



A Benchmark for Epithelial Cell Tracking

Jan Funke¹, Lisa Mais², Andrew Champion¹, Natalie Dye³,
and Dagmar Kainmueller²(✉)

¹ HHMI Janelia Research Campus, Ashburn, USA

² BIH/MDC, Berlin, Germany

`dagmar.kainmueller@mdc-berlin.de`

³ MPI-CBG, Dresden, Germany

Abstract. Segmentation and tracking of epithelial cells in light microscopy (LM) movies of developing tissue is an abundant task in cell- and developmental biology. Epithelial cells are densely packed cells that form a honeycomb-like grid. This dense packing distinguishes membrane-stained epithelial cells from the types of objects recent cell tracking benchmarks have focused on, like cell nuclei and freely moving individual cells. While semi-automated tools for segmentation and tracking of epithelial cells are available to biologists, common tools rely on classical watershed based segmentation and engineered tracking heuristics, and entail a tedious phase of manual curation. However, a different kind of densely packed cell imagery has become a focus of recent computer vision research, namely electron microscopy (EM) images of neurons. In this work we explore the benefits of two recent neuron EM segmentation methods for epithelial cell tracking in light microscopy. In particular we adapt two different deep learning approaches for neuron segmentation, namely Flood Filling Networks and MALA, to epithelial cell tracking. We benchmark these on a dataset of eight movies with up to 200 frames. We compare to Moral Lineage Tracing, a combinatorial optimization approach that recently claimed state of the art results for epithelial cell tracking. Furthermore, we compare to Tissue Analyzer, an off-the-shelf tool used by Biologists that serves as our baseline.

1 Introduction

Epithelia are thin tissues that cover body structures like organs and cavities. They are commonly composed of a single layer of cells tightly connected by junctions. In a developing organism, cells in epithelia are very dynamic. They move greatly, to the extent that they change neighbors over time despite the tight connections. They undergo cell divisions, cell death, and a large variety of shape changes.

Biologists are interested in a mechanistic understanding of the principles that govern epithelial development on the level of individual cells. A powerful approach towards this goal are high throughput studies, where membrane-stained

J. Funke and L. Mais—Shared first authors.

© Springer Nature Switzerland AG 2019

L. Leal-Taixé and S. Roth (Eds.): ECCV 2018 Workshops, LNCS 11134, pp. 437–445, 2019.

https://doi.org/10.1007/978-3-030-11024-6_33

developing epithelia are imaged over time with fluorescence light microscopy (LM), and cells are segmented and tracked in the resulting movies to allow for quantitative analysis.

Current computational tools that Biologists use to segment and track epithelial cells rely on non-learned methods watershed or thresholding for 2d segmentation, combined with heuristics for tracking cells over time [1–3]. So far the computer vision community has shown little interest in this problem despite its abundant importance for Cell Biologists. Recent cell tracking challenges [4, 5] have excluded epithelial cells from their benchmarks, and to our knowledge, only one recent approach has tackled epithelial cell tracking with modern methodology, namely the Moral Lineage Tracing (MLT) approach and its fast variants [6–8].

However, the related image analysis problem of segmenting neurons in 3d electron microscopy (EM) images has recently become a subject of study for a number of computer vision groups. Neurons are tree-structured densely packed cells, and the membranes that separate them are so thin that they are prone to fall prey to partial volume effects. In essence, it cannot be assumed that even a single pixel of membrane separates two neurons in EM imagery. These properties hold analogously for 2d+t epithelial cell tracks in light microscopy data.

There are however two key structural difference between neurons in 3d EM data and epithelial cell tracks in 2d+t LM data: First, when a neuron branches, the two emerging branches diverge immediately, making it unlikely to share a boundary later on. Conversely, when an epithelial cell divides, the daughter cells may reside next to each other for a long time. Consequently, in general, methods for 3d neuron EM segmentation are not directly applicable to cell tracking. Second, the location of the root of a tree-shaped neuron is not known a priori, whereas the root of a cell lineage tree is known to lie in the first frame of a movie. This known direction of object branchings can be leveraged for cell tracking.

In this work, we propose extensions of two state of the art deep learning based 3d neuron segmentation methods to make them applicable to 2d+t tracking of densely packed objects, namely MALA [9] and Flood Filling Networks (FFN) [10]. The core contribution of our work is a benchmark on a dataset of eight epithelial cell movies. We benchmark against a current state of the art tool used by Biologists, namely the Tissue Analyzer [1] (formerly known as Packing Analyzer [11]). Furthermore we benchmark against the current state of the art for epithelia tracking in the computer vision literature, namely the MLT approach [6–8]. Here, to allow for a fair comparison of MLT to our deep learning based methods, we have “modernized” MLT to be based on a U-Net [12] as opposed to a Random Forests [13] as employed in the original MLT. In particular, we use cell candidates and tracking scores as predicted by the MALA approach. This allows us to investigate the difference between a combinatorial optimization method (MLT) compared to greedy agglomeration and matching (MALA).

Our comparative evaluation of the standard segmentation and tracking error metrics SEG and TRA [5] reveals that MALA, FFN and the U-Net based MLT significantly outperform Tissue Analyzer. Differences among the three deep

learning based methods are more subtle, yet their run-times differ significantly. Our results suggest that MALA, which is the fastest approach, performs en par with MLT and slightly better than FFN in terms of accuracy. We therefore suggest MALA as our method of choice to be employed for epithelial cell tracking by biologists. Our code for both MALA and FFN is freely available on github.¹

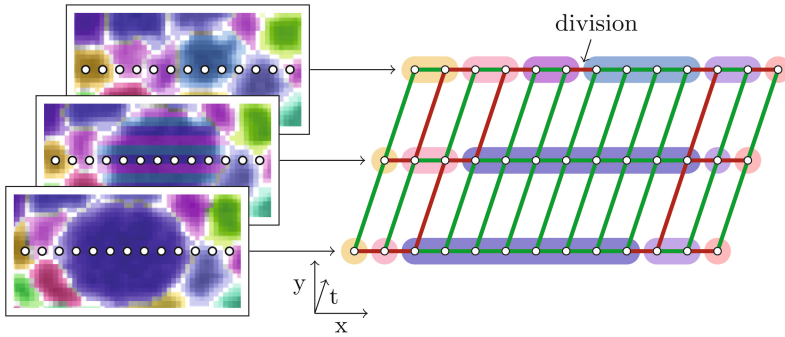


Fig. 1. The MALA method for cell tracking. A lineage (left, cell divisions are indicated by horizontally striped labels in the frame preceding the division) is represented by labelling affinity edges (right) as “cut” (red) or “connect” (green) between neighboring pixels within and across frames. Given that cells in subsequent frames overlap, this formulation allows to segment (in 2d) and track cells (over time). Shown on a coarsened pixel grid for illustration purposes. (Color figure online)

2 Methods

MALA for Cell Tracking. MALIS plus agglomeration (MALA [9]) is a method that has recently been proposed to address the problem of segmenting neurons in EM volumes: There, a 3d U-Net is trained to predict affinities on edges between voxels using a variant of the MALIS loss [14], such that edge affinities are high if the incident voxels are part of the same object and low otherwise. The prediction of affinities is followed by a simple hierarchical agglomeration, where initial fragments obtained by a watershed are iteratively merged according to the predicted affinities between them until a given threshold.

Acknowledging the similarity between the problem addressed here and the segmentation of neurons in (anisotropic) volumes, we modified the MALA method to segment and track cells in 2d+t videos. For that, we treat a movie sequence as a 3d volume, indexed by (x, y, t) . The affinities between voxels are now spatial (in x and y) or temporal (in t). Affinities in space represent a segmentation of a single image into individual cells, whereas affinities across time encode the continuation of a lineage (see Fig. 1).

¹ github.com/anonymous/repository.

Since agglomeration on the resulting 2d+t affinity graph is not guaranteed to produce a *moral* lineage (i.e., cells can only split, not merge), we diverge from the MALA approach and create a segmentation and tracking in two steps: First, we perform a watershed segmentation and agglomeration for each frame independently to find cells (thus ignoring affinities in t). Second, we track the found cells with a simple forward heuristic: we enumerate potential links between each pair of spatially overlapping cells in consecutive frames and score them with the mean t -affinity predicted on the overlap. A lineage is then found by greedily accepting links as long as they do not result in a merge.

Flood Filling Networks for Cell Tracking. Flood Filling Networks (FFNs) for segmenting neurons in EM volumes [10] are based on a CNN for pixel-wise foreground-background segmentation. They differ from traditional CNNs in that they are applied iteratively, with previous predictions recurrently fed into the CNN as additional input. The input predictions are initialized to 0.5. To be able to cope with densely packed objects, FFNs are applied individually to each object to be segmented, starting from automatically determined seed points.

At test time, to segment a single object, foreground probability is predicted in a field of view (FoV) around a seed location. Then a set of next FoV locations is determined and stored in a queue to be processed until empty: If a prediction one step away from the seed exceeds a user-defined probability threshold, it forms the center of a subsequent FoV. Here, a step is an axis-aligned offset of user-defined length. A binary segmentation of a single object is derived by thresholding the final full foreground probability image. Figure 2 shows a few exemplary FoV steps at test time. At training time, input FoVs are first formed at seed locations. Once training yields predictions above a user-defined threshold, additional training data is formed one step, but never further, away from the seed, employing previous predictions as input.

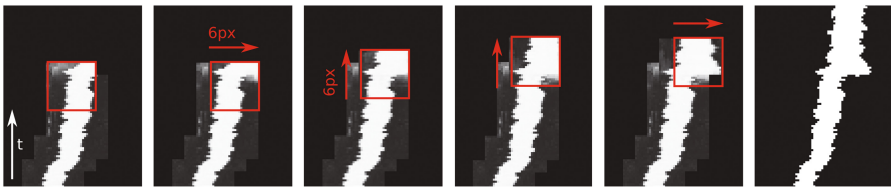


Fig. 2. FFN at test time: From left to right, a few steps of FoV movement are shown, with resulting predictions. These steps cover a cell division, visible as a sudden wider diameter of the object. The FoV can only move backwards in time, and hence does not jump into the second daughter cell that stems from this division (not depicted). Very right: Resulting binary segmentation.

FFNs are designed for pixel-wise foreground-background segmentation of individual objects. This makes them not well-suited for objects that touch themselves: For the example of a dividing epithelia cell, the mother cell as well as the two daughter cells belong to the same foreground object, where the daughter

cells often reside right next to each other, and partial volume effects often cause little to no boundary to be visible in the time dimension. Hence FFNs would not allow for distinguishing these daughter cells.

We overcome this limitation of FFNs for cell tracking applications by leveraging the knowledge that cells can divide over time but not merge. This leads us to propose the following modification: (1) We pick seeds reversely in time. (2) During training as well as testing, we allow for the FoV to only move backwards in time. This yields non-branching leaf-to-root tracks for all cells that form leafs in the lineage forest: When the FoV encounters a cell division site while filling a daughter cell, it is not allowed to move into the respective sibling cell. (3) Given the single-cell tracks resulting from (2), we detect cell divisions, i.e. pairs of single-cell tracks that stem from the same mother cell, by a simple Dice-score based overlap criterion. As with MALA, the resulting tracking graph is not guaranteed to be moral, as it may contain cell merges. Analogous to MALA, we greedily enforce morality by adding edges from our tracking graph to our final feasible solution one by one as long as morality is not violated, while rejecting edges that do lead to a violation.

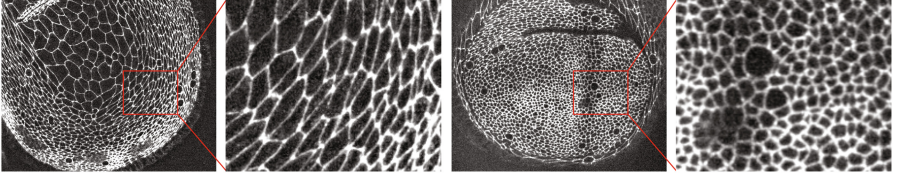


Fig. 3. Overview of the used datasets: Left: peripodial tissue with large cells in the center and thin elongated cells at the edge. Right: proper disc tissue with small roundish cells.

3 Results and Discussion

We evaluate MALA, FFN, MLT, and Tissue Analyzer (TA) on 2d+t movies of developing *Drosophila* wing epithelia. These epithelia consist of two layers, namely the proper disc layer that develops into the fly’s wing, and the peripodial layer whose biological function is a current research topic. Each layer can be captured individually as a 2d+t movie as described in [15]. Our dataset contains 8 movies: 5 proper disc and 3 peripodial layers (see Fig. 3), containing between 160 and 200 frames. We perform leave-one-out cross-validation of MALA, FFN, and MLT on all 8 movies. For each fold, a validation movie is held out from the training set to check for early stopping during training. The validation movie is chosen to be of the same class (i.e. proper disc or peripodial) as the test movie.

For MALA, we trained a 3D U-Net with a receptive field of 88×88 pixels in xy, and 28 time frames. The U-Net follows the architecture proposed in [12], consisting of four layers with downsampling factors of 2. We trained the network

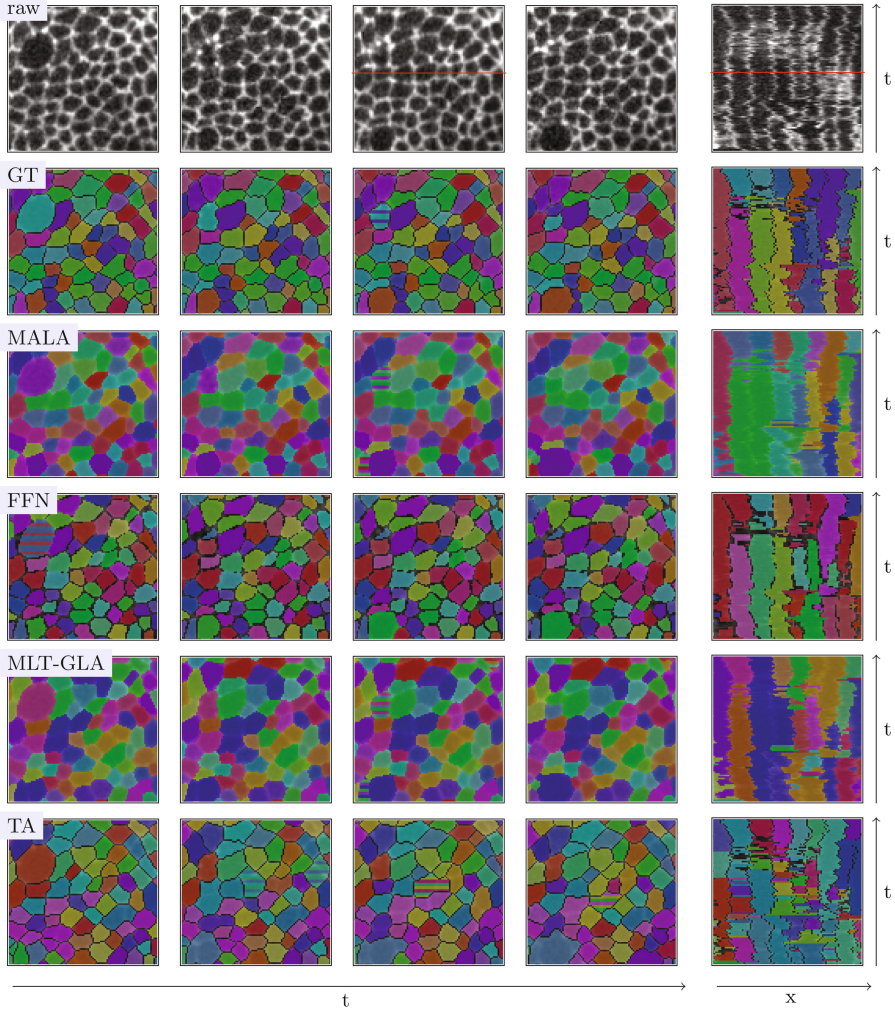


Fig. 4. Qualitative results of the investigated methods. Cell divisions are indicated by horizontally striped labels in the frame preceding the division. The red highlight in the first row indicates the same pixel line in the xy and xt views. (Color figure online)

for 150,000 iterations using the Adam optimizer [16] with an initial learning rate of $\alpha = 5 \cdot 10^{-5}$, $\beta_1 = 0.95$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.

We used the same network as for MALA to generate affinities for MLT. We employed the fast MLT solver GLA (see [7]) since neither other MLT solver terminated within 100 hours of runtime on a single movie.

For FFN, we trained a 3D Conv Net with 8 layers and 32 filters, FoV of $25 \times 25 \times 25$ pixels, step size of 6 pixels for moving the FoV, and probability threshold 0.9.

For Tissue Analyzer results on proper disc movies, we evaluate the automated tracking results that formed the basis for our ground truth data via manual curation by biologists. In this sense, our reported Tissue Analyzer results are biased towards an unfair advantage. On peripodial movies however, automated Tissue Analyzer results were deemed too inaccurate to form a suitable basis for manual curation. Instead, biologists used the interactive mode of Tissue Analyzer to generate ground truth, iteratively curating each frame individually and then using it to initialize segmentation of the next frame. For this reason we do not report automated Tissue Analyzer results for peripodial movies.

We evaluate the standard tracking error metrics SEG and TRA [4] to measure 2d segmentation and tracking accuracy, respectively (see Table 1). Both measures range from 0 to 1, where 1 indicates a perfect result. Figure 4 shows some exemplary results. The average run-times of our methods for processing a single movie are 35 min (MALA), 42 min (MLT) and 90 min (FFN).

Table 1. Quantitative results of the investigated methods. Shown are the mean and standard deviation of the SEG and TRA measure over each of the used datasets (peripodial: 3 movies, proper disc: 5 movies).

	PERIPODIAL		PROPER DISC	
	SEG	TRA	SEG	TRA
MALA	0.907 \pm 0.029	0.997 \pm 0.001	0.817 \pm 0.009	0.995 \pm 0.001
FFN	0.879 \pm 0.035	0.985 \pm 0.012	0.796 \pm 0.013	0.987 \pm 0.006
MLT-GLA	0.904 \pm 0.026	0.993 \pm 0.004	0.818 \pm 0.010	0.995 \pm 0.001
TA	—	—	0.758 \pm 0.009	0.865 \pm 0.003

Discussion. Concerning TRA scores, MALA and MLT yield almost perfect tracking performance, with scores very close to 1 on all eight individual movies. FFN yields slightly lower TRA scores than MALA and MLT. We note two technicalities that put FFN at a disadvantage w.r.t. the TRA score: (1) The ground truth is a dense tessellation of the image. While MALA and MLT produce a dense tessellation without boundary, FFN often leaves a few background pixels between objects. In case of a cell death, an FFN track of a cell might stop a few slices earlier in time than the respective ground truth track, which is counted as “missing segmentations” that contribute ten-fold to TRA compared to a superfluous segment that would be counted if the space was filled. (2) FFN maintains a queue of seeds and iteratively picks them for filling a new object if they have not been covered by a previous object. Towards the end of the seed queue, seeds can lie on the aforementioned background regions. In this case, filling will produce an object that lies mostly within a pre-existing object, and only a few pixels will be added as a new segment. This “noise” counts as false split errors in the TRA measure. We hypothesize that some more steps of simple postprocessing, like morphological growing of labels and sorting out very small components, may alleviate these disadvantages.

However, in addition to the above technical disadvantages, in some videos we did find some “catastrophic merge errors” yielded by FFN, where two neighboring cells are merged over their entire tracks. This may happen if the iterative FFN filling “leaks” into a neighboring cell early on, which cannot be amended later. In contrast, MALA and MLT first perform pure 2d segmentations and link them over time in a second step, which makes the results less prone to such large merge errors.

Concerning SEG scores, MALA and MLT are en par again, whereas FFN yields slightly lower scores, potentially again caused by the fact that it may leave a few pixels between cells blank, whereas neither ground truth nor MALA nor MLT do. All methods yield smaller scores for proper disc as compared to peripodial movies, which is simply due to the smaller size of the respective cells and the nature of the Dice score.

4 Conclusion

We have shown that FFN can be extended successfully for tracking applications. However we were unable to get FFN to outperform the cheaper MALA and MLT. MALA is fastest and most accurate method in our study and should hence make for a useful tool for biologists to segment and track epithelial cells, once integrated into a manual correction framework as e.g. provided by Tissue Analyzer. The fluorescence microscopy data analyzed in this work is of such quality that cheap greedy agglomeration of local deep learning based predictions is not outperformed by expensive combinatorial optimization as in MLT.

References

1. Aigouy, B., Umetsu, D., Eaton, S.: Segmentation and quantitative analysis of epithelial tissues. In: Dahmann, C. (ed.) *Drosophila. MMB*, vol. 1478, pp. 227–239. Springer, New York (2016). https://doi.org/10.1007/978-1-4939-6371-3_13
2. Stegmaier, J., et al.: Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Dev. Cell* **36**(2), 225–240 (2016)
3. Farrell, D.L., Weitz, O., Magnasco, M.O., Zallen, J.A.: SEGGA: a toolset for rapid automated analysis of epithelial cell polarity and dynamics. *Development* **144**(9), 1725–1734 (2017)
4. Maška, M., et al.: A benchmark for comparison of cell tracking algorithms. *Bioinformatics* **30**(11), 1609–1617 (2014)
5. Ulman, V., et al.: An objective comparison of cell-tracking algorithms. *Nat. Methods* **14**(12), 1141 (2017)
6. Jug, F., Levinkov, E., Blasse, C., Myers, E.W., Andres, B.: Moral lineage tracing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5926–5935 (2016)
7. Rempfler, M., et al.: Efficient algorithms for moral lineage tracing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4695–4704 (2017)

8. Rempfler, M., Kumar, S., Stierle, V., Paulitschke, P., Andres, B., Menze, B.H.: Cell lineage tracing in lens-free microscopy videos. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 3–11. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_1
9. Funke, J., et al.: A deep structured learning approach towards automating connectome reconstruction from 3D electron micrographs. CoRR abs/1709.02974 (2017)
10. Januszewski, M., et al.: High-precision automated reconstruction of neurons with flood-filling networks. bioRxiv, p. 200675 (2017)
11. Aigouy, B., et al.: Cell flow reorients the axis of planar polarity in the wing epithelium of drosophila. *Cell* **142**(5), 773–786 (2010)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
14. Briggman, K., Denk, W., Seung, S., Helmstaedter, M.N., Turaga, S.C.: Maximin affinity learning of image segmentation. In: Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I., Culotta, A. (eds.) *Advances in Neural Information Processing Systems 22*, pp. 1865–1873. Curran Associates, Inc. (2009)
15. Dye, N.A., et al.: Cell dynamics underlying oriented growth of the drosophila wing imaginal disc. *Development* (2017). dev-155069
16. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)