A CASCADE OF CNN AND LSTM NETWORK WITH 3D ANCHORS FOR MITOTIC CELL DETECTION IN 4D MICROSCOPIC IMAGE

Titinunt Kitrungrotsakul [5,1] *Yutaro Iwamoto* [1] *Xian-Hau Han* [2] *Satoko Takemoto* [3]
Hideo Yokota [3] Sari Ipponjima [4] Tomomi Nemoto [4] Xiong Wei [5] Yen-Wei Chen [1⋆]

[1] Graduate School of Information Science and Engineering, Ritsumeikan University, Japan
[2] Faculty of Science, Yamaguchi University, Japan
[3] Center for Advanced Photonics, RIKEN, Japan
[4] Research Institute for Electronic Science, Hokkaido University, Japan
[5] Institute for Infocomm Research, A*Star, Singapore
⋆ corresponding author

ABSTRACT

Mitotic event detection is a fundamental step in investigating of cell behaviors. The event can be used to analyze various diseases, but most mitotic event detections performed previously focused only on two-dimensional (2D) images with time information. Owing to the complex background (normal cells) and mitotic event orientations, the 2D detection methods yield many false positive and false negative results. To solve this problem, we proposed a 2.5 dimensional (2.5D) cascaded end-to-end network combined with 3D anchors for accurate detection of mitotic events in 4D microscopic images. Our proposed network uses a convolutional long short-term memory to handle issues relating to time sequence; this helps to improve the detection accuracy (reduction of false positives). Furthermore, it uses 3D anchors to capture volume information used to address the orientation problem (reduction of false negatives). The experimental results show that the proposed method can achieve higher precision and recall compared with state-of-the-art methods.

Index Terms— Mitotic detection, deep learning, 2.5D network, 3D anchors

1. INTRODUCTION

Detection of mitotic events (cell division) is a fundamental step in cell investigation. Mitotic event detection can be used to understand cell behaviors, analyze diseases, and in many biomedical applications. Furthermore, it can be used to detect an abnormal skin structure in hairless mice [1, 2, 3, 4], or development of breast cancer cell [5]. Mitotic counts are usually conduced by specialists looking at glass slides under a microscope. Since, the process involves three-dimensional (3D) data, which is the time sequence of 2D images, it is a time-consuming and labor intensive task.

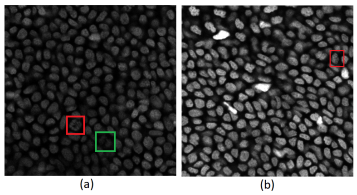

Fig. 1. Typical slice images. The red and green boxes indicate mitotic cells. (a) The division is in xy plane (easy-detection). (b) The division is along the z-axis (difficult-detection).

In an alternative study on glass slides, the researchers investigated the use of a two-photon microscope for intra-vital skin imaging. In this area, 4D data (time sequences of 3D volume images) are used to investigate cell behaviors [6]. In comparison with 3D microscopy data (time sequence of 2D image data), 4D data (time sequence of 3D volume data) can provide more information, but also requires more work to interpret. With 4D data, cell orientation becomes important, as oriented cell division is involved in determining the outcomes of various cell types, particularly during developmental periods [7, 8]. Typical slice images from two volume data are shown in Fig.1. The red and green boxes indicate mitotic cells. Fig.1(a) is a case of easy-detection, in which the cell division is clearly observed in the 2D axial plane. Fig.1(b) is a case of difficult-detection, in which the cell division cannot be clearly observed in the 2D image of each because the orientation of the cell is along the z-axis. In the case of Fig.1(b), volume information is required. Automatically identifying mitotic cells from such images is a challenging task.

Existing methods for mitotic and object detections are mostly developed by using 2D dynamic imaging techniques. Mao proposed the following two methods for identifying mitotic cells:hierarchial convolution neural network (HCNN) [9] and Two-Stream Bidirectional Long Short-Term Memory

This work is supported in part by Japan Society for Promotion of Science (JSPS) under Grant No. 16J09596 and KAKEN under the Grant Nos. 18H04747, 15H05953, 15H05954; and in part by A*STAR Research Attachment Programme.

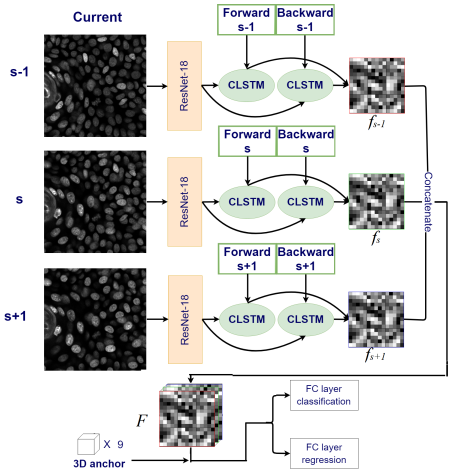

Fig. 2. Overview of CasDetNet_LSTM_3DAnchor network architecture; the top part is used to extract the 2D feature; while the bottom part uses 3D anchors to determine mitotic location.

(TS-BLSTM)[10]. Both methods feed two types of input into the network, namely appearance and motion. Applying each of the methods involves generating top-level features of appearance and motion, after which, these features are merged to form one hybrid feature. The TS-BLSTM is an improved version of the HCNN that takes advantage of Long Short-Term Memory(LSTM) to improve performance by considering temporal information. However, both methods are designed with a focus on 2D images and do not take cell orientation into consideration, thus, resulting in under-detection (false negatives).

In our previous work [11], we proposed a 2.5D cascaded convolutional neural network (CNN) with temporal information for automatic mitotic cell detection in 4D microscopic images. We used regional proposal network (RPN) to extract candidate cell from each slice and collected the RPN results of the target slice and its neighbor slices, which is known as 2.5D method, to support the classification and regression results. However, the method still lacked cell orientation information.

In this paper, we used the concept of 2.5D network [11] to propose a cascaded CNN and LSTM network with 3D anchors for mitotic cell detection in 4D microscopic images. First, we extract features from the time sequence of a selected slice and its neighbors using 2.5D convolutional and bidirectional LSTM networks, which can gather information from time sequences. To integrate spatial information, we merge the neighbor slices with the selected slice to form 3D features and apply 3D anchors to extract candidate ROIs including oriented mitotic cells. The list of ROIs is used to find the location of a mitotic cell in the image.

The paper is organized as follows. The proposed method using a 2.5D and 3D convolutional layers is introduced in Section 2, and experimental results of our proposed segmentation method are discussed in Section 3. Section 4 discusses the conclusion.

2. METHOD

The architecture of the proposed 2.5D cascaded CNN and LSTM with 3D anchors (CasDetNet_LSTM_3DAnchor) is illustrated in Fig. 2. It comprises two main parts: (1) a 2.5D cascaded CNN and LSTM which are used to extract spatial features by considering the temporal information; (2) the extracted features from the target slice are combined with features from the neighbor slices to form 3D features. Then 3D anchors are used for generating 3D ROIs of candidate mitotic cells. The ROIs are passed to the classification and regression layers to locate a mitotic cell on the target slice.

2.1. 2.5D Method for Feature Extraction

During cell development periods, not all mitotic event occur on the horizontal plane. Cell division orientation may occur on the vertical plane, in which 2D detection methods cannot be used to detect mitotic event with high accuracy. Conversely, the 3D method using volume data as input can solve the orientation problem, however, the number of training data (volume data) is very limited. This results in overfitting problem since the 4D (3D volume and time sequence) are rarely obtained in such scenario. The poor performance of the 3D method when the training data is small is discussed in Section 3.

To ameliorate this, we propose the use of 2.5D cascaded CNN and LSTM network as the detection network as shown in Fig.2. This is called as a 2.5D network because it takes images of slices $s-1, s, s+1$ as input and uses them to extract features of the target slice s. The advantage of our 2.5D network is that it can use information from neighboring slices (2.5D information) to distinguish between mitosis and normal cells. The ResNet-18 [12] is used as a backbone network. To extract features from sequential data, which can be used to reduce the false positives, we applied bidirectional convolutional LSTM (CLSTM) [13] after feature extraction using the backbone network. Eq.1 shows the formulas of CLSTM, where p_t is the input gate, f_t is the forget gate, c_t is cell state, o_t is the output gate, and h_t is the hidden state.

$$\begin{aligned} c_t &= f_t \circ c_{t-1} + p_t \circ tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \\ p_t &= \sigma(W_{xi} * x_t + W_{hp} * h_{t-1} + W_{cp} \circ c_{t-1} + b_p) \\ f_t &= \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\ o_t &= \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \\ h_t &= o_t \circ tanh(c_t) \end{aligned} \tag{1}$$

Using the proposed cascaded network, we obtained a 2D feature map represented as $f_{s+j}(j = -n, , -1, 0, 1, , n)$, where s is the target slice and $s + j$ are neighbor slices of the target slice. The number of neighbor slices is $2n + 1$. It should be noted that the 2D feature map f_s at time t is obtained using information about its neighbor slices ($s - 1$ and $s+1$) and its previous ($t-1$) and next ($t+1$) slices as shown in Fig.2. The 2D feature map $f_{s+j}(j = -n, , -1, 0, 1, , n)$

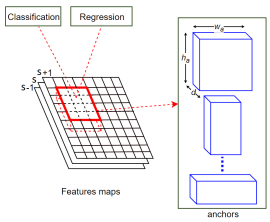

Fig. 3. 3D anchors on 3D feature map F

are concatenated into a 3D feature F, which is used as an input for the 3D anchor.

2.2. 3D Anchors for Mitotic Cell Prediction

To use the 3D feature F in an efficient way, we were motivated to employ the detection technique from RPN which is done using 3D anchors with 3 scales and 3 aspect ratios (9 anchors). We used 9 anchors sliding over the features F to generate a set of candidate ROIs. The size of the anchor is $w_a \times h_a \times d$ as shown in Fig.3. Since, the input of the network is specified by the number of neighbor slices N, the depth (d) of the anchor is equal to $2N+1$ and w_a, h_a are parameters to be estimated by regression together with the anchor center position at (x, y). Finally, the volume information (3D features) extracted from 3D anchors are fed into two sub-networks: classification $C^t_{s,i}$ and regression $R^t_{s,i}$. The output of the regression layer determines a predicted box (x, y, w, h), where i is the index of the ROI in slice s at time t. The classification layer determines the class of candidate ROIs.

2.3. Post-processing

Furthermore, to refine our network's results, we performed post-processing on the cells to remove cells erroneously-identified as normal cells and to add missing mitotic cells using information from the spatially and temporally adjacent slices, shown as follows.

$$\delta^{t+k}_{s+j,i} = \begin{cases} 1, & \text{if } C^{t+k}_{s+j,i} \geq 0.5 \\ 0, & \text{else} \end{cases} \tag{2}$$

$$W^{t+k}_{s+j,i} = 1 - \frac{\sqrt{j^2+k^2}}{\sqrt{N^2+T^2}} \tag{3}$$

$$\hat{C}^{t+k}_{s+j,i} = \frac{\sum_{j=-N}^{N}\sum_{k=-T}^{T} W^{t+k}_{s+j,i}\delta^{t+k}_{s+j,i}}{\sum_{j=-N}^{N}\sum_{k=-T}^{T} W^{t+k}_{s+j,i}} \tag{4}$$

Here, $C^t_{s,i}$ and $\hat{C}^t_{s,i}$ are probabilities before and after post-processing (refinement). The variables, N and T represent the numbers of spatially and temporally adjacent slices considered, respectively. The variable W is the weight distance between the target slice and the neighboring slice. $\hat{C}^t_{s,i}$ also represents the final classification result.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup

3.1.1. Dataset

The dataset in this work was obtained from JSPE, Technical committee on Industrial Application of Image Processing Appearance inspection algorithm contest 2017 (TC-IAIP A-IA2017), Japan [17]. In this dataset, there are 16 set of 4D data, each of which contained approximately 80 temporal frames and each frame was approximately 480×480×37. Each set contained 1-3 mitotic cells, listed in Table 2 as the ground truth. The mitotic cells were annotated, but the annotations did not indicate the stage of mitosis (unlike in many other papers). This means that we were applying all models to a binary classification task (mitotic or non-mitotic). Leave-one-out technique was used to train the model.

To increase the training samples, we augmented the data as follows. Because our network takes 2D images as input, we first sliced the 3D volumes horizontally to generate 2D images in the xy plane. Then, we rotated the data in 15° step, scaled the images to 0.8 and 1.2 using center cropping, and mirrored it to generate 96 times as much data.

Table 1. Quantitative comparison of our proposed methods and conventional methods for 2D image slices.

Method	Precision	Recall	F1 Score
2D Faster R-CNN [14]	0.0870	**0.9310**	0.509
3D Faster R-CNN	0.0592	0.4143	0.2367
SSD [15]	0.0411	0.7221	0.3816
HCNN [9]	0.7003	0.6910	0.6957
TS-BLSTM [10]	0.7883	0.7751	0.7817
2.5D Faster R-CNN [11]	0.3591	0.7532	0.5562
CasDetNet_CNN [11]	0.7228	0.70358	0.7132
CasDetNet_LSTM	0.8195	0.7974	0.8085
CasDetNet_LSTM_3DAnchor	**0.8356**	0.8442	**0.8399**

3.1.2. Implementation

Our models were trained with Keras and TensorFlow as its backend on an NVIDIA TITAN X GPU. Adam optimization method was used to train the model with an adaptive learning rate. The initial learning rate was 0.5×10^{-5} and the lowest learning rate observed was 10^{-7} with a batch size of 2. A Bidirectional technique was used in our CLSTM and the number of time sequence T was set to 6, and a total of 13 time sequences were used. The number of neighbor slices N was set to 4, which implies that 9 slices were used to determine the location of the mitotic cell along the z axis.

3.2. Detection Results on Slice Images

In this section, we present an evaluation of the detection results on 2D slice images. Each volume data was divided into multiple 2D slice images. Each slice image was considered as a sample data. The total number of 2D mitotic events in each volume data was about 800-1200 slices. The precision, recall, and F-1 score was used to evaluate the performance of the detection methods. To evaluate the performance of our network, we compared our network with the following stage-of-the-art networks: Single Shot MultiBox Detector (SSD)

Table 2. Detection results on 4D images to observe the orientation robustness.

Data		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sugano[16]	TP	1	1	2	3	2	1	0	1	2	2	1	1	0	2	2	2
	FN	0	0	0	0	1	0	1	0	0	0	1	2	1	0	0	0
	FP	0	3	0	0	0	0	0	0	0	9	0	0	0	1	17	6
Faster RCNN[14]	TP	1	1	2	3	3	1	1	1	2	2	2	3	0	2	2	2
	FN	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
	FP	8	5	9	12	16	4	4	7	8	7	6	9	8	4	15	11
CasDetNet_LSTM	TP	1	1	2	3	1	1	1	1	1	2	1	1	0	2	2	2
	FN	0	0	0	0	2	0	0	0	1	0	1	2	1	0	0	0
	FP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CasDetNet_LSTM_3DAnchor	TP	1	1	2	3	1	1	1	1	2	2	2	3	1	2	2	2
	FN	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	FP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ground truth		1	1	2	3	3	1	1	1	2	2	2	3	1	2	2	2

[15], FASTER RCNN [14], 3D FASTER R-CNN, Hierarchical Convolutional Neural Network (HCNN) [9], Two-Stream Bidirectional Long Short-Term Memory (TS-BLSTM) [10], and 2.5D network using 2D anchor [11].

3.2.1. Comparison with state-of-the-art methods

A comparison of the performance of our network with the performance of stage-of-the-art methods is shown in table 1. The conventional neural networks such as Faster RCNN and SSD did not perform well in mitotic event detection. The F1-score of both methods were only about 0.5. They could have achieved good scores in recall rate, however, most of the detected ROIs were non-mitotic cells which resulted in very low precision and F1 score. The 3D Faster RCNN did not perform well too because the number of 3D training samples were very limited. While HCNN and TS-BLSTM achieved better results than Faster RCNN and SSD, the recall was not high because they only used 2D information, and thus could not detect oriented mitotic cells. Compared with the state-of-the art methods, our proposed cascaded methods (CasDet_CNN, CasDetNet_LSTM and CasDetNet_LSTM_3DAnchor) can significantly reduce false positives and improve the precision, of mitotic events detection. Furthermore, the proposed method with 3D anchors can detect the oriented mitotic cells, improve the recall, and achieve better performance.

3.2.2. Results for 4D images

Our final goal was to detect mitotic cells using 4D data. In this case, we compared our results (with and without 3D anchors) for 4D images with the winner of the TC-IAIP AIA2017 contest [17], 2D Faster RCNN. Table 2 summarizes the detection results (TP, FN, FP) for the 4D data. The number of mitotic cells detected manually for each data is also shown in Table 2 as ground truth. Among the 16 data, No.11, 12, and 13 contained oriented mitotic cells, which were difficult to be detected using the conventional 2D method. As shown in Table 2, it can be seen that the winner of Sunagos method [16] (non-deep learning method based on machine learning), can detect almost all mitotic cells but could not detect the oriented mitotic cells in No.11, 12, and 13 data. Many false positive results were also observed. For conventional Faster RCNN, a lot of false positive result were detected. Our proposed cascaded CNN and LSTM without anchors (CasDetNet_LSTM) significantly reduced the number of false positive results (no false positives were detected), but could not detect the oriented mitotic cells in No.11, 12, and 13 data. Furthermore, some mitotic cells in No.5 and 9 were not detected. The proposed cascaded network with 3D anchors effectively detected the oriented mitotic calls in No.11, 12 and 13 data; moreover no false positive results were detected using the proposed network. It was only two mitotic cells in No.5 that were not detected by the proposed network because they were located along the border of the image.

4. CONCLUSION

We proposed a 2.5D cascaded CNN and LSTM network with 3D anchors for mitotic cell detection in 4D microscopic images. using the cascaded LSTM, we were able to significantly reduce the number of false positive results. Through the use of 3D anchors, we were able to detect oriented mitotic cells and significantly reduce the false negative results.

Our experimental results show that the proposed method can perform better than the state-of-the-art methods such as Faster RCNN, Single Shot Multi-Box Detector (SSD), Hierarchical Convolution Neural Network (HCNN), Two-Stream Bidirectional Long Short-Term Memory (TS-BLSTM), and the winner of the TC-IAIP AIA2017 contest. However, the proposed method still had issues in detecting cells that began mitosis outside the image and divided into two daughter cells that later appeared inside the image boundary. We plan to address this issue in our future work, which would aim to solve this problem and improve detection accuracy.

5. REFERENCES

[1] S. Ipponjim, T. Hibi, and T. Nemoto, "Three-dimensional analysis of cell division orientation in epidermal basal layer using intravital two-photon microscopy," *PLOS one*, 2016.

[2] YC. Hsu, L. Li, and E. Fuchs, "Emerging interactions between skin stem cells and their niches," *Nat. Med.*, 2014.

[3] P. Jones and BD. Simons, "Epidermal homeostasis: do committed progenitors work while stem cells sleep?," *Nat. Rev. Mol. Cell Biol.*, 2008.

[4] FM. Watt, "Mammalian skin cell biology: at the interface between laboratory and clinic," *Science*, 2014.

[5] P.J. van Diest, J.P.A. Baak, P. Matze-Cok, E.C.M. Wisse-Brekelmans, C.M. van Galen, P.H.J. Kurver, and et al, "Reproducibility of mitosis counting in 2,469 breast cancer specimens: Results from the multicenter morphometric mammary carcinoma project," *Human Pathology*, 1992.

[6] S. Ipponjima, T. Hibi, and T. Nemoto, "Three-dimensional analysis of cell division orientation in epidermal basal layer using intravital two-photon microscopy," *PLOS one*, 2016.

[7] D. Konno, G. Shioi, A. Shitamukai, A. Mori, H. Kiyonari, T. Miyata, and et al, "Neuroepithelial progenitors undergo lgn-dependent planar divisions to maintain self-renewability during mammalian neurogenesis," *Nat Cell Biol.*, pp. 93–110, 2008.

[8] M. Wu, CL. Smith, JA. Hall, I. Lee, K. Luby-Phelps, and MD. Tallquist, "Epicardial spindle orientation controls cell entry into the myocardium," *Dev Cell.*, pp. 114–125, 2010.

[9] Y. Mao and Z. Yin, "A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images," *MICCAI*, pp. 685–692, 2016.

[10] Y. Mao and Z. Yin, "Two-stream bidirectional long short-term memory for mitosis event detection and stage localization in phase-contrast microscopy images," *MICCAI*, 2017.

[11] T. Kitrungrotsakul, X.H. Han, Y. Iwamoto, S. Takemoto, H. Yokota, S. Ipponjima, T. Nemoto, X. Wei, and Y.W. Chen, "A 2.5d cascaded convolutional neural network with temporal information for automatic mitotic cell detection in 4d microscopic images," *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2018.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[13] S. Xingjian, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, and W.C. Woo, "Convolutational lstm network: A machine learning approach for precipitation nowcasting," *NIPS*, 2015.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *NIPS*, 2015.

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, and C. Alexander, "SSD: Single shot multibox detector," *ECCV*, 2016.

[16] J. Sugano, "mitotic cell division event detection using classification of temporal feature histogram," *Visual Inspection Algorithm Competition*, 2017.

[17] "TC-IAIP AIA2017," *http://www.tc-iaip.org/index-e.shtml*.