KIT-Loe-GE

Authors: Katharina Löffler

Email: katharina.loeffler@kit.edu

Platform: Linux (tested on Ubuntu 16.04 and 18.04)

Prerequisites: ≥ 16 GiB RAM, ≥ 12 GiB VRAM (CUDA = 10.2)

KIT-Loe-GE: SUMMARY

Based on the instance segmentation approaches of Neven *et al.* [1] and Lalit *et al.* [2] we propose a deep learning approach that segments and tracks cells through learning to predict offsets of pixels belonging to a cell to its corresponding cell centers at t and $t-1$ respectively [3]. As model a (branched) ERFNet [4] with one shared encoder and three decoder paths is used - two decoders for segmentation and one for tracking. The model receives pairs of successive image frames t and $t-1$ and predicts for each frame the segmentation as well as the offsets of the pixels belonging to cells at t to the center of its corresponding predecessor cell in $t-1$. The code repository is available at https://git.scc.kit.edu/kit-loe-ge/embedtrack.

KIT-Loe-GE: PREPROCESSING

We generate image crops of size *crop_size* and min-max normalize each image crop to range [0, 1], where the minimum and maximum are set to the percentiles 1 and 99 respectively.

KIT-Loe-GE: SEGMENTATION

In the branched ERFNet two decoders are trained for the segmentation task. The first decoder learns to predict for each pixel its offset to its corresponding cell center at time point t and a cluster bandwidth which is related to the estimated cell size. The cluster bandwidth allows for a less precise estimation of the cell center of large cells compared to smaller cells. The first decoder predicts four maps - two for segmentation offsets (x and y direction) and two for the clustering bandwidth (x and y direction). The second decoder learns to predict which cell pixels estimate their offsets to their cell center well and regresses to 0 for background - which serves as a foreground-background prediction.

For training the model, we use a similar loss function as in [1] and extend the loss by an additional tracking loss term, which enforces the predicted tracking offsets of pixels belonging to cells at t shift the pixels to their corresponding cell center in the previous frame $t-1$. We use for all data sets the available silver truth segmentation and combine it with the gold truth tracking annotation to generate reasonably labeled, fully annotated training data. For **Fluo-N2DH-SIM+** we use the fully annotated ground truth. We

split the annotated data sets into a train and validation data set where we keep the last 10% of the sequences for validation. For the train and validation data set we generate overlapping crops of patch size *crop_size*, and remove pairs of crops from the data sets if both image crops show no cells at all. During training we use flipping, rotation, contrast limited adaptive histogram equalization (CLAHE) and blur as image augmentations and add random shifts to the image crops to simulate larger cell movements. We use the validation data to calculate IOU and use the model with the best IOU for submission. Each model is trained for 15 epochs using Adam optimizer with learning rate $5 \cdot 10{-}4$ and a one cycle learning rate scheduler.

The predicted maps from the two decoders are converted into an instance segmentation by applying a clustering step. For the clustering, based on the predicted foreground-background map, cell pixels with a high score, which corresponds to a good prediction of their cell center, are selected and their estimated offsets are added to their pixel positions. The resulting map is then thresholded to find potential cell centers which are pixel positions with many clustered pixels in their neighborhood. The found centers are then sorted based on their initial score in the foreground-background map. Starting with the cell center with the highest score, pixels are iteratively assigned the found cell centers if their distance is small enough and the number of assigned pixels is larger than a minimum cell size. The minimum cell size, *min_cell_size*, is set to one half of the 1% percentile of the cell sizes of the train data split.

For inference, we generate overlapping crops of size *crop_size* from the images. As test time augmentation we flip and rotate each image crop and calculate the mean prediction over the augmented crop. The predicted patches are stitched to a prediction over the full image, then the clustering step is applied.

KIT-Loe-GE: TRACKING

The third decoder in the ERFNet learns, based on the forwarded pairs of image frames t and $t - 1$, to predict for each cell pixel its offset to its corresponding cell center in the previous frame $t - 1$. The decoder predicts two offset maps (x and y direction). Cell tracking is applied after predicting the segmentation masks by processing the segmentation masks and the predicted tracking offsets backwards in time. For each segmented cell at t its estimated offset is extracted from the offset maps and added to the cell pixel positions at t. The shifted pixel positions estimate the corresponding cell center at $t - 1$. Each segmentation mask at t is linked to the segmentation mask at $t - 1$ which contains the most shifted pixels. In case that the shifted pixels of two segmentation masks from t lay in the same segmentation mask at $t - 1$, the segmentation mask at $t - 1$ is marked as their predecessor. In all other cases – shifted pixels of more than two segmentation masks of $t - 1$ lay in a segmentation mask at t or

the shifted pixels lay in no segmentation mask at *t* at all – the segmentation mask from *t* is marked as a new track starting in *t* with predecessor 0.

REFERENCES

1. Neven D, De Brabandere B, Proesmans M, Van Gool L. Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth. In *Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, 8837-8845 (2019).
2. Lalit M, Tomancak P, Jug F. Embedding-based instance segmentation in microscopy. In *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*. 399-415 (2021).
3. Löffler K, Mikut R. EmbedTrack – Simultaneous cell segmentation and tracking through learning offsets and clustering bandwidths. arXiv: 2204.10713, 2022.
4. Romera E, Álvarez JM, Bergasa LM, Arroyo R. ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems* **19**, 263-272 (2018).