USYD-AU

Authors: Yuqian Chen, Chaoyi Zhang

Email: yche7883@uni.sydney.edu.au

Platform: Linux

Prerequisites: Python

USYD-AU: SUMMARY

In our method, we employ Mask R-CNN [1] as our backbone network for conventional cell detection and segmentation. In addition to the three original branches for predicting classification scores, bounding boxes and segmentation masks, we design a fourth branch to perform cell tracking across frames by comparing similarities [2] between cell instances at adjacent frames. Both spatial and visual features are fed into the tracking branch for better tracking performance.

USYD-AU: PREPROCESSING

Random flip is performed with a ratio of 0.5 on input images to expand the dataset. The images are normalized before put into the neural network.

USYD-AU: SEGMENTATION

Cell detection and segmentation is performed by Mask R-CNN pipeline [1], which is a two-stage instance segmentation algorithm. Instances are first detected and then segmentation is performed on each detected instance. First of all, feature maps are extracted with backbone networks such as resnet-50 and Feature Pyramid Network (FPN). Then in the first stage, Region Proposal Network (RPN) is adopted to propose candidate boxes and crop visual features of each proposal. In the second stage, the visual features are put into classification head, detection head and segmentation head to predict classifications scores, box locations and segmentation masks. Mask R-CNN has demonstrated state-of-the-art detection and segmentation performance. Please refer to [1] for more details.

USYD-AU: TRACKING

First of all, we extract both visual and spatial features for the tracking task. Visual features are extracted in the same way as Mask R-CNN, which are insufficient for cell tracking task. Therefore, we also incorporate spatial information into the network for tracking task. The spatial feature vector of each cell instance is presented as coordinate values of cell instances on images. They are normalized before put into the tracking head. In the first stage, we have obtained features of candidate proposals in frame *t* and *t-1*. In the second stage, we add a tracking head to the network in parallel to the classification, box regression and mask segmentation branches. Then the previously obtained features are fed into the tracking head, which is adopted to calculate the similarity score [2] of two inputs. The tracking head is composed of two fully-connected layers. For every candidate in frame *t*, we calculate its similarity score with candidates in the frame *t-1* respectively. In the training process. If the candidate pair are instances of the same cell, the label is set to 1, 0 the otherwise. After feeding features into the tracking head, similarity scores within [0, 1] are obtained and are used to calculate loss with a cross entropy function. Tacking head is trained in a way that the same instances have scores close to 1 while different ones close to 0. Our network is trained in an end-to-end fashion with a total loss including classification loss, bounding box loss, segmentation loss as well as tracking loss. For inference, each frame in the testing images are processed in sequence with our proposed pipeline. After generating set of instance hypothesis, Non-Max Suppression (NMS) is performed to eliminate overlapping boxes and obtain the final instance detections. Both of mask segmentation and instance tracking are performed on the detected instances. For tracking task, all instances of the first frame are regarded as new instances of the video and assigned with different tracking ids. To identify the same instance across frames, instances of the current frame are matched to those of the last frame by calculating their similarity scores. The tracking id of the cell instance with the largest similarity score in frame *t-1* is assigned to the target cell at the current frame. In the situation where two target cells match to one cell at the previous frame, we regard it as cell mitosis. With this strategy, our method is able to match cell instances across frames and also detect cell mitosis. After processing all frames, our method is able to generate a set of instance hypothesis and obtain bounding box, classification score, segmentation mask and tracking id for each instance.

USYD-AU: POST-PROCESSING

No post-processing is carried out after tracking.

REFERENCES

1. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 2961-2969 (2017).

2. Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 539-546 (2005).