

A 2.5D Cascaded Convolutional Neural Network with Temporal Information for Automatic Mitotic Cell Detection in 4D Microscopic Images

Titinunt Kitrungrotsakul*, Xian-Hau Han[†], Yutaro Iwamoto*, Satoko Takemoto[‡], Hideo Yokota[‡],
Sari Ipponjima[§], Tomomi Nemoto[§], Xiong Wei[¶] and Yen-Wei Chen*

*Graduate School of Information Sci. and Eng., Ritsumeikan Univ., Shiga 525-8577, Japan

[†]Faculty of Science, Yamaguchi University, Yamaguchi, Japan

[‡]Center for Advanced Photonics, RIKEN, Saitama, Japan

[§]Research Institute for Electronic Science, Hokkaido University, Hokkaido, Japan

[¶]Institute for Infocomm Research, Singapore

Abstract—In recent years, intravital skin imaging has been increasingly used in mammalian skin research to investigate cell behaviors. A fundamental step of the investigation is mitotic cell (cell division) detection. Because of the complex backgrounds (normal cells), the majority of the existing methods cause several false positives. In this paper, we proposed a 2.5D cascaded end-to-end convolutional neural network (CasDetNet) with temporal information to accurately detect automatic mitotic cell in 4D microscopic images with few training data. The CasDetNet consists of two 2.5D networks. The first one is used for detecting candidate cells with only volume information and the second one, containing temporal information, for reducing false positive and adding mitotic cells that were missed in the first step. The experimental results show that our CasDetNet can achieve higher precision and recall compared to other state-of-the-art methods.

Keywords—mitotic cell detection; cascaded convolutional neural network; end-to-end training; 4D microscopic images; 2.5D Fast R-CNN

I. INTRODUCTION

Division of the cell in adult mammalian epidermis is important for maintaining the epidermal structure as these cells are important for replenishing eliminated keratinocytes [1]. Cancer, atopic dermatitis, ichthyosis vulgaris, and skin diseases disrupt the balance between the proliferation and elimination of keratinocytes and create abnormal skin structures [2], [3], [4]. Though detecting the mitotic cell (cell division) is essential in investigating cell behaviors, the majority of the methods and experiments were performed with 2D dynamic images that may overlook the important information can result in wrong detection. 3D live cell dynamic images (4D images) can be obtained by using a two-photon microscopy [1]. A typical slice image of an observed 3D dynamic image is shown in Fig.1, with blue bounding boxes indicating the mitotic cells (cell division). Automatic detection of mitotic cells from such 3D dynamic images (4D images) is a challenging task. Recently, deep learning architecture has demonstrated the powerful ability of computer vision tasks by automatically learning hierarchies of relevant features directly from the input data. The deep convolutional neural network has been successfully

applied for image classification and object detection, especially for ImageNet classification competition, which has been the most successful network for image classification since 2012 [5]. Moreover, Fast Region-based Convolutional Networks (Fast R-CNN) for object detection and Single Shot MultiBox Detector (SSD) are powerful methods, both of which have outperformed several other methods, that use CNN as base network to perform object detection [6], [7]. However, these methods are designed for 2D natural image detection. In the field of mitotic cell detection, various methods have been proposed, most of which are based on image binarization [8] or segmentation of cells [9]. Though those methods is that they do not require training dataset to train the model, they require proper alignment between each slice or time sequence to obtain good results, which is time-consuming. Anat et al. [10] used a deep learning method called pixel-wised method to improve detection accuracy and accelerate the computation time. This method is based on 2D patch classification using a simple CNN network and takes considerable computation time. Though we can apply Fast R-CNN and SSD, which are widely used for object detection in natural images, they will cause several false positives because the object (mitotic cell) is similar to the background image (normal cells), as shown in Fig.1. In this paper, we proposed a 2.5D cascaded end-to-end convolutional neural network (CasDetNet) with temporal information for accurate automatic detection of 4D (x, y, z, t) mitotic cell division events in epidermal basal cells with few training data. The CasDetNet consists of two 2.5D networks. The first is used for detecting candidate cells with only volume information and the second one, with temporal information, is used for reducing false positives (normal cells) and adding mitotic cells that are missing in the first step. We also intend to use a 2.5D CNN as a base network. Compared to conventional 2D CNN, our 2.5D CNN (2D image with neighbor slices) can include more information for detection (the first step) and reduction of false positives (the second step). Though the 3D CNN can include more information than 2D and 2.5D CNN, it can use limited number of training samples

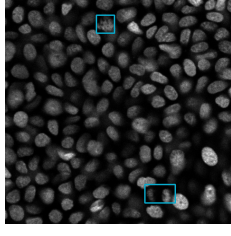


Fig. 1. One typical slice image of an observed 3D live cell dynamic image(4D image) and mitotic cells (cell division) are indicated by blue bounding boxes.

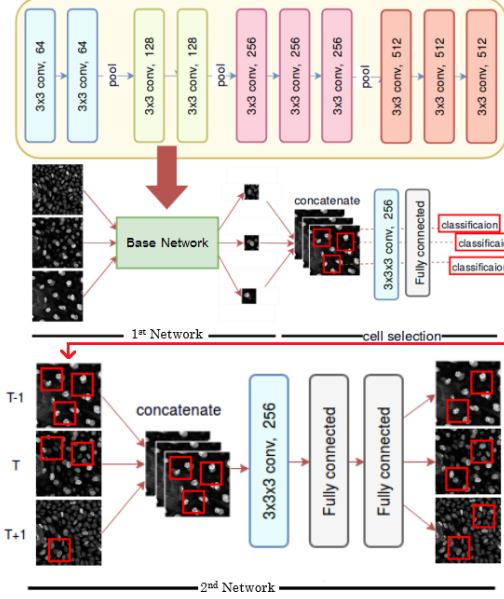


Fig. 2. Overview of our proposed CasDetNet.

(3D images) and thus cause overfitting. Results show that CasDetNet can deliver higher precision and recall comparing to other advanced methods.

The paper is organized as follows. Section II introduces the proposed CasDetNet for mitotic cell detection method is introduced in section II. Section III describes the experimental results. Finally, Section IV presents the conclusion.

II. THE PROPOSED NETWORK

The proposed CasDetNet for detection of mitotic cells is shown in Fig.2. It comprises two 2.5D networks. The first network is used to detect candidate cells using only volume information and the second, which contains temporal information, is used to reduce false positives and to add mitotic cells that were missing in the first step. The second network is cascaded to the first network and the two networks are then trained simultaneously (end-to-end training). The details regarding the first and second networks will be described in subsections II-A and II-B, respectively.

A. The first network for detection of candidate cells using volume information

The first network for detecting candidate cells is motivated by Fast R-CNN to determine the local features for establishing the region of interest (ROI). The goal is to cause the networks hidden layers to detect candidate of mitotic cells. The original Fast R-CNN requires 2D image as input and produces a set of ROI as detection results. The number of training set and network architectures determine the quality of detection result. Further, the conventional Fast R-CNNs drawback is that it loses 3D spatial information, which is important for accurate mitotic cell detection. Though we can extend the conventional Fast R-CNN to a 3D version for 3D volume images, the number of training samples will be considerably limited and result in over-fitting. Thus, we propose a 2.5D Fast R-CNN for our first detection network. As shown in Fig.1, three slice images $\{s_{-1}, s, s_{+1}\}$ are used as input to detect the candidate cells in the target slice image $\{s\}$, which is called 2.5D network. The outputs (ROIs) are indicated as $\{o_1, o_2, o_3, \dots, \text{andsoon}\}$. The advantage of our 2.5D network is that we can use neighbor slice information (2.5D information) to distinguish between the mitotic cell and normal cells, which is important for detecting mitotic cells divided along z-axis.

Figure 2 (upper part) illustrates our first 2.5D network. For our base network, we use the VGG network architecture. To enhance the accuracy of the network, we use transfer learning from ImageNet data to VGG. Each slice is processed individually and the processed slices are concatenated to form a 3D volume. We replace 3D convolutional layer with $3 \times 3 \times 3$ kernel size, followed by a ReLU non-linearity layer instead of 2D convolutional layer of original Fast R-CNN to obtain network results o_i . In each output o_i , the network will use nearby output o_{-1} and o_{+1} to generate the concatenated output. It will also be used in cell selection process to generate volume selected output O_i^t . Thus, the network will generate a set of output O_i^t consisting of 3 outputs $\{o_{i1}, o_{i2}, o_{i3}\}^t$ for each image slice s_i . Further, t indicates the time sequence in 4D data. For each set of output O_i^t , we calculate the mean to obtain first volume output V_i^t , as shown in Fig.3.

B. The second Network for reduction of false positive

We propose using the second network to reduce the false positives generated by the first network. Results from the first network V_i^t contain both correct and incorrect detection results. In this section, we used the second network to refine the results (reduce the false positives) by using temporal $\{time_1, time_2, time_3, \dots, time_n\}$ information.

There are several methods to manage extra dimensional information (temporal information). Taking mean or thresholding from image sequence is a common method to smoothen the image sequence and removing or adding over-/under-detection. We concatenate the volume output V_i^t at time t , previous output V_i^{t-1} , and next output V_i^{t+1} time sequence together and then apply the second CNN classification (for reducing false positives), as shown in Fig.2 (lower part). The network will generate the time set of output consisting of three outputs of

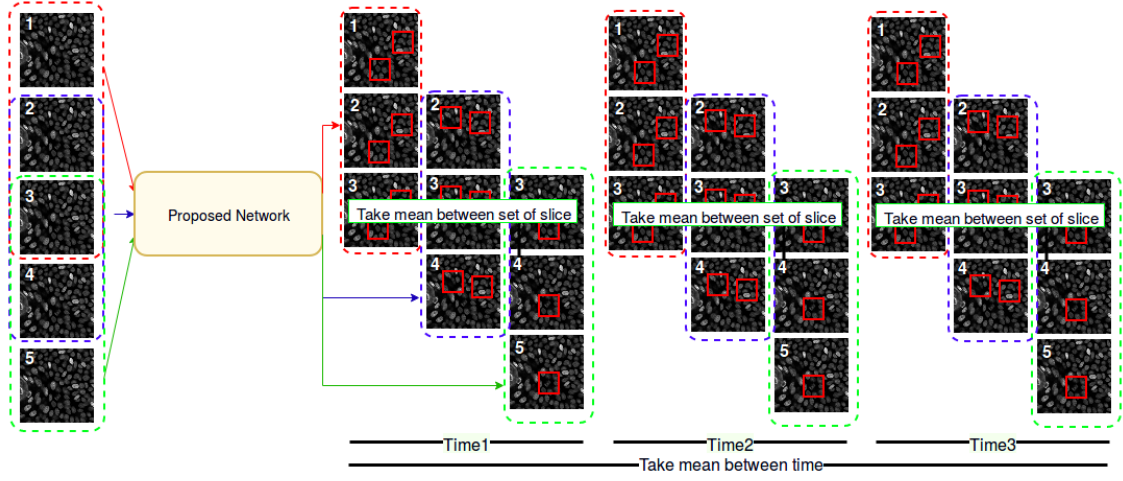


Fig. 3. Taking mean result to eliminate normal cell in detected result from both volume and time information detection network.

TABLE I
DETECTION PERFORMANCE OF OUR PROPOSED CASDETNET ON 2D SLICE IMAGE.

Data	true positive	false positive	ground truth
1	662	183	711
2	628	756	745
3	1296	216	1717
4	1215	895	1576
5	183	229	1563

temporal frames $\{time_i^{t-1}, time_i^t, time_i^{t+1}\}$. The final result F_i^t is obtained by taking the mean of three frames, as shown in Fig.3.

III. EXPERIMENTAL RESULTS

To validate the effectiveness of our proposed method, we perform experiments on 4D (temporal 3D volume sequence) data from JSPE, Technical committee on Industrial Application of Image Processing Appearance inspection algorithm contest 2017 (TC-IAIP AIA2017) [11]. There are five datasets, each containing approximately 80 temporal frames. The data size is approximately $480 \times 480 \times 37$. Each data contains 13 mitotic cells, as listed in Table III (ground truth). Data augmentation is added in the training phase to increase the number of training set so that overfitting that normally occurs in small datasets can be avoided and the model can be induced to learn to detect the mitotic cells that will generally be under-detected in the 2D network. Cropping, rotation, translation, mirror imaging, noising, and resizing methods are used in our study. The parameters for cropping, rotation, translation, noising, and resizing are randomly selected. We determine the parameter for each augmentation method as follows: 224×224 cropping size with random location; random rotation angle in the range of 0180; random percentage of Gaussian noise in the range of 1%-3%; random resizing scale in the range of 0.91.1. Using data augmentation methods helps to generate varied combination images to train the model.

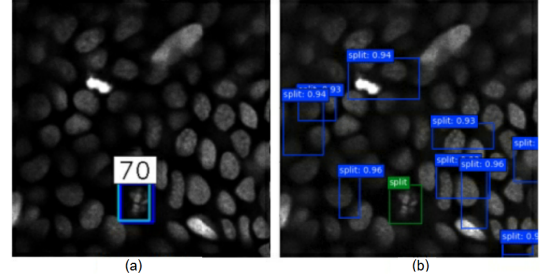


Fig. 4. Typical detection results on 2D slice image by our proposed CasDetNet (a) and SSD (b).

TABLE II
QUANTATIVE COMPARISON OF OUR PROPOSED CASDETNET WITH THE STATE-OF-THE-ART METHODS ON 2D SLICE IMAGES.

Method	precision	recall	time(sec)
2D FAST R-CNN [6]	0.0870	0.9310	239.409
3D FAST R-CNN	0.0592	0.4143	1989.012
SSD [7]	0.0411	0.7221	102.551
Our first network	0.3591	0.7532	253.005
CasDetNet	0.7228	0.70358	329.771

In our experiments, we use leave-one-out method. Further, for training our model, we use Adam optimization method. As described in the previous section, two networks are cascaded and trained simultaneously (end-to-end). The learning rate for Adam in our network starts with 0.5×10^{-5} and changes into 0.5×10^{-6} after finishing the 10k batch, with each batch containing five image slices.

A. Detection results on 2D slice images

First, we present detection results on 2D slice images. Each slice image is considered as a sample. The total number of mitotic cells (2D slice images) is shown in Table I as ground truth, and precision and recall are used as quantitative measures. For evaluation, we compare the precision and recall of our method with SSD [7], FAST R-CNN [6], and 3D convolution FAST R-CNN, which is a modified version of

TABLE III
DETECTION RESULTS ON 4D IMAGES.

Data	Sugano[12]			Our method			ground truth
	TP	FN	FP	TP	FN	FP	
1	1	0	0	1	0	0	1
2	1	0	3	1	0	0	1
3	2	0	0	2	0	0	2
4	3	0	0	3	0	0	3
5	2	1	0	1	2	0	3

the original FAST R-CNN. All methods are calibrated from ImageNet except 3D FAST R-CNN. The detection results for 2D slice images using CasDetNet are shown in Table I. The number of true positive ROI of all data is largely the same as the number of ground truth ROI, except for Data No. 5 that cannot be detected properly as its mitotic cells were difficult to detect because they occur at the edge of the image. The detection results for 2D slice images obtained using our proposed method and SSD are shown in Fig.4. It is evident that our method can detect mitotic cell correctly. On the other hand, several false positives are detected by SSD (Fig.4(b)). Compared to the SSD result (Fig.4(b)), our proposed method (Fig.4(a)) can significantly reduce false positives. The quantitative comparisons are shown in Table II. Though both 2D FAST R-CNN and SSD present high recall, they also present low precision because of several false positives being detected. Both precision and recall for 3D FAST R-CNN are lower because of overfitting. The 3D FAST R-CNN also has high computation cost. If we only use the first 2.5D network, we can improve the precision compared to 2D FAST R-CNN and SSD because of the 2.5 D network. However, it still contains a large number of false positives. We can also significantly reduce these false positives by using the second network with temporal information. It should be noted that we do not compare our method with Anats method [10] because it is a pixel-wise method and takes more time than 3D FAST R-CNN in both training and testing.

B. Detection results on 4D data

Our aim is to detect mitotic cells on 4D data. We combine our detection results on 2D slice image, as described in the previous sub-section, for final results and compare our results with the winner of the TC-IAIP AIA2017 contest [11]. The detection results (TP, FN, FP) regarding 4D data are summarized in Table III. Except Data No. 5, perfect detection is achieved without any FP and FN. For Data No. 5, two mitotic cells are not detected, the reason for which has been described in the previous sub-section. Sugano method [12],

the winner in TC-IAIP AIA2017 contest, can also properly detect mitotic cells. However, there are 3 FP for Data No. 2.

IV. CONCLUSION

We have proposed a 2.5D cascaded convolutional neural network for automatic detection of mitotic cells in 4D image (x, y, z, and time). The proposed network consists of two networks, the first of which is a modified 2.5D Fast R-CNN for detecting candidate cells and the second is used for reducing false positives using temporal information. The results demonstrated that our proposed method is more accurate than the other established methods such as Fast R-CNN, SSD, and the TC-IAIP AIA2017 contest winners method.

ACKNOWLEDGEMENT

This work is supported in part by Japan Society for Promotion of Science (JSPS) under Grant No. 16J09596 and KAKEN under the Grant No. 18H04747, 16H01436, 15H05954, 15H05953 and also partially supported by A*STAR Research Attachment Programme.

REFERENCES

- [1] Sari Ipponjima Terumasa Hibi and Tomomi Nemoto. Three-dimensional analysis of cell division orientation in epidermal basal layer using intravital two-photon microscopy. In *PLOS one*, 2016.
- [2] Hsu YC, Li L, and Fuchs E. Emerging interactions between skin stem cells and their niches. *Nat Med*, 20(8):847–856, 2014.
- [3] Jones P and Simons BD. Epidermal homeostasis: do committed progenitors work while stem cells sleep? *Nat Rev Mol Cell Biol*, 9(1):82–88, 2008.
- [4] Watt FM. Mammalian skin cell biology: at the interface between laboratory and clinic. *Science*, 346(6212):937–940, 2014.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [6] Ross Girshick. Fast r-cnn. In *International Conference on Computer Vision (ICCV)*, 2015.
- [7] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In *ECCV*, 2016.
- [8] J. Sauvola and M. Pietikainen. Adaptive document image binarization. In *Pattern Recognition*, volume 33(2), pages 225–236, 2000.
- [9] Meijering Erik. Cell segmentation: 50 years down the road [life sciences]. In *Cell segmentation: 50 years down the road [life sciences]*, 2012.
- [10] Anat Shkolyar, Amit Gefen, Dafna Benayahu, and Hayit Greenspan. Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks. In *Engineering in Medicine and Biology Society (EMBC)*, pages 743–746, 2015.
- [11] TC-IAIP AIA2017. <http://www.tc-iaip.org/index-e.shtml>. Accessed: 2017-11-30.
- [12] Junichi Sugano. mitotic cell division event detection using classification of temporal feature histogram. In *ViEW 2017 visual inspection algorithm competition*, 2017.