

---

# Reliable Cell Trackers Say “I dunno!”

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

Cell tracking is a key computational task in live-cell microscopy, but fully automated analysis of high-throughput imaging requires reliable and, thus, uncertainty-aware data analysis tools, as the amount of data recorded within a single experiment exceeds what humans are able to overlook. We here propose and benchmark various methods to reason about and quantify uncertainty in linear assignment-based cell tracking algorithms. Our methods take inspiration from statistics and machine learning, leveraging two perspectives on the cell tracking problem explored throughout this work: Considering it as a Bayesian inference problem and as a classification problem. Our methods admit a framework-like character in that they equip any frame-to-frame tracking method with uncertainty quantification. We demonstrate this by applying it to various existing tracking algorithms including the recently presented Transformer-based trackers. We demonstrate empirically that our methods yield useful and well-calibrated tracking uncertainties.

## 1 Introduction

Uncertainty-aware cell tracking is a key requirement for fully automated data analysis of high throughput live-cell microscopy (LCM) data, where images often contain hundreds or thousands of almost indistinguishably looking, moving, growing and dividing cells and where temporal resolution of the time-lapses is limited by biological and technical considerations such as phototoxicity (Tinevez et al., 2012) or camera movement speed in multi-colony setups (Seiffarth et al., 2025). LCM enables researchers to analyze cellular behavior beyond the population level, revealing dynamics, development and multi-species interactions (Blöbaum et al., 2024; Fante et al., 2024; Burmeister et al., 2021). The recent advances in computer vision, mainly driven by the success of deep learning and ever-improving computational resources, form a substantial pillar of modern LCM analyses, as the amount of data collected in a single experiment exceeds what humans are able to overlook (Kasahara et al., 2023; Seiffarth et al., 2025; Witting et al., 2025). The strong reliance on computational tools demands for high reliability and trustworthiness, which is improved by uncertainty-aware analyses aiming to reliably estimate the confidence in their own predictions. While over-confidence in deep neural networks encountering distribution shifts is a commonly known issue in the deep learning community (Guo et al., 2017; Kristiadi et al., 2020), uncertainty estimation as a remedy has – so far – attracted only little attention in cell tracking.

In this work, we strive to make cell tracking more reliable by complementing it with uncertainty estimation. To this end, we explore two perspectives on the tracking problem, considering it as a Bayesian inference and as a classification problem. Our Bayesian perspective gives rise to the *cell tracking posterior*, motivating sampling and approximate sampling methods for uncertainty quantification, which however come at increased computational costs. The classification perspective provides less costly, nevertheless useful alternatives to quantify uncertainty. Moreover, the classification perspective also provides tools to evaluate and calibrate uncertainty estimates using known

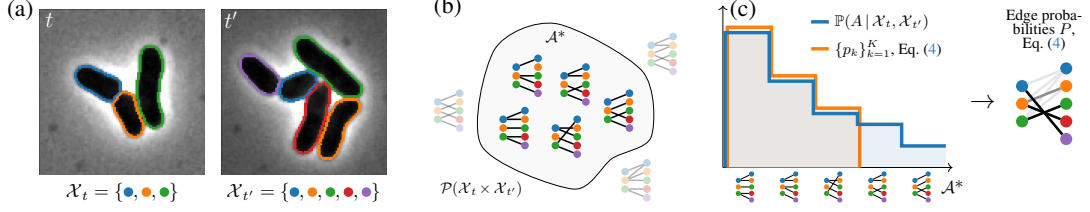


Figure 1: Schematic presentation of our Bayesian perspective and inference method for uncertainty-aware cell tracking. **a** depicts the detections  $\mathcal{X}_t, \mathcal{X}_{t'}$  of two consecutive frames. **b** The set of biologically feasible assignments  $\mathcal{A}^*$  forms a subset within the set of all possible many-to-many assignments. **c** Sorting assignment solutions by their posterior density we approximate the full posterior by a set of the  $K$  most plausible solutions.

techniques such as *temperature scaling* (Guo et al., 2017). Notably, the methods under consideration are applicable to the large family of linear assignment-based tracking methods, which are embedded naturally within our framework.

## 2 Probabilistic Cell Tracking

**Cell Tracking as Bayesian Inference** In the case of multiple cells  $x_1, \dots, x_m \in \mathcal{X}_t$  and  $x'_1, \dots, x'_n \in \mathcal{X}_{t'}$ , we aim to find an *assignment*  $A \in \mathcal{P}(\mathcal{X}_t \times \mathcal{X}_{t'})$ , such that every mother cell is represented at most twice and every daughter cell is represented at most once. We denote the subset of all assignments that adhere to those constraints as  $\mathcal{A}^* \subset \mathcal{P}(\mathcal{X}_t \times \mathcal{X}_{t'})$ , the *biologically feasible assignments* (cf. Figure 1b). The common approach to solve for a single assignment is to optimize the joint likelihood of observing cells  $\mathcal{X}_{t'}$  given  $\mathcal{X}_t$  and an assignment  $A$ , i.e.

$$A_{\text{opt}} := \arg \max_{A \in \mathcal{A}^*} \mathbb{P}(\mathcal{X}_{t'} | \mathcal{X}_t, A) = \arg \max_{A \in \mathcal{A}^*} \sum_{(x_i, x'_j) \in A} -w(x_i, x'_j) - mw_a - nw_d \quad (1)$$

where  $m$  and  $n$  are the numbers of appearing and disappearing cells,  $-w_a$  and  $-w_d$  are the respective log probabilities of the events, and  $w(x_i, x'_j)$  is some chosen cost function. Considering that the structure of the solution  $A \in \mathcal{A}^*$ , i.e. its biological feasibility, is known *a priori* without the need for any actual observations, the formulation from Equation (1) lends itself nicely to a Bayesian interpretation with  $A_{\text{opt}}$  being the *maximum a posteriori* (MAP) estimate of the posterior distribution of assignments

$$\mathbb{P}(A | \mathcal{X}_{t'}, \mathcal{X}_t) \propto \mathbb{P}(\mathcal{X}_{t'} | \mathcal{X}_t, A) \mathbb{P}(A) \quad (2)$$

where we choose the uniform distribution over valid assignments  $\mathcal{A}^*$  as prior distribution,  $\mathbb{P}(A) := \mathbb{1}_{\{A \in \mathcal{A}^*\}} / |\mathcal{A}^*|$ . Given this *cell tracking posterior* distribution, we obtain the predictive distribution of the event that  $x_i$  is the mother of  $x'_j$

$$P_{ij} := \mathbb{P}((x_i, x'_j) \in A | \mathcal{X}_t, \mathcal{X}_{t'}) = \sum_{A \in \mathcal{A}^*} \mathbb{1}_{\{(x_i, x'_j) \in A\}} \mathbb{P}(A | \mathcal{X}_t, \mathcal{X}_{t'}) \quad (3)$$

as the weighted frequency of  $x_i$  being the mother of  $x'_j$  among all possible solutions. We estimate Equation (3) by means of the self-normalized importance-weighted estimator (Tokdar and Kass, 2010)

$$P_{ij} \approx \sum_{k=1}^K \mathbb{1}_{\{(x_i, x'_j) \in A_k\}} \cdot p_k, \quad \text{with weight} \quad p_k := \frac{\mathbb{P}(\mathcal{X}_{t'} | \mathcal{X}_t, A_k) \mathbb{P}(A_k)}{\sum_{l=1}^K \mathbb{P}(\mathcal{X}_{t'} | \mathcal{X}_t, A_l) \mathbb{P}(A_l)}, \quad (4)$$

for  $k = 1, \dots, K$  and where we now assumed  $A_1, \dots, A_K$  to be the top- $K$  best solutions. We sketch this method in Figure 1c and refer to it as AS in the results. As an alternative approach, we consider perturbation of input features like the detected masks or positions. We propose two approaches: FP, which averages the costs over the perturbed features, and FP+A, which solves the assignment problem for each pair of perturbed input features and approximates the predictive posterior Eq. (3) as the frequency at which an event  $\{(x_i, x'_j) \in A\}$  is contained in the resulting assignment solutions.

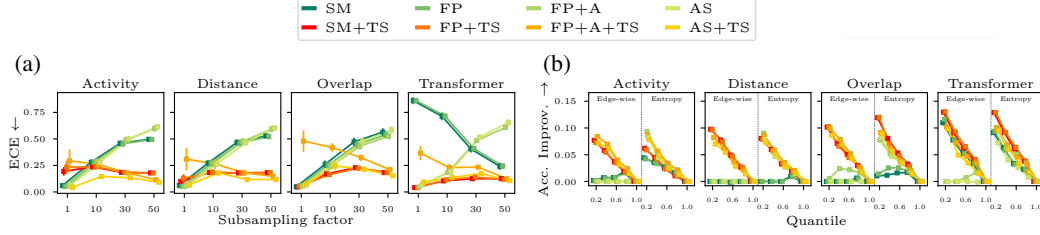


Figure 2: (a) Mean ECE and standard error thereof across datasets at decreasing temporal resolution. (b) Mean accuracy improvement when only considering the most certain tracking predictions, using either the edge-wise confidence (cf. Eq. (5)) or entropy (cf. Eq. (6)) as uncertainty estimate.

68 **Cell Tracking as Classification Task** Cell tracking can be viewed as a multi-class classification  
69 problem by considering the problem of choosing the right mother for each individual daughter. That  
70 is, for your favorite daughter  $x'_j \in \mathcal{X}_{t'}$ , the *classes* are just the mothers  $\mathcal{X}_t$  from the preceding  
71 frame. The class – or rather mother – probabilities are obtained by performing daughter-wise soft-  
72 max normalization of the predicted costs

$$P_{i|j} = \mathbb{P}((x_i, x'_j) \in A | x'_j, \mathcal{X}_t, \mathcal{X}_{t'}) = \frac{\exp\{-w(x_i, x'_j)\}}{\sum_{x \in \mathcal{X}_t} \exp\{-w(x, x'_j)\}}. \quad (5)$$

73 We can interpret the softmax probabilities  $P_{i|j}$  from Equation (5) as confidence scores, enabling  
74 their usage for uncertainty estimation. This approach is also compatible to the Bayesian approaches  
75 from Section 2 by simply column-normalizing the edge probabilities  $P_{ij}$  from Equation (4) as  $P_{i|j} =$   
76  $P_{ij} / \sum_k P_{kj}$ . Moreover, the classification view enables to estimate and improve calibration using  
77 temperature scaling (TS ) as in Guo et al. (2017). The latter however is only applicable if annotated  
78 tracking data is available. Finally, as an alternative measure of uncertainty we can also consider the  
79 daughter-wise entropy  $\mathcal{H}_j$  of the conditional distribution  $P_{i|j}$

$$\mathcal{H}_j = - \sum_i P_{i|j} \log P_{i|j}. \quad (6)$$

### 80 3 Results

81 In our experiments we investigate the calibration of tracking algorithms by Ruzaeva et al. (2022,  
82 Activity), Crocker and Grier (1996, Distance), Fukai and Kawaguchi (2023, Overlap) & Gallusser  
83 and Weigert (2025, Transformer). We consider 2D+t datasets from the Cell Tracking Challenge  
84 (Maška et al., 2014, 2023) and a large-scale microbial time-lapse microscopy sequence (Seiffarth  
85 et al., 2025). Our results show that these tracking algorithms are not *per se* well-calibrated (cf. SM ,  
86 FP , FP+A & AS  in Fig. 2a), but calibration is achievable using TS, if annotated tracking data  
87 is available (cf. SM+TS , FP+TS , FP+A+TS & AS+TS  in Fig. 2a). Further, we consider the  
88 accuracy improvement achievable by sorting out the most uncertain tracking decisions and present  
89 results in Fig. 2b. This setting reflects a possible real-world scenario where a human annotator is  
90 guided by the uncertainty estimate to manually correct the automated tracking result. We notice that  
91 using our proposed daughter-wise entropy approach for quantifying uncertainty, even the vanilla  
92 tracking algorithms can produce ‘useful’ uncertainties, *i.e.* uncertainties that positively correlate  
93 with model performance. However, in particular our approximate Bayesian methods FP+A &  
94 AS  show higher improvements in accuracy and are less reliant on temperature scaling and, thus,  
95 annotated tracking data. This comes at the price of increased computational costs of those methods.  
96 Overall, we also observe that using the entropy-based uncertainty estimation (cf. Eq. (6)) is more  
97 reliable in this setting.

## References

- Luisa Blöbaum, Luca Torello Pianale, Lisbeth Olsson, and Alexander Grünberger. Quantifying microbial robustness in dynamic environments using microfluidic single-cell cultivation. *Microbial Cell Factories*, 23(1):44, 2024. [1](#)
- Alina Burmeister, Qiratt Akhtar, Lina Hollmann, Niklas Tenhaef, Fabienne Hilgers, Fabian Hogenkamp, Sascha Sokolowsky, Jan Marienhagen, Stephan Noack, Dietrich Kohlheyer, and Alexander Grünberger. (Optochemical) Control of Synthetic Microbial Coculture Interactions on a Microcolony Level. *ACS Synthetic Biology*, 10(6):1308–1319, 2021. Publisher: American Chemical Society. [1](#)
- John C. Crocker and David G. Grier. Methods of Digital Video Microscopy for Colloidal Studies. *Journal of Colloid and Interface Science*, 179(1):298–310, 1996. [3](#)
- Niklas Fante, Christian K. Desiderato, Christian U. Riedel, and Alexander Grünberger. Time-resolved cell-to-cell heterogeneity of *Listeria innocua* after nisin exposure. *Frontiers in Bioengineering and Biotechnology*, 12, 2024. Publisher: Frontiers. [1](#)
- Yohsuke T Fukai and Kyogo Kawaguchi. LapTrack: linear assignment particle tracking with tunable metrics. *Bioinformatics*, 39(1):btac799, 2023. [3](#)
- Benjamin Gallusser and Martin Weigert. TRACKASTRA: Transformer-Based Cell Tracking for Live-Cell Microscopy. In *Computer Vision – ECCV 2024*, pages 467–484, Cham, 2025. Springer Nature Switzerland. [3](#)
- Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1321–1330. PMLR, 2017. ISSN: 2640-3498. [1](#), [2](#), [3](#)
- Keitaro Kasahara, Markus Leygeber, Johannes Seiffarth, Karina Ruzaeva, Thomas Drepper, Katharina Nöh, and Dietrich Kohlheyer. Enabling oxygen-controlled microfluidic cultures for spatiotemporal microbial single-cell analysis. *Frontiers in Microbiology*, 14, 2023. Publisher: Frontiers. [1](#)
- Agustinus Kristiadi, Matthias Hein, and Philipp Hennig. Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks. In *Proceedings of the 37th International Conference on Machine Learning*, pages 5436–5446. PMLR, 2020. ISSN: 2640-3498. [1](#)
- Martin Maška, Vladimír Ulman, David Svoboda, Pavel Matula, Petr Matula, Cristina Ederra, Ainhoa Urbiola, Tomás España, Subramanian Venkatesan, Deepak M.W. Balak, Pavel Karas, Tereza Bolcková, Markéta Štreitová, Craig Carthel, Stefano Coraluppi, Nathalie Harder, Karl Rohr, Klas E. G. Magnusson, Joakim Jaldén, Helen M. Blau, Oleh Dzyubachyk, Pavel Křížek, Guy M. Hagen, David Pastor-Escuredo, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Arrate Muñoz-Barrutia, Erik Meijering, Michal Kozubek, and Carlos Ortiz-de Solorzano. A benchmark for comparison of cell tracking algorithms. *Bioinformatics*, 30(11):1609–1617, 2014. [3](#)
- Martin Maška, Vladimír Ulman, Pablo Delgado-Rodriguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A. Guerrero Peña, Tsang Ing Ren, Elliot M. Meyerowitz, Tim Scherr, Katharina Löffler, Ralf Mikut, Tianqi Guo, Yin Wang, Jan P. Allebach, Rina Bao, Noor M. Al-Shakarji, Gani Rahmon, Imad Eddine Toubal, Kannappan Palaniappan, Filip Lux, Petr Matula, Ko Sugawara, Klas E. G. Magnusson, Layton Aho, Andrew R. Cohen, Assaf Arbelle, Tal Ben-Haim, Tammy Riklin Raviv, Fabian Isensee, Paul F. Jäger, Klaus H. Maier-Hein, Yanming Zhu, Cristina Ederra, Ainhoa Urbiola, Erik Meijering, Alexandre Cunha, Arrate Muñoz-Barrutia, Michal Kozubek, and Carlos Ortiz-de Solórzano. The Cell Tracking Challenge: 10 years of objective benchmarking. *Nature Methods*, 20(7):1010–1020, 2023. Number: 7 Publisher: Nature Publishing Group. [3](#)
- Karina Ruzaeva, Jan-Christopher Cohrs, Keitaro Kasahara, Dietrich Kohlheyer, Katharina Nöh, and Benjamin Berkels. Cell tracking for live-cell microscopy using an activity-prioritized assignment strategy. In *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*, pages 1–7, 2022. [3](#)
- Johannes Seiffarth, Luisa Blöbaum, Richard D. Paul, Nils Friederich, Angelo Jovin Yamachui Sitcheu, Ralf Mikut, Hanno Scharr, Alexander Grünberger, and Katharina Nöh. Tracking One-in-a-Million: Large-Scale Benchmark for Microbial Single-Cell Tracking with Experiment-Aware Robustness Metrics. In *Computer Vision – ECCV 2024 Workshops*, pages 318–334, Cham, 2025. Springer Nature Switzerland. [1](#), [3](#)
- Jean-Yves Tinevez, Joe Dragavon, Lamya Baba-Aissa, Pascal Roux, Emmanuelle Perret, Astrid Canivet, Vincent Galy, and Spencer Shorte. Chapter fifteen - A Quantitative Method for Measuring Phototoxicity of a Live Cell Imaging Microscope. In *Methods in Enzymology*, pages 291–309. Academic Press, 2012. [1](#)

- 152 Surya T. Tokdar and Robert E. Kass. Importance sampling: a review. *WIREs Computational Statistics*, 2(1):  
153 54–60, 2010. [2](#)
- 154 Lennart Witting, Johannes Seiffarth, Birgit Stute, Tim Schulze, Jan Matthis Hofer, Katharina Nöh, Marion  
155 Eisenhut, Andreas P. M. Weber, Eric von Lieres, and Dietrich Kohlheyer. A microfluidic system for the  
156 cultivation of cyanobacteria with precise light intensity and CO<sub>2</sub> control: enabling growth data acquisition  
157 at single-cell resolution. *Lab on a Chip*, 25(3):319–329, 2025. Publisher: Royal Society of Chemistry. [1](#)