain mitosis events, instead of detecting birth event of mitosis, which is different from the experiment setup adopted in Table 4.

Fig. 12 presents detailed cell mitosis detection results on sequence C3H10T1/2-1. Here, the centers of the yellow circles and green squares represent the detection results and annotations, respectively. To depict the detailed detection results along the temporal dimension, we also present the patch sequences between frames 122 and 129 in Fig. 13. The cell mitosis events were de-

tected within 1-frame difference, demonstrating that F3D-CNN can effectively model the spatiotemporal information of cell mitosis events.

To further quantitatively evaluate temporal localization performance of proposed method, timing error of detecting mitosis events in terms of frame difference is adopted as a metric. In particular, we keep considering the detection result as true positive only when timing error is not greater than 1. The distribution of the frame differences between annotations and true positive samples on C3H10T1/2 and C2C12 is shown in Fig. 14. The better performance on C2C12 sequences possibly because that its brightness characteristic is more distinct than that in C3H10T1/2.

6. Conclusion

This paper proposed a deep neural network based framework (F3D-CNN) for automatic cell mitosis detection from captured microscopy images. F3D-CNN employs FCNs and 3D-CNNs for candidate mitosis detection and further discrimination, respectively. Representative features are learnt automatically instead of handcraft feature designing. The F3D-CNN architecture is carefully designed to deal with the challenges of cell image processing. Specifically, the point annotations are processed by a smoothing and cropping strategy in the candidate detection stage, which reduces the memory cost and alleviates the imbalance training set problem. In the training phase, the multiple loss layers in the FCNs mitigate the gradient vanishing problem and explicitly guide the hidden layers to learn multi-scale features. In the mitosis discrimination stage, candidates with spatiotemporal context, which are further discriminated by a well-trained 3D-CNN. Finally, the F3D-CNNs were validated on C3H10T1/2 and C2C12 datasets released by the CMU cell image analysis group. In terms of performance metrics of cell mitosis detection, F3D-CNN outperformed competing state-of-the-art methods.

Acknowledgment

This work was supported by the National Science Foundation of China (Grant No. 61402306 and 61432012).

References

[1] S. Huh, D.F.E. Ker, R. Bise, M. Chen, T. Kanade, Automated mitosis detection of stem cell populations in phase-contrast microscopy images, IEEE Trans. Med. Imaging 30 (3) (2011) 586–596.
[2] N. Harder, F. Mora-Bermúdez, W.J. Godinez, A. Wünsche, R. Eils, J. Ellenberg, K. Rohr, Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time, Genome Res. 19 (11) (2009) 2113–2124.
[3] Z. Wang, W. Guo, L. Li, B. Luk'yanchuk, A. Khan, Z. Liu, Z. Chen, M. Hong, Optical virtual imaging at 50 nm lateral resolution with a white-light nanoscope, Nat. Commun. 2 (2011) 218.
[4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[5] Y. Wang, H. Mao, Z. Yi, Protein secondary structure prediction by using deep learning method, Knowl. Based Syst. 118 (2017) 115–123.
[6] Q. Guo, J. Jia, G. Shen, L. Zhang, L. Cai, Z. Yi, Learning robust uniform features for cross-media social data by using cross autoencoders, Knowl. Based Syst. 102 (2016) 64–75.
[7] J. Chen, H. Zhang, H. Mao, Y. Sang, Z. Yi, Symmetric low-rank representation for subspace clustering, Neurocomputing 173 (2016) 1192–1202.
[8] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2013, pp. 411–418.
[9] H. Chen, X. Wang, P.A. Heng, Automated mitosis detection with deep regression networks, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, 2016, pp. 1204–1207.
[10] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.
[11] A. Shkolyar, A. Gefen, D. Benayahu, H. Greenspan, Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, 2015, pp. 743–746.
[12] Y. Zhu, S. Lucey, Convolutional sparse coding for trajectory reconstruction, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 529–540.
[13] S. Ji, W. Xu, M. Yang, K. Yu, 3d convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231.
[14] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.
[15] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.
[16] CMU cell image analysis group, <http://www.celltracking.ri.cmu.edu/>.
[17] A. Liu, K. Li, T. Kanade, Mitosis sequence detection using hidden conditional random fields, in: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, 2010, pp. 580–583.
[18] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, T. Darrell, Hidden conditional random fields, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1848–1852.
[19] T. He, H. Mao, J. Guo, Z. Yi, Cell tracking using deep neural networks with multi-task learning, image and vision computing.
[20] K.E. Magnusson, J. Jaldén, P.M. Gilbert, H.M. Blau, Global linking of cell tracks using the viterbi algorithm, IEEE Trans. Med. Imaging 34 (4) (2015) 911–929.
[21] K. Zhang, L. Zhang, M.-H. Yang, Real-time compressive tracking, in: Proceedings of the European Conference on Computer Vision (ECCV), Springer, 2012, pp. 864–877.
[22] T. Gilad, M.-A. Bray, A.E. Carpenter, T.R. Raviv, Symmetry-based mitosis detection in time-lapse microscopy, in: Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, IEEE, 2015, pp. 164–167.
[23] F. Amat, W. Lemon, D.P. Mossing, K. McDole, Y. Wan, K. Branson, E.W. Myers, P.J. Keller, Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data, Nat. Meth. 11. 9 (2014) 951–958.
[24] M. Kandemir, C. Wojek, F.A. Hamprecht, Cell event detection in phase-contrast microscopy sequences from few annotations, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 316–323.
[25] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026–1034.
[26] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2014, pp. 818–833.
[27] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision (CVPR), 2015, pp. 1520–1528.
[28] H. Chen, Q. Dou, X. Wang, J. Qin, P.A. Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[29] W. Shen, K. Zhao, Y. Jiang, Y. Wang, Z. Zhang, X. Bai, Object skeleton extraction in natural images by fusing scale-associated deep side outputs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[30] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision (CVPR), 2015, pp. 1395–1403.
[31] K. Li, E.D. Miller, M. Chen, T. Kanade, L.E. Weiss, P.G. Campbell, Computer vision tracking of stemness, in: Proceedings of 5th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, 2008., IEEE, 2008, pp. 847–850.
[32] A.Z.K. Simonyan, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015.
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.
[34] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3d convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, 2015, pp. 4489–4497.