3D Convolutional Networks-Based Mitotic Event Detection in Time-Lapse Phase Contrast Microscopy Image Sequences of Stem Cell Populations

Wei-Zhi Nie†,Wen-Hui Li†, An-An Liu†*, Tong Hao‡, Yu-Ting Su†
† School of Electronic Information Engineering, Tianjin University, China
‡ Tianjin Key Laboratory of Animal and Plant Resistance/College of Life Science,
Tianjin Normal University, China
*Corresponding author: anan0422@gmail.com

Abstract

In this paper, we propose a straightforward and effective method for mitotic event detection in time-lapse phase contrast microscopy image sequences of stem cell populations. Different from most of recent methods leveraging temporal modeling to learn the latent dynamics within one mitotic event, we mainly target on the data-driven spatio-temporal visual feature learning for mitotic event representation to bypass the difficulties in both robust hand-crafted feature designing and complicated temporal dynamic learning. Specially, we design the architecture of the convolutional neural networks with 3D filters to extract the holistic feature of the volumetric region where individual mitosis event occurs. Then, the extracted features can be directly feeded into the off-the-shelf classifiers for model learning or inference. Moreover, we prepare a novel and challenging dataset for mitosis detection. The comparison experiments demonstrate the superiority of the proposed method.

1. Introduction

Mitotic behavior analysis during stem cell proliferation is an important indicator for cancer screening and assessment, tissue engineering and so on. Previously, this work could only be qualitatively performed with manual annotation by biologists [17][2][21]. To realize the quantitative analysis of high-throughput cell data, it is necessary to develop sophisticated methods for automatic mitotic event identification and localization.

1.1. Background and Related Work

The state-of-the-art methods for mitotic event detection in microscopy image sequences can be generally classified into three following categories.

- Tracking-free approaches detect mitosis directly in an image sequence without relying on cell trajectories [13]. Liu *et al.* [16] regarded individual mitotic cell as a special visual pattern. They extracted the area features (area/convex are), the shape features (eccentricity/major axis length/minor axis length/orientation), and the intensity features (maximum/average/minimum intensity) and trained the SVM classifier for identification. They further leveraged the sparse coding method [12] and multi-task learning method [18] for feature representation and cell modeling. Parthipan *et al*. [20] proposed the detection method, which can measure the deformation of the embryonic tissue and then utilize the intensity changes of the salient regions to identify the locations of mitosis events. Olcay *et al*. [19] proposed a computer-assisted system which can detect the mitotic cells based on the likelihood functions estimated from the samples of manually marked regions.

- Tracking-based approaches mainly aim to extract individual cell trajectories. Then, the mitotic event can be identified based on temporal progression of cell features and/or the connectedness between the segmented mother and daughter cells. Li *et al*. [10] proposed a cell tracking method with the fusion of multiple filters to deal with cell tracking and mitosis detection in a unified framework. Ryoma *et al.* [1] proposed a cell tracking method based on partial contour matching that is capable of robustly tracking partially overlapping cells, while maintaining the identity information of individual cells throughout the process from their initial contact to eventual separation. They further proposed a cell tracking method based on global spatiotemporal data association which considered hypotheses of initialization, termination, translation, division and false positive in an integrated formulation. Sungeun *et al*. [3] proposed an automated tracking method that can facilitate the study and expansion of Hematopoietic stem cells(HSCs) by monitoring the behavior of the

cells in vitro under serum-free media conditions using phase contrast microscopy.

- Temporal modeling methods are the mainstream approaches since they can overcome the limitations of both tracking-based and tracking-free approaches [17][14]. This kind of methods typically consist of candidate detection, feature extraction, and mitosis classification as consecutive steps. Gallardo *et al.* [4] adopted a hidden Markov model (HMM) to classify candidates based on temporal patterns of cell shape and appearance features. Liang *et al.* [11] utilized a conditional random field (CRF) model [9] to identify cell cycle phases. This method can effectively handle two challenges, including the small size of training samples and the datasets obtained under different conditions. Zhou *et al*. [23] used Markov model to proposed an automated quantitative analysis system that can be used to segment, track, and quantize cell cycle behaviors. Liu *et al*. [15] developed the chain-structured hidden-state conditional random fields model for mitosis sequence classification by learning the latent states and the transition between adjacent states within one mitosis progression. Seungil *et al*. [6] proposed the even-detection conditional random fields model which can not only determine whether a mitotic event occurs, but also provide the time at which the mitosis is completed and daughter cells are born. Most recently, Liu *et al*. [17] designed a semi-Markov model (SMM) to model four biological states within one mitotic event and proposed the max-margin theory-based method for model learning. This method is the first one that can simultaneously identify mitosis events and localize four main stages within cell cycle.

1.2. Our Approach

Although the recent novel methods advance in the way of temporal modeling, which is independent on the accurate localization and trajectory extraction of cells and has show superiority to the other two methods, they still meet a serious problem whether the hand-crafted visual features (such as HoG, SIFT, GIST [15]) can really be implemented for temporal modeling by the proposed models (such as HMM, CRF, HCRF, EDCRF, SMM). These features are subjectively designed for object detection, salient point detection and description, scene classification and so on but not specifically designed for cells/mitotic cells, which are non-rigid objects with drastic and random shape variation. Therefore, it is impossible to discover whether the feature sequence representation of one mitotic event lies in a manifold space which exactly fits the biological state space of the mitotic event. Consequently, we argue that it is essential to develop the data-driven feature learning method for mitotic event representation. Particularly, we hope that the designed feature can be directly utilized together with any off-the-shelf and practical classifier for mitotic event modeling and inference to avoid the complicated temporal state discovery and state transition modeling. Inspired by the deep learning breakthroughs in the image domain where rapid progress has been made in the past few years, we propose the 3D convolutional neural networks (3D CNN)-based method to extract the holistic feature of the volumetric region where individual mitosis event occurs. Although various pre-trained convolutional network models are made available for extracting image features, these models working on 2D images can not be directly utilized for mitotic event since it is a dynamic progression with obvious appearance evolution and motion characteristic. In this paper we elaborate the architecture of 3D CNN for spatio-temporal feature extraction. Especially, we aim to discover the structure of spatio-temporal filter and the pooling strategy to learn a hierarchy of increasingly complex features for mitosis event. We empirically show that the 3D CNN features with either a simple linear SVM classifier or the softmax layer widely used in CNN can obtain significantly better performances against the hand-crafted visual features [2]. Although 3D CNN has been designed several years ago [7][22], to the best of our knowledge, this paper is the first work to explore its application for mitosis event detection. Furthermore, we prepare a new mitosis dataset in hopes that this will inspire other researchers to tackle this challenging task. The comparison experiments demonstrated its superiority for this challenging task.

The rest of this paper is structured as follows. The proposed method is detailed in Section 2. The dataset, the experimental method and results will be shown in Section 3. At last, we conclude the paper in Section 4.

2. Methodology

In this section, we will explain in detail the basic operations of 3D CNN, including the architecture of the network and the parameters setting. The extracted features will be feeded into the off-the-shelf classifier for modeling and prediction.

2.1. Feature Learning via 3D CNN

3D CNN is leveraged to extract a hierarchy of increasingly complex features for the volumetric regions where mitotic events occur. Compared to 2D CNN, 3D CNN has the ability to model temporal information better owing to 3D convolution and 3D pooling operations. The convolution and pooling layers of 3D CNN fully consider spatial and temporal information of image sequence while 2D CNN can not do it. Figure 1 shows the architecture of 3D CNN.

We first introduce the definitions of several variables. One image sequence, which corresponds to individual mi-

Figure 1. The architecture of 3D CNN. The designed 3D CNN has 5 convolution, 5 max-pooling, and 2 fully connected layers. All 3D convolution kernels are $3 \times 3 \times 3$ with stride 1 in both spatial and temporal dimensions. The pooling layers are denoted from pool1 to pool5. All pooling kernels are $2 \times 2 \times 2$. Each fully connected layer has 2048 output units.

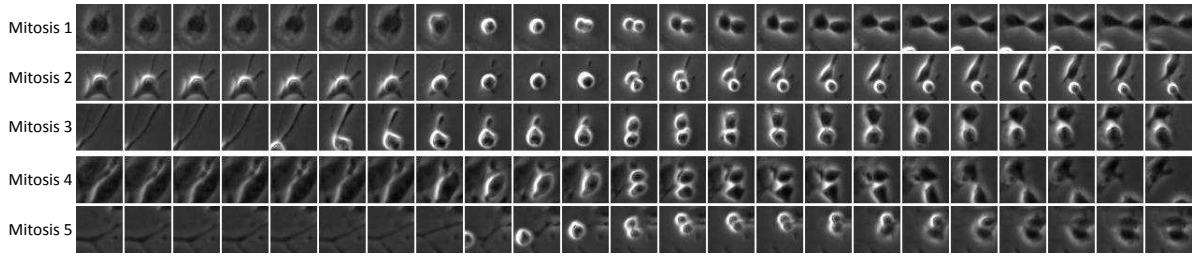


Figure 2. Mitosis samples from the C2C12 dataset.

totic event, is represented as $l \times h \times w$, where l is the number of frames, h and w are the height and width of each frame, respectively. The 3D convolution and pooling kernel sizes are represented by $d \times k \times k$, where d is the temporal depth of one kernel and k is the spatial size of one kernel. When $k = 1$, it means that 3D CNN is identical to 2D CNN.

The networks are set up to take the image sequence of individual mitotic candidate as inputs and predict the label (mitosis or non-mitosis). The networks have 5 convolution layers and 5 pooling layers (each convolution layer is immediately followed by a pooling layer), 2 fully-connected layers. The output of the final fully-connected layer is extracted as the final feature representation to represent each mitosis candidate. The number of filters for 5 convolution layers from 1 to 5 are 64, 128, 256, 256, 256, respectively. All convolution kernels have a size of d, which is the kernel temporal depth[1]. All of these convolution layers are applied with appropriate padding (both spatial and temporal) and stride 1, thus there is no change in term of size from the input to the output of these convolution layers. All pooling layers are Max pooling[2] with kernel size $2 \times 2 \times 2$ and stride 1 which means the size of output signal is reduced by a factor of 8 compared with the input signal. Finally, two fully connected layers have 2048 outputs. We train the networks from scratch using mini-batches of 30 clips, with initial learning rate of 0.003. The learning rate is divided by 10 after every 4 epochs. The training is stopped after 16 epochs.

[1]In Section 3.2, we will tune d to search the best parameter to represent the temporal dynamics within one mitosis event

[2]The strategy of pooling will be discussed in Section 3.4

2.2. Mitosis Modeling

One important goal of this work is that the extracted deep features for mitosis events can be directly feeded into the off-the-shelf and practical classifiers for event modeling without specific graph structure designing as did for the temporal modeling methods [6][5]. In our work, we applied the classic SVM classifier with the linear kernel and the softmax layer widely used in CNN for mitosis modeling and prediction. We will directly compare the performances of the proposed method against the SVM classifiers trained with other popular visual features.

3. Experiment

3.1. Dataset

Eight phase contrast image sequences of C2C12 myoblastic stem cells populations were acquired, each containing 1013 images. C2C12 myoblastic stem cells (ATTC, Manassas, VA) have the capacity to differentiate into osteoblasts and myocytes and were grown in DMEM, 10% bovine serum (Invitrogen, Carlsbad, CA) and 1% penicillin-streptomycin (PS; Invitrogen, Carlsbad, CA). All cells were kept at 37 C, 5% CO in a humidified incubator. Phase contrast images of growing stem cells were acquired every 5 min using a Zeiss Axiovert T135V microscope (Carl Zeiss Microimaging, Thornwood, NY). The microscope is equipped with a phase contrast objective (5X, NA 0.15), a custom-stage incubator, and the InVitro software (Media Cybernetics Inc., Bethesda, MD). Every image contains 1392×1040 pixels with a resolution of 1.3 $\mu m/pixel$.

This dataset was prepared by the CMU cell image analysis group for cell tracking [8]. Since there is lack of large-scale dataset to evaluate the algorithms for mitotic event de-

# Sequence	F0002	F0005	F0006	F0007
# Mitosis	501	304	225	320
# Sequence	F0008	F0014	F0015	F0016
# Mitosis	329	398	222	166

Table 1. Statistic of mitotic events in each sequences

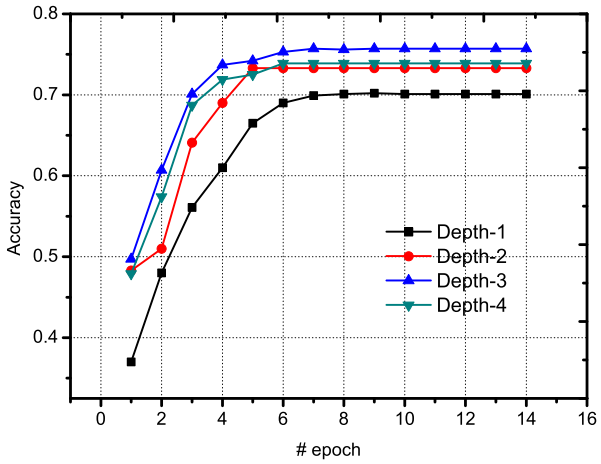
tection, we manually annotated the mitotic events on these C2C12 image sequences. For each mitotic event, the center of the boundary between two daughter cells was marked when the boundary is clearly observed. The numbers of annotated mitotic events in individual sequences are shown in Table 1. Some mitosis events are shown in Figure 2. This dataset is much more challenging than the previous C3H10T1/2 dataset [15][17] since: 1) C2C12 myoblasts were cultured to a much higher level of confluence as shown in Figure 7; 2) there exist much more mitosis events in this dataset comparing against the C3H10T1/2 dataset only contains 41-128 mitotic events in the five sequences [17].

In our experiment, we utilized the F0002 sequence, which contains the most mitotic events, as the training data and treated all the others as the test data. We extracted the mitotic samples (positive samples) by centering on the annotated locations and extracting a $23 \times 50 \times 50$ volumetric regions. Consequently, the dimension of individual input candidate is $23 \times 50 \times 50$. 1024 negative samples are generated randomly with the same spatial and temporal scales in the F0002 sequence.

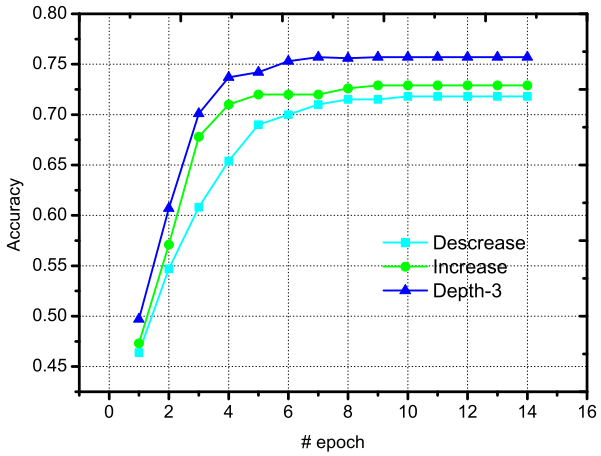
3.2. Varying kernel temporal depth

The main contribution of this work is that 3D CNN is applied to explicitly extract temporal features of each mitosis in the data-driven manner. Thus, the goal of this section is to explore how to aggregate temporal information through the deep networks. In order to handle this problem, we vary kernel temporal depth d of the convolution layers while keeping all other common settings fixed. Here, we fix the kernel spatial size with $k = 3$ and fix the kernel size of pooling with $2 \times 2 \times 2$.

In the experiment, we evaluated two types of networks: 1) homogeneous temporal depth: all convolution layers have the same kernel temporal depth; 2) varying temporal depth: kernel temporal depth is changing across the layers as did in [22]. For the first setting, we experiment with 4 networks and the kernel temporal depth, d, equal to 1, 2, 3, 4. These networks are terms as **depth-d**, where d is their temporal depth. Note that **depth-1** has the same architecture with 2D CNN. For the varying temporal depth setting, we evaluated two networks with temporal depth, **increasing**: 2-2-3-3-5 and **decreasing**: 5-3-3-2-2 from the first to the fifth convolution layer, respectively. The classic SVM is utilized model learning and inference. In this study, we randomly selected the positive and negative samples from F0002 for model learning and the mitotic samples from F0008 for test. The corresponding experimental results are shown in Figure 3. Figure 3(a) shows the results of 3D CNN with homogeneous temporal depth and Figure 3(b) shows the results of 3D CNN with varying kernel temporal depth.



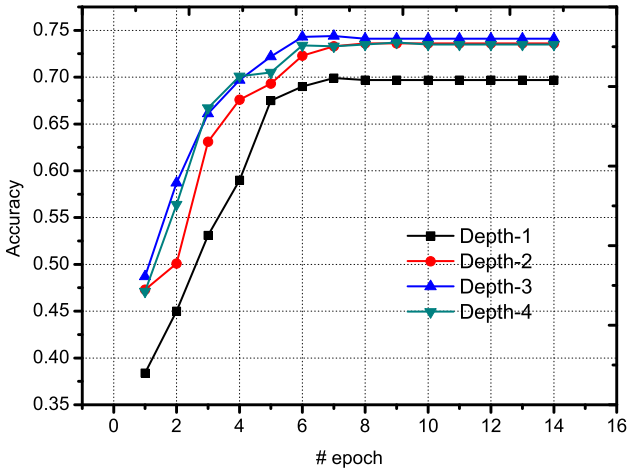
(a)



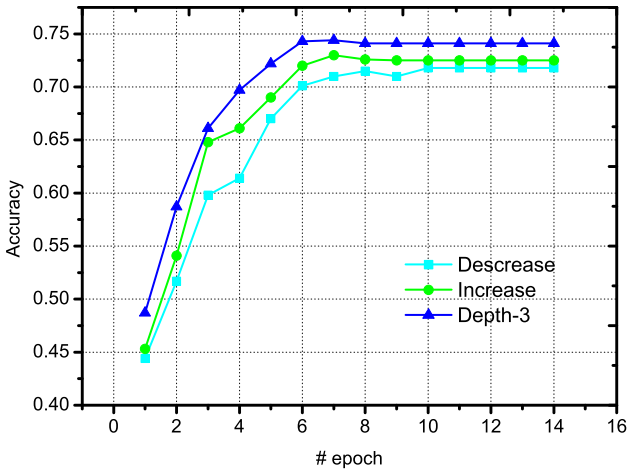
(b)

Figure 3. Tuning kernel temporal depth with 3×3 spatial kernel size.

From these experimental results, we can find that **depth-3** has the best performance among other networks. As expected, **depth-1** (2D CNN) has the worst results since it is lack of temporal information for feature learning. Compared to the varying temporal depth nets, **depth-3** in the homogenous manner can consistently outperform the others. In order to further demonstrate this conclusion, we change the spatial size (4×4) to repeat the previous work. The corresponding experimental results are shown in Figure 4. Figure 4(a) shows the results of networks with homogeneous temporal depth and Figure 4(b) shows the results of networks with varying kernel temporal depth. The same trends can be observed. These comparison suggests $3 \times 3 \times 3$ is



(a)



(b)

Figure 4. Tuning kernel temporal depth with 4×4 spatial kernel size.

the best kernel for 3D CNN-based mitotic event detection on this dataset.

3.3. Varying kernel spatial size

Based on the aforementioned experiments, we have found that 3D CNN with the kernel temporal depth $d = 3$ has the best performance. In this section, we vary the spatial size of network to find the best kernel spatial size k. Specifically, we tuned the kernel spatial size of k with 2 and 3. SVM is utilized for classification. We utilized the same training and test datasets in Section 3.2. From Figure 5, it is obvious that the kernel with the spatial size $k = 3$ can achieve the best performance.

3.4. Varying pooling strategy

The previous experiments discovered that the homogeneous setting with the 3D convolution kernels, $3 \times 3 \times 3$, is the best configuration of 3D CNN for mitotic event detection on this dataset. On the other hand, the pooling strategy is also one important element for spatio-temporal feature learning. In this experiment, we empirically compared two widely used pooling kernel sizes, $2 \times 2 \times 2$ and $3 \times 3 \times 3$. Meanwhile, we applied two common pooling styles, including Max and Average, to find the best pooling strategy. SVM was selected as the classifier and the same training and test datasets were utilized as Section 3.2. The experimental results are shown in Figure 6. From Figure 6, it is obvious that the pooling kernel, $2 \times 2 \times 2$, with Max style can outperform the others.

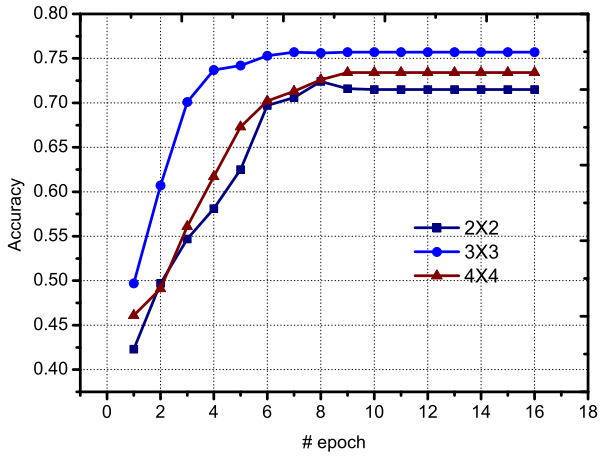
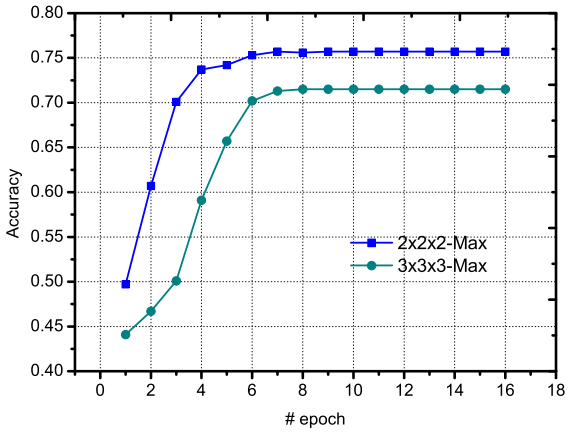


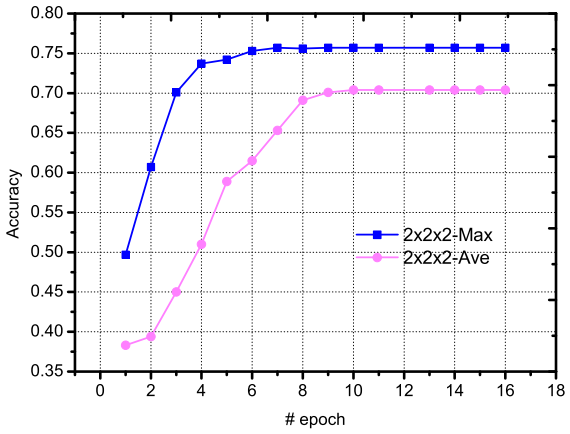
Figure 5. Tuning kernel spatial size with the temporal kernel size $d = 3$.

3.5. Quantitative comparison with different visual features

Based on the previous experiments, we discovered that the homogeneous setting with the 3D convolution kernel of $3 \times 3 \times 3$ and the pooling kernel size with $2 \times 2 \times 2$ together with the pooling style of Max is the best configuration for spatio-temporal feature learning for this task. In order to demonstrate the superiority of the extracted 3D CNN features, we compared its performances with the SVM classifier and the softmax layer against several representative visual features, including Sift, Gist, and HoG. Since Sift, Gist, and HoG are designed for the description of the 2D region and can not directly represent temporal information, we leveraged the popular bag-of-visual-word strategy to convey the spatio-temporal information. Specifically, each mitotic candidate sequence is considered as a bag of N images, each of which corresponds to one frame in the sequence. We trained a dictionary with 500-D codewords. Then each sequence can be represented by a 500-D visual features which implicitly convey both spatial and temporal characteristics. We also utilized the SVM classifier for model learning. In our experiment, F0002 was utilized for model learning and all the others were implemented



(a)



(b)

Figure 6. 3D CNN with different kernels size and types of pooling.

for test. The accuracy ($\#Turepositive/\#Groundtruth$) was computed for evaluation. The experimental results are shown in Table.2.

From Table 2, it is obvious that the performances with 3D CNN features can consistently outperform HoG, Gist, and Sift in terms of Accuracy. Even with the simple softmax classifier, 3D CNN features can still work better than the others. A students paired t-test on the Accuracies shows that the improvements of P-value are statistically significant at the significance level 0.01 except the case of "3D CNN+Softmax versus HoG+SVM" (3D CNN+SVM versus HoG+SVM, Gist+SVM, Sift+SVM: 0.0105 0.0033 0.0017; 3D CNN+Softmax versus HoG+SVM, Gist+SVM, Sift+SVM: 0.066 0.0055 0.0032). It denotes that 3D CNN can automatically mine the spatial and temporal characteristics with the preset architecture in the data-driven manner. This advantage can especially fit the task of mitosis detection since it is difficult to design a reasonable features for the mitotic cell, which is non-rigid object with random shape variance but without obvious contour, texture and intensity information.

3.6. Qualitative comparison and discussion

To further show that 3D CNN feature is discriminative for spatio-temporal representation, we design a more straightforward comparison for qualitative analysis. We selected an image segment from F0005 with 23 frames ($\#frame: 905-927$, image resolution: 1392×1040) and high cell crowd density as shown in Figure 7. To test the generalization ability of individual classifiers trained with different features and also bypass the difficulty in mitotic candidate extraction [17], we implemented the sliding window strategy (step = 10 pixels) to generate all possible $23 \times 50 \times 50$ volumes for classification. If the output score of the classifier is above 0.5, the candidate is classified as mitosis. The detection results by 3D CNN+SVM, HoG+SVM, and Gist+SVM are shown in Figure. 7.

From Figure 7(a,c,d), it is obvious to see that 3D CNN features are more robust and discriminative to distinguish each candidate region as mitosis or not by leveraging both spatial and temporal context. Therefore, the classifier trained with 3D CNN features only generated 4 false positive samples and 4 false negative samples. Comparatively, the classifiers trained with HoG and Gist respectively generated 9 and 5 false negative samples and too many false positive samples. Figure 7(b) shows the TP/FP/FN samples. The TP sample contains the spatio-temporal characteristics of mitosis and can be correctly classified. The FP sample shows a quite similar spatio-temporal appearance due to the irregular variation and consequently was falsely considered as a mitotic event. Due to the high confluence, two mitotic events occur near to each other. Consequently, both candidate sequences (the FN samples) contain the visual interference from each other and were falsely regarded as non-mitosis.

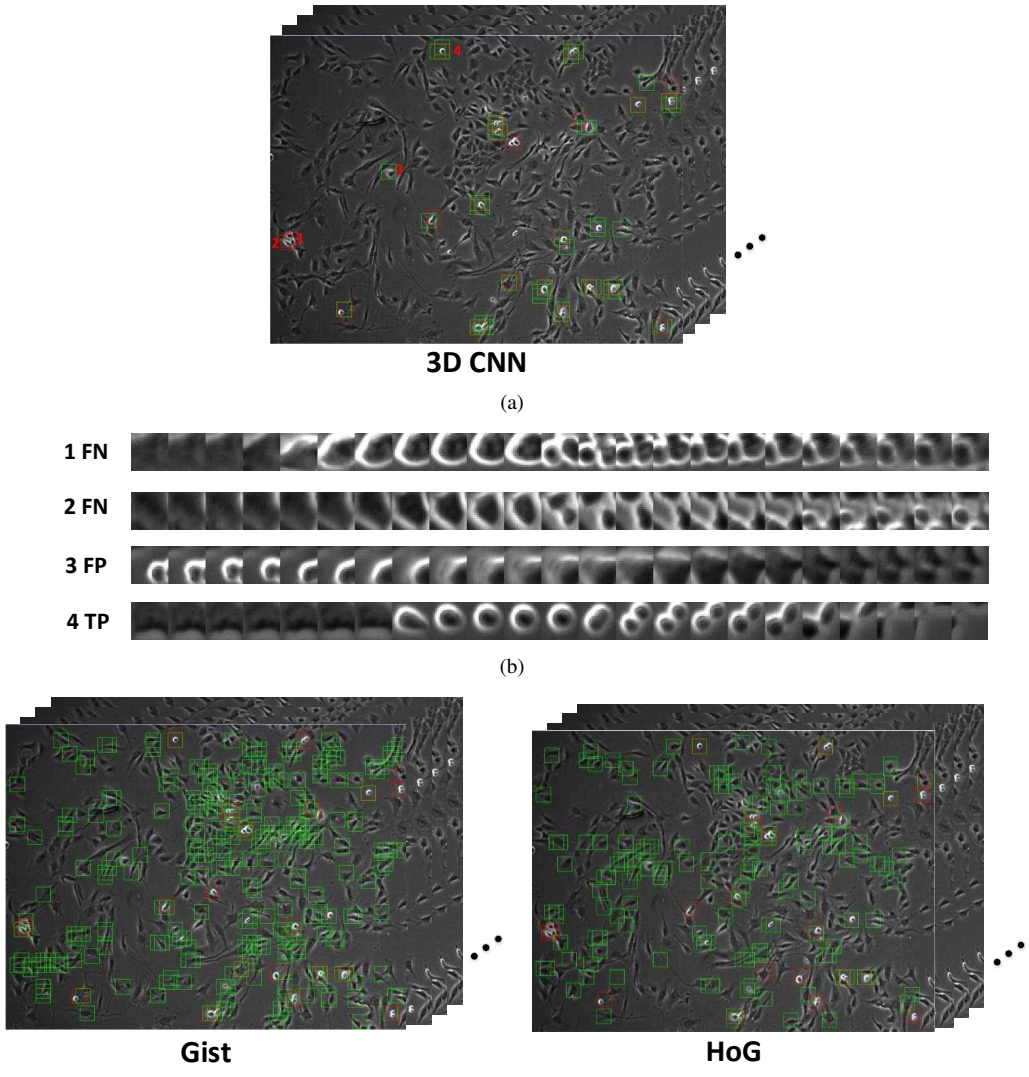
4. Conclusion and Future Work

In this work, we aim to address the problem of mitosis event detection based on the spatiotemporal features learned by 3D CNN. We conducted a systematic study to find the optimal temporal kernel length, the optimal spatial kernel size and the optimal pooling strategy for 3D CNN. We experimentally demonstrated that 3D CNN can outperform the 2D CNN features and the popular hand-crafted features for visual representation.

Our future work will focus on the direct comparison between the 3D CNN-based method and the temporal modeling methods to test whether the proposed method can explicitly outperform the state-of-the-art performances. Moreover, we plan to integrate both 3D CNN for spatio-temporal feature learning and LSTM for temporal context learning together to tackle this task.

Sequence#	HoG+SVM	Gist+SVM	Sift+SVM	3D CNN+Softmax	3D CNN+SVM
F0005	75.5	68.5	65.7	76.4	**79.4**
F0006	73.1	70.6	71.1	73.6	**75.5**
F0007	72.0	70.7	66.9	74.8	**77.7**
F0008	72.6	69.8	50.2	74.6	**75.7**
F0014	73.7	71.6	61.8	72.9	**73.8**
F0015	74.7	68.7	62.2	78.7	**84.7**
F0016	74.3	71.0	69.3	82.6	**82.6**
Average	73.7	70.1	63.9	76.3	**78.5**

Table 2. Performance comparison with different features (%).



(a)

(b)

(c)

(d)

Figure 7. Detection results by different methods (Green: detected region; Red: groundtruth; TP: true positive; FP: false positive; TN: true negative; FN: false negative).

5. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China

(61472275,61502337,61100124), the Tianjin Research Program of Application Foundation and Advanced Technology (15JCYBJC16200), the grant of China Scholarship Council (201506255073), the grant of Elite Scholar Program of Tianjin University (2014XRG-0046).

Tong Hao's work was supported by National High-Tech Research and Development Program of China (863 programs, 2012AA10A401 and 2012AA092205), Grants of the Major State Basic Research Development Program of China (973 programs, 2012CB114405), National Natural Science Foundation of China (21106095), National Key Technology R&D Program (2011BAD13B07 and 2011BAD13B04), Tianjin Research Program of Application Foundation and Advanced Technology (15JCYBJC30700), Project of introducing one thousand high level talents in three years, Foundation of Introducing Talents to Tianjin Normal University(5RL123), "131" Innovative Talents cultivation of Tianjin, Academic Innovation Foundation of Tianjin Normal University (52XC1403).

References

[1] R. Bise, K. Li, S. Eom, and T. Kanade. Reliably tracking partially overlapping neural stem cells in dic microscopy image sequences. In *MICCAI Workshop on OPTIMHisE*, volume 5, 2009.

[2] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 411–418. Springer, 2013.

[3] S. Eom, S.-i. Huh, D. Ker, R. Bise, and T. Kanade. Tracking of hematopoietic stem cells in microscopy images for lineage determination. *J Latex Class Files*, 6(01-09):9, 2007.

[4] G. M. Gallardo, F. Yang, F. Ianzini, M. Mackey, and M. Sonka. Mitotic cell recognition with hidden Markov models. volume 5367, pages 661–668, 2004.

[5] S. Huh and M. Chen. Detection of mitosis within a stem cell population of high cell confluence in phase-contrast microscopy images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1033–1040. IEEE, 2011.

[6] S. Huh, D. F. E. Ker, R. Bise, M. Chen, and T. Kanade. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *Medical Imaging, IEEE Transactions on*, 30(3):586–596, 2011.

[7] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):221–231, 2013.

[8] T. Kanade, Z. Yin, R. Bise, S. Huh, S. Eom, M. F. Sandbothe, and M. Chen. Cell image analysis: Algorithms, system and applications. In *IEEE Workshop on Applications of Computer Vision (WACV 2011), 5-7 January 2011, Kona, HI, USA*, pages 374–381, 2011.

[9] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289, 2001.

[10] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell. Cell population tracking and lineage construction with spatiotemporal context. *Medical image analysis*, 12(5):546–566, 2008.

[11] L. Liang, X. Zhou, F. Li, S. T. Wong, J. Huckins, and R. W. King. Mitosis cell identification with conditional random fields. In *Life Science Systems and Applications Workshop, 2007. LISA 2007. IEEE/NIH*, pages 9–12. IEEE, 2007.

[12] A. Liu, Z. Gao, T. Hao, Y. Su, and Z. Yang. Sparse coding induced transfer learning for hep-2 cell classification. *Biomedical Materials and Engineering*, 24(1):237–243, 2014.

[13] A. Liu, T. Hao, Z. Gao, Y. Su, and Z. Yang. Nonnegative mixed-norm convex optimization for mitotic cell detection in phase contrast microscopy. *Comp. Math. Methods in Medicine*, 2013:176272:1–176272:10, 2013.

[14] A. Liu, T. Hao, Z. Gao, Y. Su, and Z. Yang. Sequential sparse representation for mitotic event recognition. *Electronics Letters*, 49(14):869–870, 2013.

[15] A. Liu, K. Li, and T. Kanade. Mitosis sequence detection using hidden conditional random fields. In *ISBI*, pages 580–583, 2010.

[16] A. Liu, K. Li, and T. Kanade. Spatiotemporal mitosis event detection in time-lapse phase contrast microscopy image sequences. In *IEEE Int. Conf. Multimedia and Expo*, pages 161–166, 2010.

[17] A. Liu, K. Li, and T. Kanade. A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *IEEE Trans. Med. Imaging*, 31(2):359–369, 2012.

[18] A. L. nad Yao Lu, W. Nie, Y. Su, and Z. Yang. Hep-2 cells classification via clustered multi-task learning. *Neurocomputing*, 2016.

[19] O. Sertel, U. V. Catalyurek, H. Shimada, and M. Guican. Computer-aided prognosis of neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 1433–1436. IEEE, 2009.

[20] P. Siva, G. W. Brodland, and D. Clausi. Automated detection of mitosis in embryonic tissues. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*, pages 97–104. IEEE, 2007.

[21] Y. Su, J. Yu, A. Liu, Z. Gao, T. Hao, and Z. Yang. Cell type-independent mitosis event detection via hidden-state conditional neural fields. In *11th International Symposium on Biomedical Imaging*, pages 222–225, 2014.

[22] D. Tran, L. D. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 4489–4497, 2015.

[23] X. Zhou, F. Li, J. Yan, and S. T. Wong. A novel cell segmentation method and cell phase identification using markov model. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):152–157, 2009.