THU-Hu-CN

Authors: Tao Hu, Xiaowo Wang

Email: hut16@mails.tsinghua.edu.cn

Platform: Linux

Prerequisites: Pytorch

THU-HU-CN: SUMMARY

Our approach for multi-cell segmentation and tracking is based on a deep convolutional neural network for instance segmentation and a optimization cell association using Hungarian algorithm for tracking. The network, which takes the raw images as input and provides the final segmentation masks for each cell as output, consist of spatial-aware temporal feature extraction network and instance mask generation network. By this design, our network is able to assemble appearances and motion contexts of various scales in a time period, resulting in better segmentation performance compared to a single static image. The network can be trained end-to-end from the annotated images. More detailed explanations can be found in related papers [1, 2].

THU-Hu-CN: SEGMENTATION

Our approach proposes a polar coordinates based instance segmentation method which aggregates historical features in a spatial-aligned and scale-aware paradigm. We present a novel spatial-aware temporal feature aggregation network for feature extraction to utilize temporal video information. The intuition is to find out a pixel-level alignment among historical frames and explore an elegant fusion of multi-resolution feature maps. In addition, we also use a feature pyramid structure to handle motion estimation at different scales. The features of different layers from historical frames are aligned to their corresponding feature layers of the template frame through deformable convolution network, and then aggregated by another vanilla convolution network. The features from different layers are combined gradually top-down to produce hierarchical features for various cell size. Our network architecture is able to balance the need of segmentation for shallow layer features and the need of translation invariance for deep semantic features aroused by deformable offset learning. Referring previous work [2], the instance segmentation layer is used to predict cell markers and masks, which takes the feature pyramid of input image and predicts the distance from a sampled positive location (candidates of the instance center) to the instance contour at each angle, and after assembling, outputs the final mask.

THU-Hu-CN: TRACKING

Our method follows the online tracking-by-detection paradigm, which generates the trajectories by associating detection results across frames. For cell label propagation, we conduct data-association with Hungarian algorithm, which optimizes the total overlap (measured as intersection over union of segmentation mask) between the cells in frame t and frame t+1.

Tracking process follows rules as below:

- The tracklet for each target cell starts from the detection that it first appears, and will be extended by appending new detection in the next frame.
- If more than one cells are covering at least 15% of known tracklet in the previous frame, then all of them are daughters.
- Each tracklets without daughters cells corresponds to one detection.
- If a detection receives no label, a new label is assigned.
- This process is repeated recurrently until all trajectories are constructed completely.

REFERENCES

1. Hu T, Huang L, Liu X, Shen H. Real time visual tracking using spatial-aware temporal aggregation network. arXiv:1908.00692 (2019).
2. Xie E, Sun P, Song X, Wang W, Liang D, Shen C, Luo P. PolarMask: Single shot instance segmentation with polar representation. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, (2020).