Cell Tracking based on Multi-Frame Input Detection

Zibin Zhou, Fei Wang
Harbin Institute of Technology, Shenzhen, China

Abstract. The tracking-by-detection strategy is a common tracking method. The detector selects the candidate area of the target, and the tracker performs association, e.g. using overlap rate, between targets in different frames. In our method, states of cells can be categorized into normal states and mitotic states, and multi-frames are used as inputs of neural network. When tracking, matching between preceding and succeeding frames is performed using overlapping rate. Finally, cell regional state matrix is used to identify whether mitosis occurs, and cell lineage trees are assigned.

1 Method

1.1 Preprocessing

In this stage, four tasks are performed.

(a) After cell semantic segmentation using the U-Nnet [1], sizes between input and output images are inconsistent. To address this problem, the original image is expanded by adding a flipping version, so as to make sure the size of output image is consistent with that in the original image.

(b) Since the ground truth of the edge part of the image is ignored, the image border of the ignored part is completely black filled.

(c) To expand the train data set with limited size, horizontalor vertical flipping operations are performed.

(d) The ground truth of the cell tracking can be categorized into a normal state and a mitotic state. Here the mitotic state is defined as the $n_{mitotic}$ frames before and after mitosis. In our method, TRA ground truth for each pixel are redifined, i.e., “0” for background, “1” for normal cells, “2” for motitic cells.

1.2 Detection

Sample images which have all the three kinds of lebels are used to train, and U-Net is trained for cell detection. Since there is a frame-to-frame association in the mitotic stage, multi-frame images are used as inputs, which enables the network to more robust to identify mitotic cells. When perform U-Nnet test, the multi-frame input images are also adopted. Here, the current frame is associated to the previous n_{input} frames as inputs. The experiment results show that multi-frame input method has higher accuracy on mitotic cell detection, compared with single-frame input method.

In our experiments, the loss weight $weight_{loss}$ is defined, and weights on thress categories (background, normal cells, and mitotic cells) can be adjusted. We find that, increasing the proportion of cell loss weight of the mitotic cells can make the network learn and detect better.

1.3 Tracking

Here, the preceding and succeeding frames associations are performed according to the overlap IOU (intersection-over-union) [2]. According to the mass centroid of the detected cell, a predicted

rectangular target region of size m_{size}*m_{size} is created, and the relationships of the candidate cells are matched according to the overlap IOU of the rectangular target regions between the preceding and succeeding frames.

If tracking is lost, i.e., there is no matching detection in the current frame, the position of the previous frame is used as a replacement of the position in current frame.

If there is still no matching detection in the after max_{miss} frames, the tracking is terminated.

1.4 Postprocessing

Combining the detection results of the n_{detect} frames before and after the current frame, a cell regional state matrix is created for the starting point of each trajectory, and in the matrix, each element is marked as 0, 1 or 2, correspondig to background, normal cell and mitotic cell.

When the number of '2' is greater than a threshold $min_{mitotic}$, a mitosis is thought to occurs at the starting point of the trajectory. Then we assign lineage trees, and mark tracking trajectory appropitely.

For the trajectory whose length is smaller than min_{track}, it is deleted.

2 Parameters

In the data set "PhC-C2DL-PSC", the parameter configuration is as follows:
$n_{mitotic} = 3$, $n_{input} = 2$, $weight_{loss} = [0.2, 0.3, 0.5]$,
$m_{size} = 11$, $max_{miss} = 3$, $n_{detect} = 3$, $min_{mitotic} = 3$,
$min_{track} = 3$.

References

[1] Ronneberger, Olaf, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." (2015).

[2] Bochinski, Erik, V. Eiselein, and T. Sikora. "High-Speed tracking-by-detection without using image information." IEEE International Conference on Advanced Video & Signal Based Surveillance IEEE, 2017.