INFERRING THE LOCATION AND ORIENTATION OF CELL DIVISIONS ON TIME-LAPSE IMAGE SEQUENCES

Marc Karnat[⋆], *Mehdi Saadaoui* [⋄], *Sham Tlili* [⋄], *Romain Karpinski*[†], *Jean-François Rupprecht* [⋆]

[⋆] Aix Marseille Univ, Université de Toulon, CNRS, CPT (UMR 7332),
Turing Centre for Living systems, Marseille, France.
[†] LORIA, CNRS, Nancy, France
[⋄] Aix Marseille Univ, CNRS, IBDM (UMR 7288), Turing Centre for Living systems, Marseille, France,
Contact: romain.karpinski@loria.fr and jean-francois.rupprecht@univ-amu.fr

We propose a two-stage method to characterize cell divisions. In a first stage, the division detection problem is recast into a semantic segmentation task on image sequences. In a second stage, a local regression on individual divisions yields the orientation and distance between daughter cells. We apply our formalism to confocal image sequences of neural tube formation in chicken embryos, where divisions occur within a well-defined plane. We show that our two-stage method can be implemented using simple networks, e.g. a U-Net for the segmentation and a 4-layer CNN for the regression. Optimization of the networks was achieved through a systematic exploration of hyperparameters. In particular, we show that considering several frames as inputs significantly improves the segmentation performance. We reach a performance of 96% in the F1 measure for the detection and errors for the angle, which are within the bounds of the uncertainty of the ground-truth annotation dataset.

1. INTRODUCTION

In living tissues, dividing cells exert mechanical stresses on their environment, with the two cells resulting from a division (called daughter cells) typically pushing on their environment as they grow [1].

Here, we are interested in the chicken neural development, in which neural cells first migrate along a specific direction (called apicobasal) before dividing in the apical plane, which is orthogonal to the apico-basal direction [1]. Using confocal microscopy, we obtained image sequences of the apical plane every 3 minutes with the sirActin membrane marker, see Fig. 1. Cells first appear in the apical plane as growing disks which, upon reaching a critical size, undergo a fission event: the cell membrane pinches along a diameter, eventually dividing the cell into two.

A complete cell segmentation at each time frame can, in principle, yield the location of dividing cells, using a Bayesian algorithm lineage reconstruction algorithm [2]. However, a major weakness of this approach is its sensibility to segmentation errors [3], making it inapplicable under low-photon budget constraints. Several works considered the problem of cell division detection through U-Net-based segmentation, e.g. [4], which showed that using as input an image sequence (of $N = 7$ frames) can improve the cell detection algorithm. However, the temporal, N-frame, U-Net considered in [4] cannot be used to obtain the orientation of daughter cells. Recently, [5] combined a temporal U-Net for the cell division location with a temporal U-Net for the cell orientation inference; such a second network yields a segmentation mask, which, after post-processing, is used to extract a cell division angle.

Here, we evaluate a two-stage location/orientation strategy, similar in principle to [5], but replacing the second temporal U-Net by a dedicated CNN network that yields the angle and length of the divisions directly, i.e. without any image analysis post-processing. We systematically explored the impact of the number of frames N included as input on the performance of both algorithms.

2. METHODS

2.1. Stage 1: midpoint segmentation by a N-frame U-Net

Cell division midpoint and segmentation radius r We approach the cell division detection task as a form of semantic segmentation. We initially considered a representation of the two daughter cells, either by individual centers or by using an ellipse between the two; however, pairing daughter cells was an issue within cell division clusters. Here, we instead propose a representation of the cell division midpoint (point C in Fig. 1a), i.e. the center of the cleavage furrow. This center is

The project leading to this publication has received funding from France 2030, the French Government program managed by the French National Research Agency (ANR-16-CONV-0001) and from Excellence Initiative of Aix-Marseille University - A*MIDEX, as well as by ANR-20-CE30-0023 COVFEFE. This work was granted access to the HPC resources of IDRIS under the allocation AD010314339 made by GENCI.

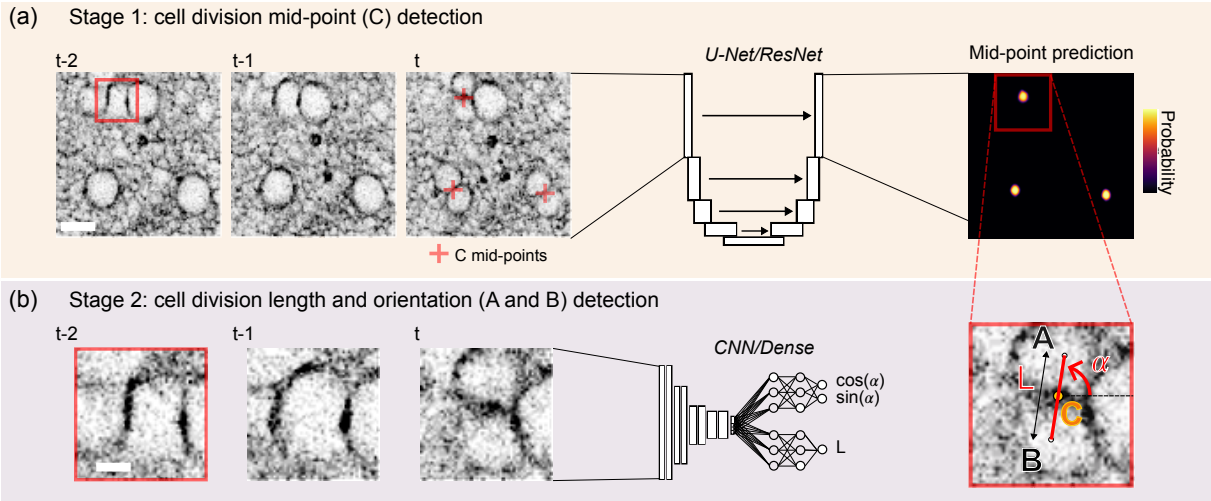

Fig. 1. Description of the proposed two-stage approach. (Top row) Stage 1: the input image sequence is processed by a U-Net to label the division centers. The image sequence is stacked at the channel dimension. Scale bar: 10μm. (Bottom row) Stage 2: around each predicted cell division center, we crop the image around the center and use a regression CNN to find the length and orientation of the division. The regression also uses stacked input images from different time steps. Scale bar: 3μm.

represented as a circular area with a radius r.

Network architecture To perform the cell detection, we choose the U-Net neural network architecture[6]. We consider a ResNet-18 as an encoder, decoded by 5 successive upsampling + conv blocks [7]. The final activation function is a sigmoid. We use a 256x256 pixels image crop as input, pre-processed through an intensity histogram equalizer. We vary from 1 to 3 the number N of consecutive frames as input. As the considered task is binary with a lot of background pixels, we assign a weight W to the midpoint division class in the following weighted binary cross-entropy (WCE) loss

$$\mathrm{WCE}(y, \hat{y}) = -Wy\ln(\hat{y}) - (1-y)\ln(1-\hat{y}), \quad (1)$$

where y for the ground truth ($y = 1$ on a division center pixel, zero otherwise) and $\hat{y}$ for the prediction. We use a standard combination of random brightness, flip, Gaussian noise, rotation, elastic distortion, and random scaling, using the Albumentation library [8], and elastic distortion of the intensity histogram [9]. We trained the network for 75 epochs, each consisting of a batch size of 32 and 250 iterations, with a learning rate of 10^{-4} with the Adam optimizer.

Post-processing We apply a threshold on the output probability map, $P > 0.5$. The centroid of the resulting connected components is the predicted midpoint position. In some instances, we detect the same midpoint division on two consecutive frames. Such redundancies could be detected and filtered by evaluating the distances between centroids on successive frames.

Evaluation metrics Sometimes, the daughter cell separation event spans several time frames, and a ± 1 frame tolerance level is justified. Therefore, to count for true positives, we implemented the following two-step matching scheme: at any time t, (1) we first count predicted centers whose distance to the ground truth set is lower than a threshold at 10 pixels (2μm, corresponding to the half the size of the smallest cells), and (2) for unmatched predicted centers, we consider the matching problem to the ground truth set at the time $t-1$ and $t+1$. Predicted centers that remain unmatched are called false positives; ground truth centers that remain unmatched are called false negatives (FN). We then estimate $F_1 = 2n_{TP}/(2n_{TP} + n_{FN} + n_{FP})$, with $n_{TP/FN/FP}$ the number of true positive, false negative and false positive, respectively.

2.2. Stage 2: orientation and length by a CNN-regression

Input processing We consider as input a crop of size 64x64 pixels that are centered on the detected cell division midpoint. Such size is chosen to encompass the two daughter cells, as the distance between daughters lies in the 20-40 pixels range.

Network architecture We use a CNN composed of 4 conv blocks. Each block comprises a convolution layer followed by a ReLU activation and a max-pooling. The first stage comprises 128 filters; this number is doubled for each subsequent block. After the 4th block, we use 2 regression heads on the flattened features: one scalar l for the length, and one vector $(\cos(2\alpha), \sin(2\alpha))$ for the orientation, which we preferred to the scalar α, as to avoid issues related to 0 and π representing the same orientation, while appearing distant in the classical MSE [10]. We used data augmentation and training protocols similar to the one described for the U-Net.

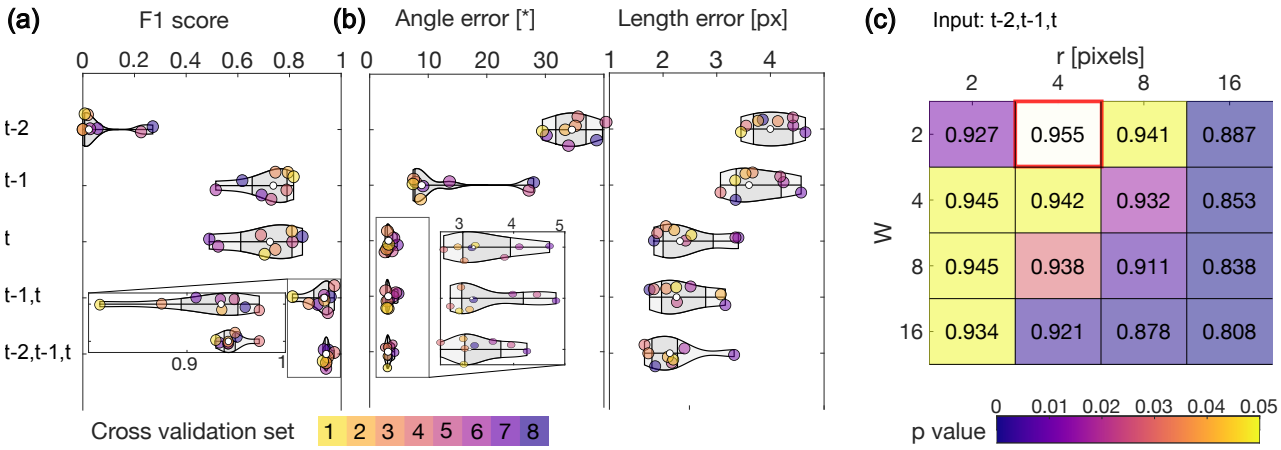

Fig. 2. (a-b) Impact of the time frames used as input on (a) the F1 score of the mid-point detection task, (b) the error on the orientation and length inference. The time t refers to the first frame at a membrane separation between the two daughter cells is apparent. Each colour refers to the set used for the validation. (c) Impact of the weight scaling W and circle radius r on the mean F1 performance of the U-Net-based division location inference with $(t-2, t-1, t)$ as input. The color code (shades of blue) refers to a p-value, estimated by a two-sample t-test, against the network's best performance at $W = 2, r = 4$px (red box). Saturation in yellow indicates that a p-value above 5% (difference with the optimal value is not statistically significant).

3. RESULTS

Our results are evaluated through a standard k-fold cross-validation method [11], in which one subset of images is chosen as the validation set, while the remaining image $k-1$ subset forms the training set. Here, we consider $k = 8$, with each of the 8 subsets originating from a different experiment. In each set, we have between 225 to 1266 divisions over 31 to 180 frames (see Supplementary Table for details).

Impact of number of frames as input We find that using the two past frames $(t-2; t-1; t)$ yields a significant enhancement of the $F1$ detection score $(0,945 \pm 0.015)$ on the midpoint detection task, as compared to all other considered time windows $(0,918 \pm 0.053$ for $t-1, t)$, see Fig. 2a. The improvement is less pronounced on the angle and length regression, i.e. as compared to $(t-1; t)$, with $(t-2; t-1; t)$ as input the reduction in the mean angle error and length error is of 0.4 degrees and 0.2 pixels, respectively, see Fig. 2b. Some information on the angle remains at $t-2$, since the average error is significantly lower than the 90 degree error baseline for randomly distributed angles.

Optimal (*r*, *W*) We perform a grid search on W (see Eq. (1)) and r to find the optimal set maximizing the F1 score of our best, N=3-frame U-Net, see Fig. 2c. We find that the couple $W = 2$ and $r = 4\,\text{px}$ is optimal; increasing W values can be compensated by lowering r.

Influence of groundtruth center vs predicted center We find that the regression performance is robust to errors in the located mid-point center. In particular, using the ground-truth or the predicted mid-point centers, the difference in the angle error amounts to $0,3$ degree, and the one on the length to $0.7\,\text{px}$ ($W = 2$ and $r = 4\,\text{px}$ model).

4. DISCUSSION

Importance of *t-2* Divisions can occur on more than 2 frames, as shown in Fig. 3a. This could explain why adding the (t-2) image as input enhances performance. In particular, by using only (t) or (t-1, t), we would have missed divisions that span several frames, Fig. 3a.

FN/FP investigation To understand the remaining error on our best N=3-frame U-Net, we investigated the 14 false negatives (FN) and 11 false positives (FP) detected over the 268 ground truth annotations of the set 3 (see Supplementary Movie). We find that 8 FN and 6 FP are located on the image edge; there, missing pixels can explain the failures. Some remaining FN are obvious divisions, but a rare feature - the presence of a high membrane intensity within the cell bulk, as in Fig. 3b - can explain the absence of detection. In contrast, the remaining FP correspond to ambiguous cases, confusing even experts. In particular, cells can quickly migrate from the basal layer to appear in the imaged apical layer; when this occurs near a growing cell, the resulting image sequence is similar to a division, causing false positives, as in Fig. 3c.

Optimal performance on the angle measure Here, we argue that our angle detection scheme cannot be more precise than the natural variability of our data labelling. Indeed, a one-pixel error on the endpoint of a 25-pixel long segment typically leads to an angle error in the $(180/\pi) \times 1/25 = 2.5$ degree range; our mean error on the angle is not significantly larger than this value.

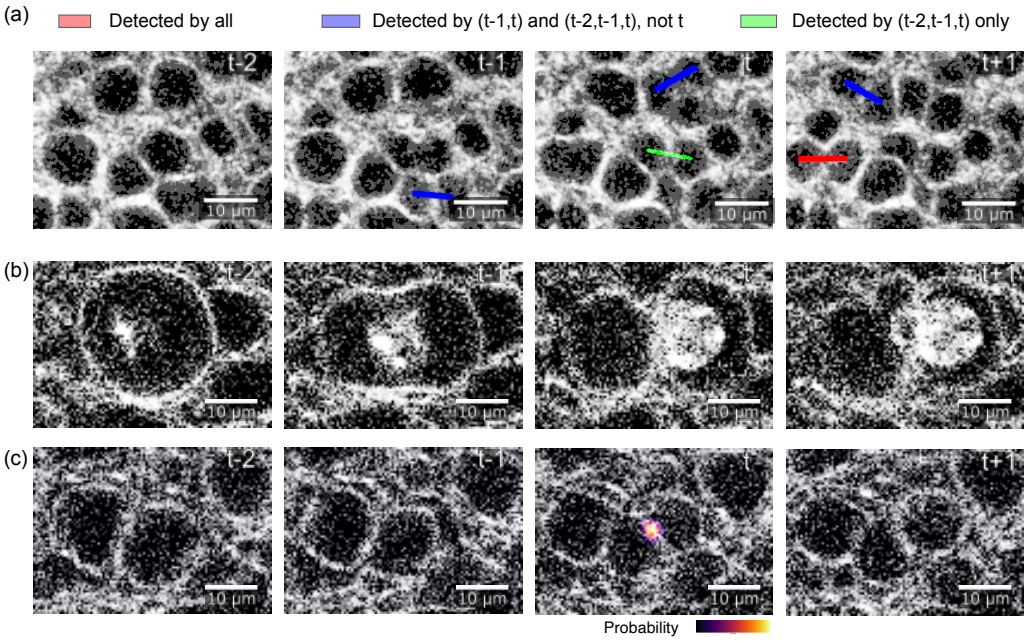

Figure 3. (a) Example of detected and missed divisions when using only one (t), two ($t-1, t$) and three ($t-2, t-1, t$) frames. Red: divisions were detected in all three cases. Blue: divisions using one ($t-1, t$) and two ($t-2, t-1, t$) frames but missed using only one t only. Green: divisions detected only using the ($t-2, t-1, t$) frames. (b-c) Examples of (b) a missed division and (c) a detection that does not belong to the ground truth.

5. CONCLUSION

In this article, we optimized a two-stage neural network for a cell division task. The proposed representation relies on easily detected objects, requiring almost no post-processing. We also investigated (1) the best representation parameters for the cell division detector and (2) the influence of past image frames around the time of the event on the prediction.

6. REFERENCES

[1] L. Bocanegra-Moreno, A. Singh, E. Hannezo, M. Zagorski, and A. Kicheva, "Cell cycle dynamics control fluidity of the developing mouse neuroepithelium," *Nature Physics*, 2023.

[2] K. Ulicna, G. Vallardi, G. Charras, and A. R. Lowe, "Automated deep lineage tree analysis using a bayesian single cell tracking approach," *Frontiers in Computer Science*, 2021.

[3] C. Malin-Mayor, P. Hirsch, L. Guignard, K. McDole, Y. Wan, W. C. Lemon, D. Kainmueller, P. J. Keller, S. Preibisch, and J. Funke, "Automated reconstruction of whole-embryo cell lineages by learning from sparse annotations," *Nature Biotechnology*, vol. 41, pp. 44–49, 1 2023.

[4] K. McDole, L. Guignard, F. Amat, A. Berger, G. Malandain, L. A. Royer, S. C. Turaga, K. Branson, and P. J. Keller, "In toto imaging and reconstruction of post-implantation mouse development at the single-cell level," *Cell*, vol. 175, pp. 859–876.e33, 10 2018.

[5] J. Turley, I. V. Chenchiah, P. Martin, T. B. Liverpool, and H. Weavers, "Deep learning for rapid analysis of cell divisions in vivo during epithelial morphogenesis and repair," *eLife Sciences Publications, Ltd*, 2023. [Online]. Available: http://dx.doi.org/10.7554/eLife.87949.1

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[8] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/125

[9] O. M. O'Connor, R. N. Alnahhas, J. B. Lugagne, and M. J. Dunlop, "Delta 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics," *PLoS Computational Biology*, 2022.

[10] L. Combe, M. Durande, H. Delanoë-Ayari, and O. Cochet-Escartin, "Small hand-designed convolutional neural networks outperform transfer learning in automated cell shape detection in confluent tissues," *PLoS ONE*, 2023.

[11] Y. Jung and J. Hu, "A k-fold averaging cross-validation procedure," *Journal of Nonparametric Statistics*, 2015.