Cell Tracking according to Biological Needs - Strong Mitosis-aware Multi-Hypothesis Tracker with Aleatoric Uncertainty

Timo Kaiser [1] **Maximilian Schier** [1] **Bodo Rosenhahn** [1]

Abstract

Cell tracking and segmentation enable biologists to extract insights from large-scale microscopy time-lapse data. Driven by local accuracy metrics, current tracking approaches often suffer from a lack of long-term consistency and an inability to correctly reconstruct lineage trees. To address this issue, we introduce a novel assignment strategy consisting of two key components. First, we propose an uncertainty estimation technique for motion estimation frameworks. This method relaxes single-point motion representations into probabilistic spatial densities using problem-specific test-time augmentations. Second, we leverage these spatial densities to define a novel mitosis-aware assignment problem formulation. This formulation allows multi-hypothesis trackers to model cell divisions and resolve false associations and mitosis detections based on long-term conflicts. Our framework integrates explicit biological knowledge into assignment costs and combines it with learned representations derived from spatial densities. We evaluate our approach on nine competitive datasets and demonstrate that it substantially outperforms the current state-of-the-art on biologically inspired metrics, achieving improvements by a factor of approximately six and providing new insights into the behavior of motion estimation uncertainty.

1. Introduction

Cell tracking and lineage reconstruction allow researchers to investigate the fate of cells over extended periods, such as analyzing liver diseases (Yoon et al., 2024) or studying the interaction between redox signaling and cell migration in the context of breast cancer (Kukulage et al., 2024). Automated tracking and segmentation algorithms are therefore valuable tools for reducing the substantial effort required to analyze optical microscopy output in biomedical research (Antony et al., 2013). These algorithms aim to segment cells in images, track them over time to form trajectories, and detect mother-daughter relationships arising from mitotic cell divisions (mitosis). Figure 2 presents some example microscopy image sequences that visualize the setting of cell tracking and lineage reconstruction. By enabling the analysis of large datasets, these methods facilitate studies such as (Malin-Mayor et al., 2023), which reconstructs cell lineages of whole-embryo development encoded in 4.7 terabytes of images. However, despite no-

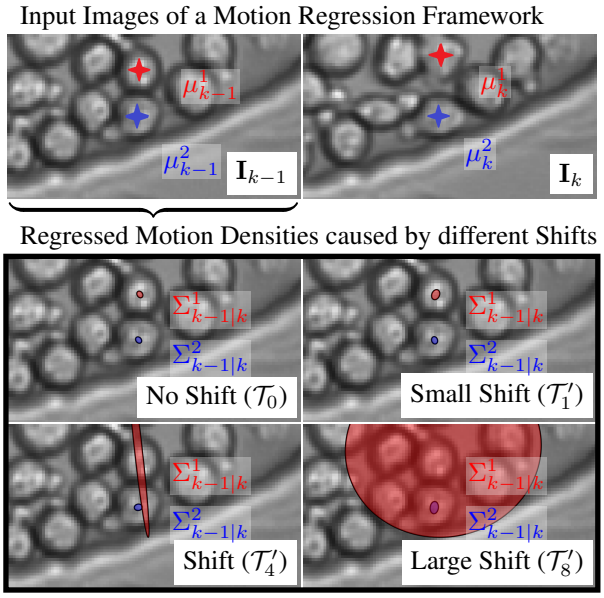

Figure 1. Uncertainty in motion regression. The red and blue Gaussians represent the distribution of cell motion estimations when applying our test-time shifts to the input image $\mathbf{I}_{k-1}$. The estimation variance $\mathbf{\Sigma}^{\mathcal{Z}}_{k-1|1}$ is small if standard augmentations are applied ($\mathcal{T}_0$). It increases for the red cell if $\mathbf{I}_{k-1}$ is shifted by one ($\mathcal{T}'_1$), four ($\mathcal{T}'_4$), or eight pixels ($\mathcal{T}'_8$) while the blue remains small. This indicates a certain motion estimation for the blue but an uncertain one for the red cell.

[1]Institute for Information Processing, Leibniz University Hanover, Hanover, Germany. Correspondence to: Timo Kaiser <kaiser@tnt.uni-hannover.de>.

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2025.358314.

table advancements, (Malin-Mayor et al., 2023) reports that only 50% of cells are correctly tracked in the long term, and mitosis detection requires further improvement to enable deeper lineage tree analysis.

The reported cell tracking improvements in (Malin-Mayor et al., 2023), which still lack long-term consistency and robust mitosis detection, can be systematically analyzed. Revisiting the so-called *technical* metrics (Bernardin & Stiefelhagen, 2008; Matula et al., 2015) used in leading tracking benchmarks (Maška et al., 2014; Anjum & Gurari, 2020), the isolated focus on local correctness appears reasonable at first glance. Local errors, such as missing segmentations, are penalized more heavily than rare association errors or missed mitosis detections. In fact, cell tracking metrics penalize the correction of faulty associations when cells are re-associated in future frames. As a result, some cell tracking methods achieve near-perfect benchmark scores, approaching 100% tracking accuracy (Maška et al., 2023), while primarily relying on local cues within only a few consecutive frames (Löffler & Mikut, 2022; Gallusser & Weigert, 2024; Malin-Mayor et al., 2023). Although conventional technical metrics suggest that the tracking problem is nearly solved, the aforementioned limitations persist, undermining the practical usability of cell tracking algorithms.

The discrepancy between technical metrics and biologically relevant indicators (Ulman et al., 2017), which are essential for biomedical research, has been highlighted in several studies (Maška et al., 2023; Kaiser et al., 2024; O'Connor & Dunlop, 2024). Recognizing this gap, the primary tracking benchmark (Maška et al., 2014) has recently begun incorporating these biologically relevant metrics into specialized challenges.

To bridge the gap between technical and biological measures, this work addresses the challenges of missing long-term consistency and mitosis detection in modern cell tracking algorithms. We tackle these issues by introducing an extended association framework that enhances existing heuristic approaches with probabilistic models and explicitly incorporates the likelihood of cell division during mitosis. Specifically, we first introduce an advanced test-time augmentation (Wang et al., 2019) using spatial shifts to estimate position and motion densities of cells. These densities capture aleatoric uncertainty (Gawlikowski et al., 2023) of tracking methods, which arises from ambiguous image data where the correct tracking result is not obvious. To demonstrate our contribution, we apply this approach to the heuristic state-of-the-art framework *EmbedTrack*, relaxing its discrete position and motion predictions into probabilistic densities.

As our second contribution, we leverage these densities to introduce a novel mitosis-aware *Multi-Hypothesis Tracking (MHT) framework*. MHT solves the tracking task by integrating association costs across all possible hypotheses and selecting the most likely one *a posteriori*. Compared to standard MHT (Reid, 1979), our framework supports non-bijective assignments to model mitosis—the process by which a single parent cell divides to produce two genetically identical daughter cells. While mitosis timing depends on various factors, the interval between successive mitotic events can be approximated by an Erlang distribution in homogeneous cell cultures (Yates et al., 2017; Paul et al., 2024). Thus, we explicitly model mitosis probability based on expected lifetimes and define mitosis costs derived from the Erlang distribution. These globally inferred mitosis costs are combined with local assignment costs, derived from position and motion densities, to determine the most likely tracking hypothesis. By integrating global proliferation aspects with local aleatoric uncertainty, our framework resolves long-term conflicts *a posteriori*, correcting both association errors and false or missing mitosis detections.

We evaluated our method on nine well-established *Cell Tracking Challenge* (Maška et al., 2014) datasets and demonstrate that it achieves substantial improvements in biologically relevant metrics—by up to a factor of 5.75 compared to the state-of-the-art. At the same time, our method maintains performance on technical measures without notable differences, confirming our initial assertion that current algorithms do not align with biomedical metrics while showing that long-term consistency can significantly enhance tracking quality.

Our contributions can be summarized as follows:

- We estimate position and motion densities of cells using a novel test-time augmentation strategy.
- We introduce a mitosis-aware MHT tracker to model cell splits.
- We define mitosis costs to incorporate biological knowledge and ensure long-term consistency.
- By integrating these components, we establish a new state-of-the-art based on biologically inspired metrics.

The remainder of this article is structured as follows: Section 2 reviews prior work relevant to cell tracking and uncertainty estimation. Section 3 provides background on the baseline methods, *EmbedTrack* and MHT tracking. Next, Section 4 introduces our novel uncertainty estimation technique and demonstrates its implementation in *EmbedTrack*. These densities are then utilized in our mitosis-aware MHT framework, presented in Section 5, where we introduce non-bijective assignment and mitosis costs. Finally, we evaluate and discuss our results in Section 6 and conclude in Section 7. **Code is available at:** https://github.com/TimoK93/BiologicalNeeds .

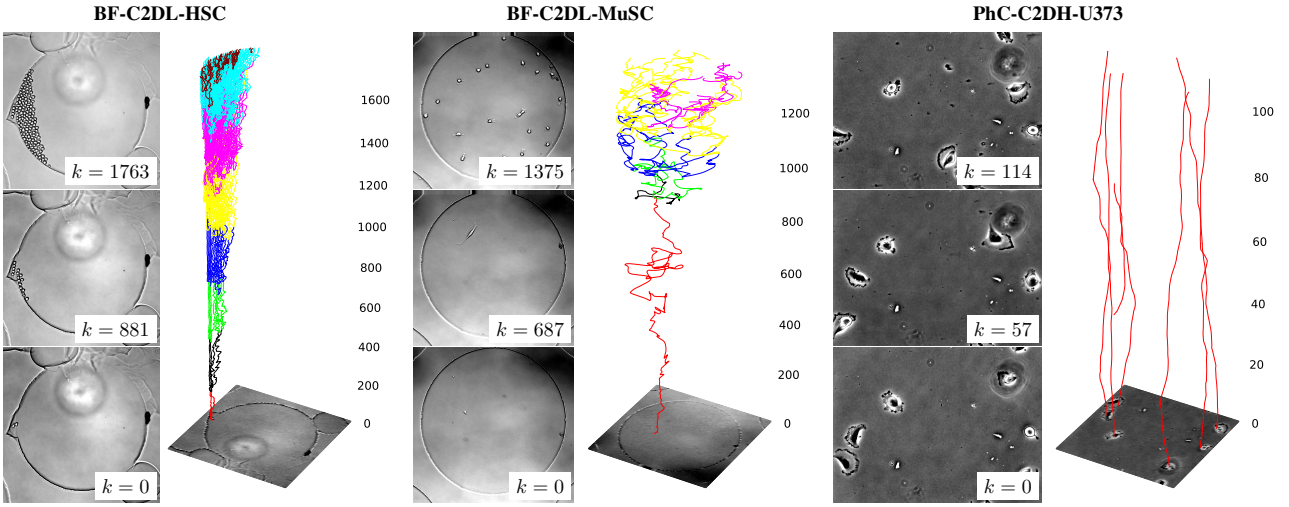

Figure 2. Lineage trees of varying complexities, where the trajectory color indicates the cell generation. For instance, mouse hematopoietic stem cells (BF-C2DL-HSC) exhibit extensive proliferation, providing rich cell cycle information. The mouse muscle stem cells (BF-C2DL-MuSC) dataset combines strong proliferation with high cell motility. In contrast, other settings, such as Glioblastoma-astrocytoma U373 cells (PhC-C2DH-U373), show little to no proliferation, with minimal mitosis and immobile cells. Our method addresses the more complex scenario involving proliferation and motility by incorporating cell cycle information and introducing a robust implicit motion model.

2. Related Work

Prior works related to our contribution can be divided into cell tracking and uncertainty estimation techniques, especially for multiple object tracking. This section discusses both.

2.1. Cell Tracking

The prevalent *Tracking-by-Detection* (TbD) paradigm employs a two-stage approach, where proposals of object detections obtained from detectors like (Ronneberger et al., 2015; Lalit et al., 2022) are associated, *e.g.* , using simple strategies such as greedy matching (Bao et al., 2021; Stegmaier et al., 2012), based on spatial similarity scores or Siamese networks (Gupta et al., 2019). More sophisticated association strategies aim to find globally optimal associations between all detections by constructing a graph and optimizing a minimum-cost flow, for example, with the *Viterbi* algorithm (Magnusson et al., 2015). The graph's edge costs adhere to the Markov property and rely only on local features between two temporal positions. These features can be spatial or appearance-based, such as distances between the positions (Löffler et al., 2021) or phase correlation of detections (Scherr et al., 2020). Additionally, over- and under-segmentations are addressed with graph models (Schiegg et al., 2013). Classical graph-based methods disregard the global cell life cycle because they do not include features between detections with long temporal distances. A more advanced graph-based approach uses graph neural networks that share feature information over longer temporal gaps with autoregressive message passing (Ben-Haim & Raviv, 2022).

Other methods following the TbD paradigm attempt to maintain global consistency by extracting and clustering appearance embeddings, such as with spatio-temporal Mean Shift (Payer et al., 2019) or Siamese Networks (Chen et al., 2021), or by using recurrent neural networks like LSTMs in the object detector (Arbelle & Raviv, 2019). Recent trends use transformers to link or identify detections, like *Trackastra* (Gallusser & Weigert, 2024) or *Cell DINO* (Liao et al., 2024). *Trackastra* uses visual or spatial features and potentially learns global dependencies. However, *Trackastra* performs well on short sequences but not on very large ones, where proliferation and cell division are more frequent. *Cell DINO* uses mask tokens to detect former cells in future frames.

Another paradigm is *Tracking-by-Regression* (TbR), where the position and motion of potential objects are regressed jointly based on local features using neural networks (Bergmann et al., 2019). For example, *EmbedTrack* (Löffler & Mikut, 2022) introduces a multitask regression head that takes two subsequent frames as input and estimates instance segmentation, center position, and motion of cells. By its very nature, this paradigm can only preserve local dependencies. Nevertheless, TbR proves to be an effective and competitive data-driven tracking approach, leading to state-of-the-art results on technical tracking metrics.

The most relevant class of global optimal association strategies in our context follows the TbD paradigm and is known

as *Multi-Hypothesis Tracking* (MHT) (Reid, 1979). These trackers re-evaluate multiple hypotheses *a posteriori* based on new information and can use prior sub-optimal hypotheses to resolve errors. The MHT framework is more frequently used in general multiple object tracking and is improved with random finite sets (Mori et al., 1986), such as Poisson multi-Bernoulli mixtures (Granström et al., 2018). MHT is also applied in cell tracking (Rezatofighi et al., 2015; Theorell et al., 2019; Nguyen et al., 2021; Hossain et al., 2018). For instance, these methods model mitosis as a hypothesis and score it based on local appearance (Nguyen et al., 2021; Hossain et al., 2018), but do not incorporate cell lifetime or other long-term temporal features.

2.2. Uncertainty Estimation

Without focusing on tracking, a lot of research is dedicated to estimating the uncertainty in NN-driven predictions (Kaiser et al., 2023; Wehrbein et al., 2025). In medical imaging, Bayesian approaches (Wang & Lukasiewicz, 2022) approximate epistemic uncertainty, while test-time augmentation (Wang et al., 2019) is used to quantify aleatoric uncertainty. However, NN-derived uncertainty estimation is rarely applied, neither in cell tracking nor general object tracking. To the best of our knowledge, the only commonly utilized NN-derived uncertainty in tracking pertains to the probability of being clutter to filter detection noise, as seen in (Hornakova et al., 2021) and other works. A few approaches employ advanced strategies like normalizing flows (Mancusi et al., 2023) for optimizing association during training or fuzzy logic (Stegmaier & Mikut, 2017). More recently, methods have employed uncertainty quantification for the association task in general object tracking (Zhou et al., 2024; Solano-Carrillo et al., 2024).

3. Preliminaries

The task of cell tracking and segmentation involves detecting and segmenting each visible cell while consistently assigning unique IDs over time. Additionally, cell division must be detected to reconstruct the lineage tree.

Formally, the input is an image sequence $\mathcal{I} = \{\mathbf{I}_k\}_{k=1}^{K}$, containing images $\mathbf{I}_k \in \mathbb{R}^{I_H \times I_W}$, where k denotes the temporal position and $I_H \times I_W$ represents the spatial resolution. The resulting tracking and segmentation outputs are discrete ID maps $\mathbf{M}_k \in \mathbb{N}_0^{I_H \times I_W}$ corresponding to the images. A map $\mathbf{M}_k$ labels pixels based on whether they belong to a specific cell ID or the background. Pixels belonging to a specific cell j are denoted as $\mathcal{M}_k^j$. Segmented areas from different images that belong to the same cell instance should be labeled with the same unique ID to form temporally consistent tracks. To construct a lineage tree, every cell ID should be assigned to its parent ID from which it descends, or assigned to 0 if the parent does not exist in the given image sequence.

Table 1. Notation

General		Cell Detections and Representation	
$\mathbf{I}$	Image	j	Index for Detections
I_W	Image Width	i	Index for previous Objects
I_H	Image Height	$\mathcal{Z}$	Set of Detections
$\mathcal{I}$	Image Sequence	$N^{\mathcal{Z}}$	Num. of Detections in $\mathcal{Z}$
K	Sequence Length	μ	Mean of Spatial Density
k	Time Index	Σ	Variance of Spatial Density
		λ^C	Clutter Probability
		Age(i)	Age of Cell i
		r	Probability of Existence
Uncertainty in *EmbedTrack*		MHT	
$\mathbf{M}$	Label Map	$\mathcal{H}$	Hypotheses Parameter
$\mathcal{M}^j$	Pixels belonging to label j	H	Num. of Hypotheses
$\mathbf{D}$	Segmentation Score	H_max	Max. Num. of Hypotheses
$\mathbf{O}^\mathrm{C}$	Centroid Estimation	h	Hypothesis
$\mathbf{O}^\mathrm{M}$	Motion Estimation	l^h	Hypothesis Likelihood
$\mathbf{O}^{\mathrm{C},\mathbf{\Sigma}}$	Centroid Covariance	N^h	Num. of Cells in h
$\mathbf{O}^{\mathrm{M},\mathbf{\Sigma}}$	Motion Covariance	P^B	Probability of Birth
$\mathbf{p}$	Index for a Pixel	P^D	Probability of Detection
$\mathcal{T}$	Image Augmentation	Ψ	Assignment Hypotheses
		A_max	Max. Num. of Assignments
		$\mathbf{C}$	Assignment Cost Matrix
		c	Cost Value
		u	Abbreviation for *Unassigned*

To solve the described task, this paper presents a fully probabilistic algorithm without heuristics, which can be applied to existing baseline frameworks. Next, we recapitulate two baseline methods that aim to solve the described tasks in different ways. First, we discuss the heuristic neural network *EmbedTrack* (Löffler & Mikut, 2022), which is enhanced to predict probabilistic position and motion densities through our first contribution. Second, we describe the probabilistic multi-hypothesis tracking framework (MHT) (Reid, 1979), which is not directly suitable for cell tracking, but is later extended using the densities to model cell division and lineage reconstruction in our second contribution. Our novel fully probabilistic tracking framework leverages the local expressiveness of the first method while adapting the global optimal association strategy of the latter. The notations used are summarized in Table 1.

3.1. Motion Regression with EmbedTrack

Our baseline *EmbedTrack* (Löffler & Mikut, 2022) follows the TbR paradigm and solves the tracking and segmentation tasks by jointly estimating cell segments and motion with a single neural network (NN). It operates in two steps: (1) estimating pixel-wise segmentation scores, offsets to cell centroids, and motion to the last frame, and (2) clustering these into cell instances linked to cells in the previous frame. The architecture of *EmbedTrack* is visualized in the upper part of Figure 3, excluding the red elements.

More precisely, using two subsequent images $(\mathbf{I}_{k-1}, \mathbf{I}_k)$ as input, the first step estimates a pixel-wise probability of belonging to a cell $\mathbf{D}_k \in (0,1)^{I_H \times I_W}$, a relative offset to

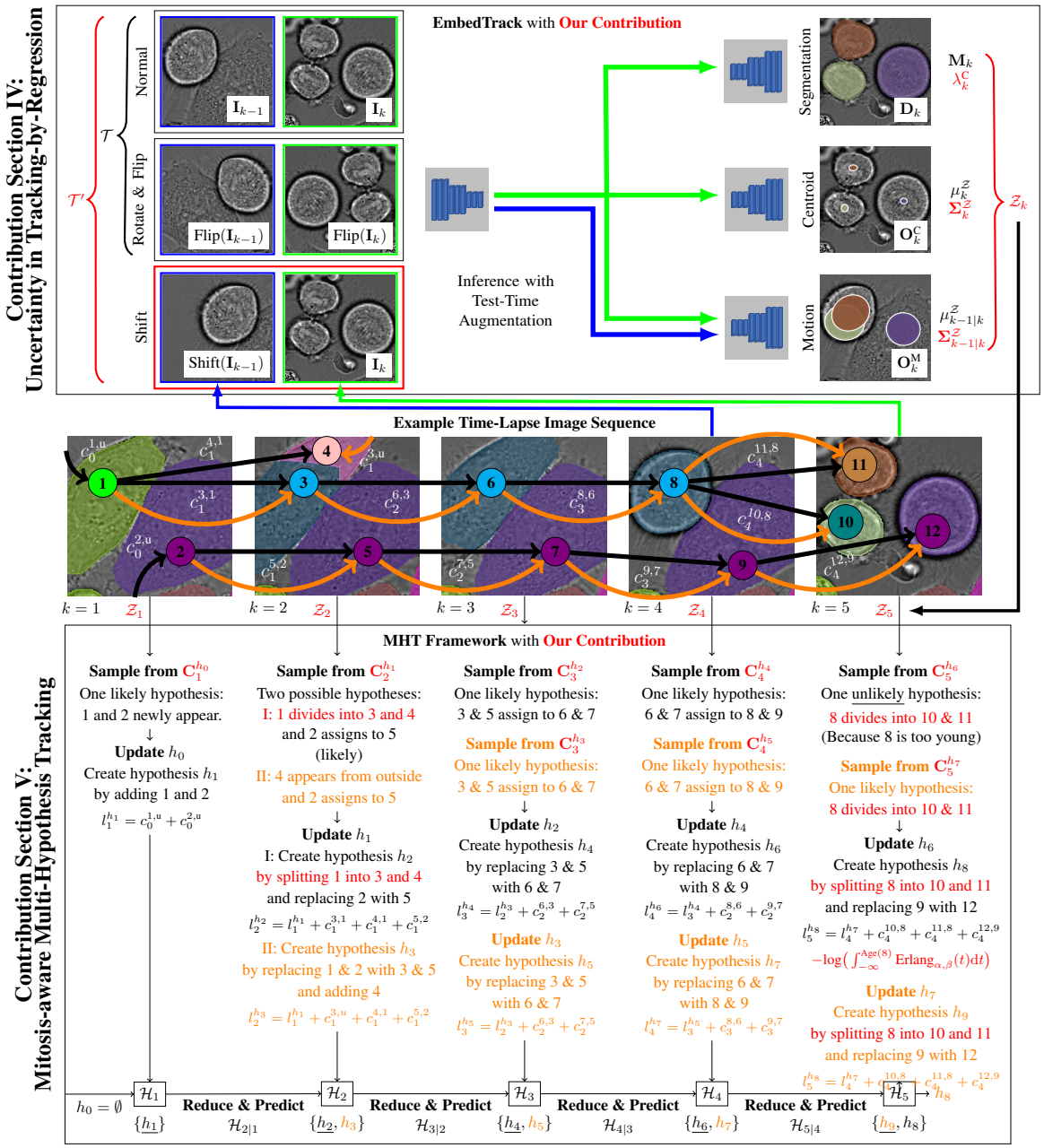

Figure 3. Overview of our tracking framework applied to a time-lapse image sequence. Compared to vanilla frameworks, red elements highlight our contributions and novelties. The centered image sequence shows the result of vanilla *EmbedTrack* (Löffler & Mikut, 2022) with an over-segmentation in frame $k = 2$, which is resolved using our framework. Our test-time augmentation strategy (top) generates position and motion estimation densities that are required to calculate association costs in our extended MHT framework (Reid, 1979) (bottom). First, to estimate cell segments (M_k) as well as their Gaussian-distributed position ($\mu_k^{\mathcal{Z}}$, $\mathbf{\Sigma}_k^{\mathcal{Z}}$) and motion densities ($\mu_{k-1|k}^{\mathcal{Z}}$, $\mathbf{\Sigma}_{k-1|k}^{\mathcal{Z}}$), we augment the two subsequent input images ($\mathbf{I}_{k-1}$, $\mathbf{I}_k$). We extend standard augmentations ($\mathcal{T}$) with problem-specific shifts ($\mathcal{T}'$) that increase the estimation variance in ambiguous images. Second, the densities are used as input detections $\mathcal{Z}_k$ for iterations of our MHT framework. Hypothesis states from the last frame ($\mathcal{H}_{k|k-1}$) are compared with the estimated position and motion densities of $\mathcal{Z}_k$ to calculate matching costs (C). The costs are utilized to sample new potential associations, which are then used to update the hypotheses in the new state ($\mathcal{H}_k$). Unlike other MHT frameworks, we model cell splits and design mitosis costs using the Erlang distribution. Association and mitosis costs are used to calculate hypothesis likelihoods l_k^h. In the shown example, the hypothesis state ($\mathcal{H}_k$) mostly contains two hypotheses (h), where the underline indicates the most likely hypothesis. In frame $k = 2$, a mitosis (black track) is reasonable and perhaps more likely than a randomly appearing cell or clutter (orange track). However, using posterior knowledge from frame $k = 5$ and our novel mitosis costs, our extended MHT framework corrects the falsely detected mitosis in frame $k = 2$ by identifying an implausible short life cycle for the blue cell in the black graph.

the corresponding cell centroid $\mathbf{O}_k^{\mathrm{C}} \in (-1,1)^{I_H \times I_W \times 2}$, and a motion offset $\mathbf{O}_k^{\mathrm{M}} \in (-1,1)^{I_H \times I_W \times 2}$ that estimates the motion, i.e., the offset relative to the preceding frame $\mathbf{I}_{k-1}$. The second step uses $\mathbf{D}_k$ and $\mathbf{O}_k^{\mathrm{C}}$ to cluster pixels into cell instances, where a cell instance j is represented by a two-dimensional centroid μ_k^j and pixels $\mathcal{M}_k^j$ in the ID map $\mathbf{M}_k$. Compared to other frameworks (Ronneberger et al., 2015), *EmbedTrack* uses the regressed motion offsets $\mathbf{O}_k^{\mathrm{M}}$ to warp the cell centroids to the previous frame, denoted as $\mu_{k-1|k}^j$. This is done by selecting the predicted offset $o_k^{\mathrm{M},j} \in \mathbf{O}_k^{\mathrm{M}}$ at the same two-dimensional position as the respective cell centroid μ_k^j and adding it to the current position. Thus, the estimated position in the previous frame is the sum of the current position and the motion regression: $\mu_{k-1|k}^j = \mu_k^j + o_k^{\mathrm{M},j}$. The output per frame can be described as a set of detected cell instances

$$\mathcal{Z}_k = \left\{ \left(\mu_k^j, \mu_{k-1|k}^j, \mathcal{M}_k^j \right) \right\}_{j=1}^{N_k^{\mathcal{Z}}}. \tag{1}$$

Finally, *EmbedTrack* solves the tracking task with nearest-neighbor matching between the warped cell positions $\mu_{k-1|k}$ in frame k and previous cell centroids μ_{k-1} in frame $k-1$. If multiple cells from k are matched to a single cell from $k-1$, it is classified as mitosis detection.

The lightweight method *EmbedTrack* is state-of-the-art, requires few hyperparameters, and can be learned end-to-end from existing data. However, the association is done only with local visual features and heuristic nearest-neighbor matching, which does not ensure consistency over long temporal periods. This can lead to errors, especially when cells appear similar in ambiguous image data. To overcome this limitation, we introduce a method to extend the single-point position and motion estimation to density estimations in subsequent sections.

3.2. Multi-Hypothesis Tracking with Random Finite Sets

In contrast to *EmbedTrack*, the well-known probabilistic MHT (Reid, 1979) framework follows the TbD paradigm, *i.e.* , it relies on precomputed cell instance detections, and finds the most likely assignment hypothesis between all detections over the entire image sequence. In this section, we present a realization of the MHT framework, namely the *Multi-Bernoulli Mixture* (MBM) tracker. We discuss the high-level concepts and formalize necessary details. A high-level visualization of the iterative MHT framework applied to an example image sequence is visualized in the lower part of Figure 3. Note that the visualization also includes contributions that are introduced in subsequent sections, highlighted in red. For the full framework, we refer to the literature or our code.

In the MBM framework, the spatio-temporal position of objects is described by 2D Gaussians. A potential object i is defined by its mean position μ_k^i, the corresponding covariance $\mathbf{\Sigma}_k^i$, and the probability r_k^i, which describes the likelihood that object i is indeed present. At a specific time k, there are N_k potentially existing objects. The correct tracking result is assumed to be one of H_k potential hypotheses h that form a multi-Bernoulli mixture:

$$\mathcal{H}_k = \left\{ \left(l_k^h, \left\{ \left(r_k^{i,h}, \mu_k^{i,h}, \mathbf{\Sigma}_k^{i,h} \right) \right\}_i^{N_k^h} \right) \right\}_h^{H_k}, \tag{2}$$

in which l_k^h is the log-likelihood of the hypothesis being correct.

To find the most likely tracking solution for an image sequence, the iterative MBM framework generates new hypotheses $\mathcal{H}_k$ at every time step k up to the last time step K, selecting the most likely hypothesis with the highest log-likelihood l_K^h. The input consists of a set of potential object detections $\mathcal{Z}_k$ for each time step, typically provided by an external detector (*e.g.* , *EmbedTrack*). A detection j is defined by its Gaussian position density $\mathcal{N}(\cdot; \mu_k^{j,\mathcal{Z}}, \mathbf{\Sigma}_k^{j,\mathcal{Z}})$, the probability of being clutter $\lambda_k^{\mathrm{C},j}$, and the probability of newly appearing in the scene ("born") $P^{\mathrm{B},j}$, *i.e.* , not existing in the image sequence before. Additionally, the MBM framework requires a model for the probability of detection $P_h^{\mathrm{D},i}$, which describes the likelihood that an object is detected by the detector, *i.e.* , is in $\mathcal{Z}_k$. Since P^{D} and P^{B} are difficult to model, they are typically set to constant values.

Given these inputs and starting with an empty initial hypothesis state $\mathcal{H}_0 = \{(l_0^1 = 1, \emptyset)\}$, $\mathcal{H}_K$ is derived by applying an iterative filter recursion to every hypothesis $h \in \mathcal{H}_k$ beginning at frame $k = 0$ and ending at $k = K$. The recursion, which is divided into prediction, sampling, update, and reduction, is elaborated next in detail. To illustrate the algorithm alongside the abstract formulation, we showcase a filter recursion applied to hypothesis h_4 in image $k = 3$, where it receives new detections $\mathcal{Z}_4$ from frame $k = 4$ in Figure 3 (black graph).

3.2.1. Predict

The first step is to estimate the expected position densities of all objects that are in $h \in \mathcal{H}_k$ in the next frame $k+1$. This is done with a motion model such as the linear Kalman filter (Kalman, 1960) that predicts the new state based on prior motion patterns. When the prediction is applied to all hypotheses in $\mathcal{H}_k$, the result is a warped state $\mathcal{H}_{k+1|k}$.

In our example hypothesis h_4, we model random-walk-like cell motion and increase the variances $\mathbf{\Sigma}_3^{6,h_4}$ and $\mathbf{\Sigma}_3^{7,h_4}$ of the existing objects 6 and 7.

3.2.2. Sample Association Hypotheses

Next, the predicted state $h \in \mathcal{H}_{k+1|k}$ is mapped to the new detection data $\mathcal{Z}_{k+1}$. A known object $i \in h$ is either represented by a detection $j \in \mathcal{Z}_{k+1}$ or remains undetected by the detector, for instance, if it moves outside the field of view. If a detection j does not represent an object i, it is either a newly appeared object or clutter. Note that mitosis, where two new detections are assigned to a former object, is not modeled in the vanilla MBM. However, to find the correct mapping, multiple likely assignments Ψ_k^h between new detections and known objects are possible. This is modeled by association costs $c_k^{j,i,h}$ between detections and objects, and unassignment costs $c_k^{j,\mathrm{u},h}$ for leaving detection j unassigned. Higher costs reflect a lower likelihood that an assignment is reasonable. The exact definition of these costs is not essential for this paper, but for completeness, they are formalized below using a score $n_k^{i,j,h}$ that evaluates the spatial similarity between the Gaussian-distributed positions of object i and detection j:

$$c_k^{j,i,h} = -\log\left(\left(1-\lambda_{k+1}^{\mathrm{C},j}\right)\frac{P^{\mathrm{D}} r_{k+1|k}^{i,h} n_k^{i,j,h}}{P^{\mathrm{B}} + \sum_{i'} P^{\mathrm{D}} r_{k+1|k}^{i',h} n_k^{i',j,h}}\right) \tag{3}$$

$$\text{and} \quad c_k^{j,\mathrm{u},h} = -\log\left(\left(1-\lambda_{k+1}^{\mathrm{C},j}\right) - \sum_i e^{-c_k^{j,i,h}}\right) \tag{4}$$

$$\text{with} \quad n_k^{i,j,h} = \mathcal{N}\left(\mu_{k+1|k}^{i,h}; \mu_{k+1}^{j,\mathcal{Z}}, \mathbf{\Sigma}_{k+1|k}^{i,h} + \mathbf{\Sigma}_{k+1}^{j,\mathcal{Z}}\right). \tag{5}$$

To finally sample assignments, the costs are arranged in a matrix

$$\mathbf{C}_k^h = \left[\ \mathbf{C}_k^{j,i,h} \ \middle|\ \mathrm{Diag}_\infty\left(\mathbf{c}_k^{j,\mathrm{u},h}\right)\ \right] \tag{6}$$

in which Diag_∞ maps the values to the diagonal of a square matrix with all other elements set to ∞. A row represents a detection and a column an object. Potential object to detection assignments are located in the left submatrix, unassigned detections in the right submatrix. A bijective sampling algorithm like Murty (Murty, 1968) or Gibbs (Geman & Geman, 1984) is applied to $\mathbf{C}_k^h$ to sample the $A_{\max}$ most likely assignments Ψ_k^h. The assignments minimize the total assignment costs and always assign a column to a row but at most a single row to a column. Thus, mitosis is not allowed in this formulation.

In our example hypothesis h_4 in Figure 3, multiple assignments are possible, *e.g.* , 8 to 6, 8 to 7, 8 is unassigned, and so further. Since the positions of detections $\mathcal{Z}_4$ and objects $\mathcal{H}_{4|3}$ are very similar, only the assignment that assigns 6 to 8 and 7 to 9 has low costs and is therefore likely.

3.2.3. Update

For each sampled assignment in Ψ_k^h, a new hypothesis h^* is created by refining the object states from h, under the assumption that the corresponding assignment is true. The object states from h are updated with their assigned detections and the motion model (e.g., the Kalman filter). Unassigned detections are added as new objects to h^*. The log-likelihood of the new hypothesis, $l_{k+1}^{h^*}$, is adjusted by adding the assignment costs corresponding to the assignment. Finally, the updated hypotheses are added to $\mathcal{H}_{k+1}$, contributing to the state density of the next frame.

For the example hypothesis h_4, we create the new hypothesis h_6 and update the positions of objects 6 and 7 by applying Kalman's update rule using the predicted positions and variances of objects 8 and 9. The log-likelihood of the new hypothesis is computed as $l_4^{h_6} = l_3^{h_4} + c_2^{8,6} + c_2^{9,7}$ and is added to $\mathcal{H}_4$.

3.2.4. Reduction

Since the number of hypotheses $H_k = |\mathcal{H}_k|$ grows exponentially, it becomes computationally expensive and impractical. To address this, hypotheses that describe the same state can be merged, and hypotheses with low probability (i.e., high l_k^h) are pruned until the number of hypotheses satisfies $H_k \leq H_{\max}$.

In our example, after the recursion, the two hypotheses h_6 and h_7 remain. They only differ in the age of object 8: it is 3 in the black graph of h_6 and unknown in the orange graph of h_7. Without utilizing this information in standard MHT frameworks, one of the hypotheses might be deleted to save computational time. However, in our approach presented later, we preserve this distinction to enable long-term lineage tracking.

The primary advantage of MBM/MHT lies in its ability to re-evaluate the likelihoods of hypotheses a posteriori. Hypotheses that initially appeared likely can be rejected if they lead to unlikely outcomes in subsequent frames. MBM and other MHT-based approaches are effective because they provide a globally optimal tracking solution, incorporating all available detection information. However, the framework does not model cell division and heavily depends on a strong motion model, which is challenging to define for random-walk-like cell movements in time-lapse sequences. In later sections of this paper, we introduce a strong implicit motion model and extend the standard MBM framework to model cell division and ensure long-term consistency in cell tracking. It is important to note that the core of our contribution can be similarly applied to other MHT-based frameworks, such as (Granström et al., 2018; Rezatofighi et al., 2015). For simplicity, we use the term MHT instead of specifically referring to MBM.

4. Uncertainty in Tracking-by-Regression

In this section, we introduce our first contribution and estimate the uncertainty of regression frameworks, exemplarily applied to *EmbedTrack*, to derive continuous spatial position and motion estimation distributions that are necessary for our extended association strategy in MHT frameworks (Section 5). We want to emphasize that our concepts apply to arbitrary TbR frameworks that estimate motion. Our contributions applied to *EmbedTrack* are visualized in Figure 3.

4.1. Test-Time Augmentation for Motion Regression

Test-Time Augmentation is a widely used strategy to reduce data noise during inference by applying a set of transformations, $\mathcal{T}$, to the input image $\mathbf{I}$ and averaging the inferred estimations (Wang et al., 2019). For example, *EmbedTrack* utilizes rotations of $0°$, $90°$, $180°$, and $270°$, as well as reflections, resulting in a set of $|\mathcal{T}| = 8$ Euclidean transformations that are equally applied to both input images $\mathbf{I}_{k-1}$ and $\mathbf{I}_k$.

While this strategy tackles geometrical variances, motion regression networks that aim to estimate the position of the cell instances in the previous frame (i.e., the motion), based on visual cues, suffer from two problems that are not addressed by standard augmentations: Cell appearance is similar within a population and can change between consecutive frames, as highlighted in the upper row of Figure 1. The appearance of the cell marked in blue is more or less static, but the red one changes shape and pixel intensities. Since reliable re-identification is sometimes impossible even for humans, we suspect regression networks to perform a heuristic and random assignment, such as a simple nearest-neighbor matching, in those cases.

To overcome this heuristic, we add spatial transformations to $\mathcal{T}$ that keep $\mathbf{I}_k$ unchanged but shift $\mathbf{I}_{k-1}$. The shift should induce slight variances in the spatial arrangement without creating implausible motions of cells. Thus, we shift the images in all vertical and horizontal directions by the average cell radius $\overline{r}_{\text{Cell}}$, which can be extracted from training data or with a segmentation framework. Using training data, one can approximate the average cell radius by modeling cell shapes as perfectly round circles and employing the circle areas

$$\overline{r}_{\text{Cell}} = \frac{1}{\sum_{k=1}^{K^{\text{GT}}} N_k^{\text{GT}}} \sum_{k=1}^{K^{\text{GT}}} \sum_{n=1}^{N_k^{\text{GT}}} \sqrt{\frac{1}{\pi} \|\mathcal{M}_k^{n,\text{GT}}\|_1}, \tag{7}$$

in which K^{GT} is the number of ground truth images, N_k^{GT} denotes the number of ground truth segmentations in frame k, and $\mathcal{M}_k^n$ is the respective binary segmentation map of cell n. It's worth noting that this approximation is only useful if training data is available and has the same cell size distribution as the test data. If no data is available, the ground truth segmentation $\mathcal{M}_k^{n,\text{GT}}$ can be substituted with the results of high-performing segmentation frameworks, as, for example, shown in Equation (1). In this work, we use training data that is available for the respective datasets employed in our experiments presented in Section 6. The shift parameter $\overline{r}_{\text{Cell}}$ is static for the entire tracking sequence and therefore not adaptive to changing cell sizes in the sequence.

The resulting set of transformations, $\mathcal{T}'$, consists of the original transformations as well as shifted transformations in all spatial directions with the shift $\overline{r}_{\text{Cell}}$, such that $|\mathcal{T}'| = 5 \cdot |\mathcal{T}|$. Using the new shifts, all motion estimations should be similar when visual cues are unambiguous but may vary in uncertain situations. The variance of the prediction reflects ambiguity and uncertainty and is used in the next section.

4.2. Uncertainty in Centroid and Motion Distributions.

While averaging test-time augmentations reduces the influence of geometrical variances in tasks like instance segmentation, simple averaging is not useful for multi-modal distributions in centroid or motion regression. The networks implicitly try to find the object location in the (subsequent) image based on visual cues. If multiple similar cell instances exist, averaging the respective predictions leads to a random average prediction that may point to a position with no cell. Thus, we transform the uncertainty revealed by shifted test-time augmentation applied to *EmbedTrack* into continuous spatial distributions.

Before we apply our contribution, the estimated centroid and motion of a detection j are described by discrete 2D positions, μ_k^j and $\mu_{k-1|k}^j$. To represent spatial estimation uncertainties, we describe the detection centroids and motion with multivariate Gaussian densities $\mathcal{N}\big(\cdot; \mu_k^{j,\mathcal{Z}}, \mathbf{\Sigma}_k^{j,\mathcal{Z}}\big)$ and $\mathcal{N}\big(\cdot; \mu_{k-1|k}^{j,\mathcal{Z}}, \mathbf{\Sigma}_{k-1|k}^{j,\mathcal{Z}}\big)$ with respective mean and variance parameters for specific detection instances. To get the parameters, we first calculate pixel-wise (co)variances of the centroid and motion offsets, denoted as $\mathbf{O}_k^{\text{C},\mathbf{\Sigma}} \in \mathbb{R}_+^{H \times W \times 2 \times 2}$ and $\mathbf{O}_k^{\text{M},\mathbf{\Sigma}} \in \mathbb{R}_+^{H \times W \times 2 \times 2}$. Then, every pixel $\mathbf{p} \in \mathcal{M}_k^j$ that belongs to a specific cell j defines a weighted Gaussian with the weights $\mathbf{D}_{k,\mathbf{p}}$ and the parameters $(\mathbf{O}_{k,\mathbf{p}}^{\text{M}}, \mathbf{O}_{k,\mathbf{p}}^{\text{M},\mathbf{\Sigma}})$ (the index $\mathbf{p}$ denotes the matrix value at the respective pixel position). The Gaussians can be seen as a Gaussian mixture (GM) and merged according to (Crouse et al., 2011), such that the spatial distribution parameters of a cell instance j are defined as

$$\Big\{ \big(\mathbf{D}_{k,\mathbf{p}}, \mathbf{O}_{k,\mathbf{p}}^{\text{C}}, \mathbf{O}_{k,\mathbf{p}}^{\text{C},\mathbf{\Sigma}}\big) \mid \mathbf{p} \in \mathcal{M}_k^j \Big\} \xrightarrow{\text{Merge}} \big(\mu_k^{j,\mathcal{Z}}, \mathbf{\Sigma}_k^{j,\mathcal{Z}}\big) \tag{8}$$

$$\left\{ \left(\mathbf{D}_{k,\mathbf{p}}, \mathbf{O}^{\mathrm{M}}_{k,\mathbf{p}}, \mathbf{O}^{\mathrm{M},\mathbf{\Sigma}}_{k,\mathbf{p}} \right) \mid \mathbf{p} \in \mathcal{M}^j_k \right\} \xmapsto{\text{Merge}} \left(\mu^{j,\mathcal{Z}}_{k-1|k}, \mathbf{\Sigma}^{j,\mathcal{Z}}_{k-1|k} \right) \quad (9)$$

Moreover, we estimate the probability that detection j is a false detection (*a.k.a.* clutter). The clutter probability $\lambda^{\mathrm{C},j}_k$ is defined using the inverted segmentation score $1 - \mathbf{D}_{k,\mathbf{p}}$ at the centroid pixel $\mathbf{p} = \mu^{j,\mathcal{Z}}_k$.

With our contribution, the new output of *EmbedTrack* extends Equation (1) to

$$\mathcal{Z}_k = \left\{ \left(\lambda^{\mathrm{C},j}_k, \mu^{j,\mathcal{Z}}_k, \mathbf{\Sigma}^{j,\mathcal{Z}}_k, \mu^{j,\mathcal{Z}}_{k-1|k}, \mathbf{\Sigma}^{j,\mathcal{Z}}_{k-1|k}, \mathcal{M}^j_k \right) \right\}_{j=1}^{N^{\mathcal{Z}}_k}. \quad (10)$$

The additional variance indicates situations in which *EmbedTrack* may causes errors and and needs to be corrected.

5. Mitosis-aware Multi-Hypothesis Tracking

Our first contribution allows strong neural regression models to estimate position densities that are required by MHT frameworks. However, the presented MHT tracker suffers from several drawbacks for cell tracking: There is no accurate motion model due to the often unpredictable cell motion in time-lapse videos, and the bijective one-to-one association does not allow modeling mitosis. Thus, the following sections introduce A) our novel implicit motion model for MHT trackers based on our uncertainty-aware regression system, and B) a model that also accounts for mitosis and enables long temporal consistency with our novel cell cycle-preserving mitosis costs. Furthermore, without going into further detail, we replace the typically handcrafted probability of clutter with our uncertainty-based $\lambda^{\mathrm{C},j}_k$ introduced in the previous section. An example MHT filter recursion with our novelties highlighted in red is visualized in Figure 3.

5.1. Implicit Motion Model

The MHT framework predicts the motion of objects from the current frame k to the subsequent frame $k+1$. The estimated positions are then matched to the positions of new detections and updated according to the Kalman filter. Instead of only using the spatial position of detection proposals, we propose the additional use of the motion estimation output from our uncertainty-aware regression framework, as given in Equation (10). Motion regression is an implicitly learned, solid appearance-based motion model and replaces the linear and spatial Kalman prediction step. The association costs can be calculated directly between the previous object positions and the estimated motion prediction of the detections. To achieve this, Equation (5) needs to be modified to

$$n^{i,j,h}_k = \mathcal{N}\left(\mu^{i,h}_k; \mu^{j,\mathcal{Z}}_{k|k+1}, \mathbf{\Sigma}^{i,h}_k + \mathbf{\Sigma}^{j,\mathcal{Z}}_{k|k+1} \right). \quad (11)$$

Consequently, we also do not apply the Kalman filter during the update step and instead directly use the estimation of the usually high-performing centroid estimation model to update the object positions with $\mu^{i,h}_{k+1} = \mu^{j,\mathcal{Z}}_{k+1}$ and $\mathbf{\Sigma}^{i,h}_{k+1} = \mathbf{\Sigma}^{j,\mathcal{Z}}_{k+1}$. Objects that are not assigned to a detection keep their mean but have their covariance increased as $\mathbf{\Sigma}^{i,h}_{k+1|k+1} = \mathbf{\Sigma}^{i,h}_k + \bar{\mathbf{\Sigma}}$ by the mean cell motion per frame, $\bar{\mathbf{\Sigma}}$. Using our regressed motion estimation densities leads to large estimation variances only for cells with high aleatoric uncertainty.

5.2. Mitosis-aware Association Costs

Due to the bijective assignment, the vanilla MHT does not allow mitosis, where object i from $h \in \mathcal{H}_{k+1|k}$ has an association with two detections, j_1 and j_2, from measurement $\mathcal{Z}_{k+1}$. This limitation arises from the design of the cost matrix $\mathbf{C}^h_k$ in Equation (6) for the assignment problem, where only one association per row and column is allowed. To enable the assignment of j_1 and j_2 to an object i during cell mitosis, we extend the cost matrix $\mathbf{C}^h_k$ by adding a submatrix $\mathbf{C}'^{j,i,h}_k$ to the right, such that

$$\mathbf{C}^h_k = \left[\; \mathbf{C}^{j,i,h}_k \;\middle|\; \mathrm{Diag}_\infty\left(\mathbf{c}^{j,\mathrm{u},h}_k \right) \;\middle|\; \mathbf{C}'^{j,i,h}_k = \mathbf{C}^{j,i,h}_k \; \right]. \quad (12)$$

In the new cost matrix, a detection is represented in a single row, while objects are represented in two columns. If the costs $c^{j_1,i,h}_k$ and $c^{j_2,i,h}_k$ are relatively small, solving the assignment problem leads to assignments of j_1 in $\mathbf{C}^{j,i,h}_k$ and j_2 in $\mathbf{C}'^{j,i,h}_k$, or vice versa, such that both are assigned to i.

The new cost matrix enables the MHT framework to model mitosis, but it does not explicitly incorporate biological knowledge about the cell life cycle and mitosis. Thus, the likelihood of mitotic events is not assessed in the rating l^h_k of the corresponding hypotheses. To close this gap, we reformulate $\mathbf{C}'^{j,i,h}_k$ and add biologically inspired costs such that $\mathbf{C}'^{j,i,h}_k = \mathbf{C}^{j,i,h}_k + \mathbf{C}^{\mathrm{M},i,h}_k$ to guide the mitosis detection. A value $c^{\mathrm{M},i,h}_k$ in column i reflects the probability that cell i from hypothesis h should not split at this moment. We set

$$c^{\mathrm{M},i,h}_k = \begin{cases} -\log\left(\int_{-\infty}^{\mathrm{Age}(i)} \mathrm{Erlang}_{\alpha,\beta}(t) \mathrm{d}t \right) & \text{if known} \\ 0 & \text{else} \end{cases} \quad (13)$$

with the current lifetime $\mathrm{Age}(i)$ of the cell (which may also be unknown) and the cumulative $\mathrm{Erlang}_{\alpha,\beta}(t)$ distribution, which describes the expected lifetime distribution of cells (Yates et al., 2017). Adding these costs penalizes hypotheses that imply implausibly short cell life cycles. For example, in Figure 3, even if the black hypothesis is more likely before frame 5, mitosis in frame 5 causes high association costs. This leads the orange track, which does not involve mitosis in frame 2, to appear more likely a-posteriori. The Erlang distribution is parameterized by α and β, which

Table 2. Benchmark on biological measures on the test data following the official Cell Tracking Challenge (Maška et al., 2023). Colored numbers represent the performance of the official top #k state-of-the-art methods. Applied to identical input detections, the baseline compares heuristic association of EmbedTrack (Löffler & Mikut, 2022) and *Trackastra* (Gallusser & Weigert, 2024) transformer-based association to our association approach using . Numbers highlighted in **bold** indicate that we surpass other the association methods on identical inputs and underlined numbers that our approach is the new state-of-the-art over all participating methods.

	[%] ↑	BF-C2DL-HSC	BF-C2DL-MuSC	DIC-C2DH-HeLa	Fluo-C2DL-MSC	Fluo-N2DH-GOWT1	Fluo-N2DL-HeLa	PhC-C2DH-U373	PhC-C2DL-PSC	Fluo-N2DH-SIM+
CT	Top #1	5.94	5.32	24.06	29.29	36.58	67.45	57.33	17.13	59.58
	Top #2	5.4	1.65	16.88	20.29	34.13	63.34	50.40	16.79	58.53
	Top #3	4.42	1.53	12.20	19.77	31.63	59.59	50.22	14.23	53.09
	Cell DINO	N/A	N/A	24.06	19.77	31.63	N/A	26.19	N/A	34.17
	Baseline	5.94	1.00	12.20	6.79	27.26	58.76	50.22	14.23	59.88
	Trackastra	16.23	1.73	20.54	14.29	25.57	64.91	30.08	17.30	61.33
	Ours	**20.59**	**2.32**	**29.19**	**20.75**	**29.10**	**65.42**	**56.94**	**17.68**	56.09
TF	Top #1	62.26	68.46	87.14	75.29	94.16	98.05	100.0	84.16	93.66
	Top #2	59.62	68.16	80.93	74.96	93.57	96.95	99.85	83.65	92.72
	Top #3	57.14	63.49	75.30	68.12	89.76	96.85	99.82	82.65	91.71
	Cell DINO	N/A	N/A	87.14	66.11	89.76	N/A	92.15	N/A	90.03
	Baseline	62.26	68.16	75.30	59.59	84.00	96.95	97.07	82.65	93.66
	Trackastra	75.85	68.73	81.02	69.14	84.93	97.94	99.82	85.99	93.82
	Ours	**79.88**	**75.67**	**82.08**	65.44	**88.34**	97.21	94.20	84.18	**94.74**
BC(i)	Top #1	44.05	65.10	N/A	N/A	N/A	88.21	N/A	60.04	92.16
	Top #2	32.46	55.07	N/A	N/A	N/A	88.12	N/A	57.59	91.79
	Top #3	27.68	39.20	N/A	N/A	N/A	81.10	N/A	53.58	89.67
	Cell DINO	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	69.59
	Baseline	32.46	24.68	N/A	N/A	N/A	77.09	N/A	48.17	92.16
	Trackastra	56.77	30.59	N/A	N/A	N/A	85.76	N/A	54.04	91.69
	Ours	**58.82**	**60.40**	N/A	N/A	N/A	85.05	N/A	**59.29**	83.07
CCA	Top #1	56.33	85.18	N/A	N/A	N/A	93.12	N/A	85.34	94.76
	Top #2	51.79	34.02	N/A	N/A	N/A	91.37	N/A	64.89	91.71
	Top #3	43.33	25.37	N/A	N/A	N/A	89.71	N/A	63.49	90.52
	Cell DINO	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	70.92
	Baseline	12.23	11.82	N/A	N/A	N/A	62.75	N/A	34.75	85.29
	Trackastra	34.35	15.04	N/A	N/A	N/A	77.71	N/A	49.68	89.61
	Ours	**69.63**	**36.05**	N/A	N/A	N/A	**85.10**	N/A	**57.65**	**90.15**

Method

- KIT-GE (1) (Stegmaier et al., 2012)
- KIT-GE (2) (Löffler et al., 2021)
- KIT-GE (3) (Scherr et al., 2020)
- KIT-GE (4) (Löffler & Mikut, 2022)
- KTH-SE (1) (Magnusson et al., 2015)
- KTH-SE (1*) (Magnusson et al., 2015)
- KTH-SE (3) (Magnusson et al., 2015)
- KTH-SE (5) (Magnusson et al., 2015)
- BGU-IL (1) (Arbelle & Raviv, 2019)
- BGU-IL (5) (Ben-Haim & Raviv, 2022)
- HD-GE (BMCV) (1)
- HD-GE (IWR) (Schiegg et al., 2013)
- FR-GE (2) (Ronneberger et al., 2015)
- HKI-GE (5) (Belyaev et al., 2021)
- THU-CN (2) (Hu et al., 2021)
- TUG-AT (Payer et al., 2019)
- USYD-AU (Chen et al., 2021)
- ND-US (1)
- DREX-US
- IMCB-SG (1)
- UVA-NL (Gupta et al., 2019)
- HIT-CN (1) (Zhou et al., 2019)
- CVUT-CZ (Sixta et al., 2020)
- NUDT-CN (Liao et al., 2024)

can be approximated from the data by regressing the exponential proliferation rate, as done, for instance, in (Paul et al., 2024). For short sequences of length K that rarely contain cell splits and do not allow reasonable approximations, we set $\alpha = K$ and $\beta = \frac{1}{K}$ to penalize cell splits with relatively high costs. The novel cost matrix, together with mitosis costs, allows modeling cell proliferation and explicitly uses statistical biological knowledge to identify the most likely hypothesis in MHT frameworks.

6. Experiments

This section presents experimental results analyzing our uncertainty estimation and tracking framework. We evaluated our method on nine publicly available and competitive datasets provided by the *Cell Tracking Challenge* (CTC) (Maška et al., 2023), that cover a wide range of cell types and modalities as summarized in table 3. The data consists of mouse muscle stem cells (BF-C2DL-HSC/MuSC, Fluo-N2DH-GOWT1), HeLa cells (DIC-C2DL-HeLa, Fluo-N2DL-HeLa), rat mesenchymal stem cells (Fluo-C2DL-MSC), glioblastoma-astrocytoma U373 cells (PhC-C2DH-U373), pancreatic stem cells (PhC-C2DL-PSC), and simulated HL60 nuclei (Flup-N2DH-SIM+). They are captured in short (*e.g.* , Fluo-C2DL-MSC) and long microscopic video sequences with different distinction of proliferation trees (*e.g.* , BF-C2DL-HSC). To provide an intuition about the differing complexity with respect to the tracking task, Table 3 presents metadata about the cell culture derived from our results. This clearly shows that the mouse stem cell data (BF-C2DL-HSC and BF-C2DL-MuSC) have the longest proliferation trees, beginning with a few initial cells and growing exponentially to hundreds or thousands of cells in the colony. The lineage tree of a sequence from BF-C2DL-HSC and BF-C2DL-MuSC compared to the much smaller PhC-C2DH-U373 is shown in Figure 2. As such, BF-C2DL-HSC and BF-C2DL-MuSC are the most important datasets to evaluate the long-term tracking capabilities of our method.

The CTC undisclosed ground truth for the test data used for benchmarking and only publishes a limited set of evaluation measures. Since our method aims to enhance long-term consistency, which is essential for monitoring entire cell life cycles, we assess its performance using biologically relevant metrics (Ulman et al., 2017) that are most suitable for evaluating long-term tracking. Specifically, we report *Complete Tracks* (CT), *Track Fractions* (TF), *Branching Correctness* (BC(i)), and *Cell Cycle Accuracy* (CCA). While CT indicates the fraction of tracks that are fully reconstructed without error, TF reports the average fraction of a track that is continuously reconstructed correctly. For evaluating mitosis detection, BC(i) indicates the fraction of correctly

Table 3. Additional information to the test datsets presented in Table 2. Each dataset includes two video sequences for evaluation. The number of cell instances describes the total number of cell detections captured by our method, that are then clustered to a distinct number of trajectories. The average cell size in pixel is the area of cell detections, together with the average cell motion per frame in pixel and relative cell radian, the initial cells the number of cells in the first frame, and cell splits the branching events. From the technical perspective, the data differs in the used microscope and lens, pixel grid size, image resolution, number of frames and the time elapsed between two consecutive frames.

	BF-C2DL -HSC	**BF-C2DL -MuSC**	**DIC-C2DH -HeLa**	**Fluo-C2DL -MSC**	**Fluo-N2DH -GOWT1(Bártová et al., 2011)**	**Fluo-N2DL -HeLa(Neumann et al., 2010)**	**PhC-C2DH -U373**	**PhC-C2DL -PSC(Rapoport et al., 2011)**	**Fluo-N2DH -SIM+(Svoboda & Ulman, 2017)**
Cell Instances	185475	14830	2631	723	5243	31099	1244	146902	10077
Trajectories	1475	330	105	56	107	943	24	4211	263
Cell Splits	652	130	15	6	10	309	3	1638	98
Initial Cells	4	2	19	13	56	164	11	140	43
Avg. Cell Size [Pixel]	299	889	21151	9768	35895	569	4240	130	1728
Avg. Motion [Pixel]	5 (46%)	12 (70%)	7 (13%)	21 (37%)	3 (9%)	2 (17%)	4 (10%)	1 (14%)	4 (16%)
Frames	3528	2752	230	96	184	184	230	600	248
Resolution [Pixel]	1010x1010	1036x1070	512x512	832x992 782x1200	1024x1024	700x1100	520x696	576x720	718x660 790x664
Microscope	Zeiss PALM/ AxioObserver Z1	Zeiss PALM/ AxioObserver Z1	Zeiss LSM 510 Meta	PerkinElmer UltraVIEW ERS	Leica TCS SP5	Olympus IX81	Nikon	Olympus ix-81	Zeiss Axiovert 100S Micromax 1300-YHS
Objective Lens	Plan-Neofluar 10x/0.30 Ph1	Plan-Neofluar 10x/0.30 Ph1	Plan-Apochromat 63x/1.4 (oil)	Plan-Neofluar 10x/0.3 (Plan-Apo 20x/0.75)	Plan-Apochromat 63x/1.4 (oil)	Plan 10x/0.4	Plan Fluor DLL 20x/0.5	UPLFLN 4XPH	Plan-Apochromat 40x/1.3 (oil)
Pixel Size [micron]	0.645x0.645	0.645x0.645	0.19x0.19	0.3x0.3	0.24x0.24	0.645x0.645	0.65x0.65	1.6x1.6	0.125x0.125
Time Step [min]	5	5	10	20	5	30	15	10	29

Table 4. Impact of augmentations on the variance of motion estimation adding shifts $\mathcal{T}'$ to standard augmentations $\mathcal{T}_0$. In addition to our shift, we halfed ($\mathcal{T}'_{\text{Half}}$) and doubled ($\mathcal{T}'_{\text{Double}}$) the amount of pixels. We present the average motion estimation per frame for $\mathcal{T}_0$ in pixels and otherwise the relative amplification. Also, the impact to CHOTA is visualized.

Mean Motion [Pixel]		Test-Time Shift			
		$\mathcal{T}_0$	$\mathcal{T}'_{\text{Half}}$	$\mathcal{T}'_{\textbf{Ours}}$	$\mathcal{T}'_{\text{Double}}$
Dataset	BF-C2DL-HSC	1.55	×1.33	×2.97	×6.32
	BF-C2DL-MuSC	9.61	×1.04	×1.19	×1.76
	DIC-C2DH-HeLa	14.13	×1.31	×2.58	×4.76
	Fluo-C2DL-MSC	30.48	×1.05	×1.28	×2.09
	Fluo-N2DH-GOWT1	3.88	×1.09	×1.87	×9.87
	Fluo-N2DH-SIM+	3.82	×1.07	×1.40	×7.50
	Fluo-N2DL-HeLa	2.21	×1.08	×1.92	×7.24
	PhC-C2DH-U373	6.32	×1.08	×1.57	×5.39
	PhC-C2DL-PSC	1.45	×1.07	×1.40	×3.07
CHOTA ↑ [%]		Test-Time Shift			
		$\mathcal{T}_0$	$\mathcal{T}'_{\text{Half}}$	$\mathcal{T}'_{\textbf{Ours}}$	$\mathcal{T}'_{\text{Double}}$
Dataset	BF-C2DL-HSC	73.27	75.40	**76.75**	73.98
	BF-C2DL-MuSC	81.36	81.26	**82.24**	80.25
	DIC-C2DH-HeLa	90.52	90.51	**91.78**	87.96
	Fluo-C2DL-MSC	83.21	83.19	**86.74**	80.69
	Fluo-N2DH-GOWT1	96.80	96.80	96.81	**97.07**
	Fluo-N2DH-SIM+	96.29	**96.37**	**96.37**	96.06
	Fluo-N2DL-HeLa	92.45	92.57	**92.61**	92.52
	PhC-C2DH-U373	**92.28**	**92.28**	**92.28**	92.05
	PhC-C2DL-PSC	82.50	82.75	**83.07**	82.74

detected cell splits, and CCA measures the overlap between predicted and ground truth life cycle distributions. Furthermore, to gain in-depth insights into long-term tracking, we use the CHOTA metric (Kaiser et al., 2024) in our ablation studies, which are performed on the CTC training/validation data with disclosed ground truth. CHOTA evaluates long-term tracking capability by rating each association based on its impact on the entire tracking result. Qualitatively, CHOTA quantifies the fraction of connected descendant and ascendant cell detections in the ground truth proliferation tree that are also connected in the tracking result for every cell detection. Therefore, CHOTA is the most comprehensive tracking metric used in this paper, as it includes both short- and long-term relationships between cells.

To ensure a fair comparison, we employ the code and pretrained models of *EmbedTrack* (Löffler & Mikut, 2022) without any modification or re-training, applying only our uncertainty estimation strategy. We use *EmbedTrack's* preprocessing during inference, generating overlapping crops of size 256x256 (512x512 for Fluo-C2DL-MSC) and applying min-max normalization to the range $[0, 1]$ using the 1% and 99% percentiles per crop. Thus, we refer to *EmbedTrack* as our baseline. If not stated otherwise, our extended MHT tracker is implemented using hyperparameters $A_{\text{max}} = 7$, $H_{\text{max}} = 150$. After tracking, we remove very small tracks and interpolate at gaps. The code is publicly available (see page 2). To precisely compare our association strategy with the current state-of-the-art, we apply *Trackastra* (Gallusser & Weigert, 2024) to the same input detections used by our method, derived from *EmbedTrack*. We use the official implementation of *Trackastra* that can be found here[1] without modifications. Inference was performed in mode *greedy* and with the provided pretrained model *general_2d* that is trained on the CTC data as used in our experiments.

6.1. Quantitative Results

Our method is specifically designed to enhance tracking results in long and complex scenarios by effectively resolving long-term conflicts through the introduction of mitosis costs. To evaluate this capability, we present the results obtained on nine diverse datasets, assessed using the CTC evaluation server. The results are summarized in Table 2, which presents the top three leading benchmark methods for each dataset in the CTC, as well as a comparison of our method to the baseline *EmbedTrack*, differing only in the association strategy. Additionally, we compare our as-

[1] https://github.com/weigertlab/trackastra

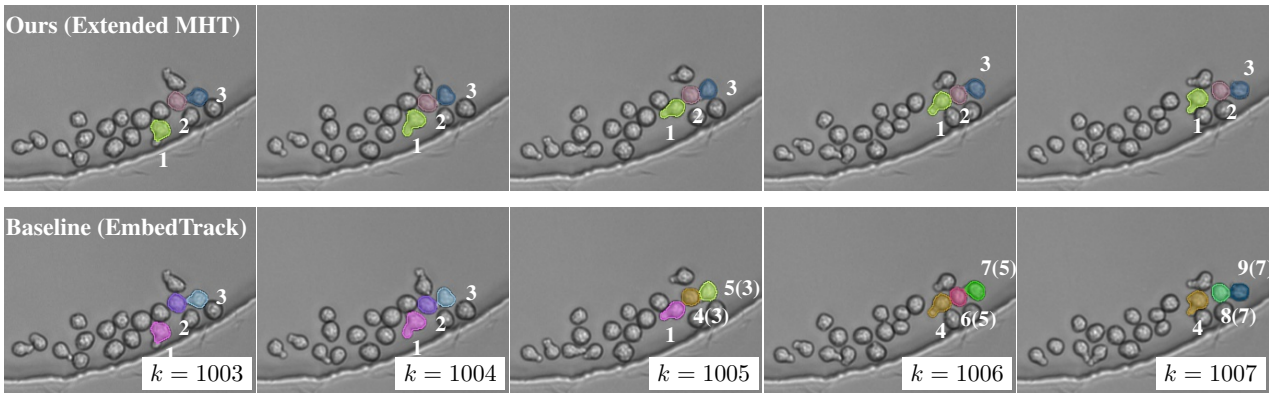

Figure 4. **Superior Case.** The image sequence illustrates a scenario in which our method achieves significant improvements over the baseline (BF-C2DL-HSC). The cells are densely packed, similar in appearance, and the sequence is relatively long, containing extensive lineage tree information. An abrupt increase in the motion of cell 1 at frame $k = 1005$ displaces cells 2 and 3. While our tracker successfully reconstructs the correct trajectories, the baseline's local assignment strategy results in several implausible cell splits. White numbers indicate cell IDs, and braces show the parent ID when available in the displayed sequence. Cells without label are not of interest and can be considered as tracked correctly.

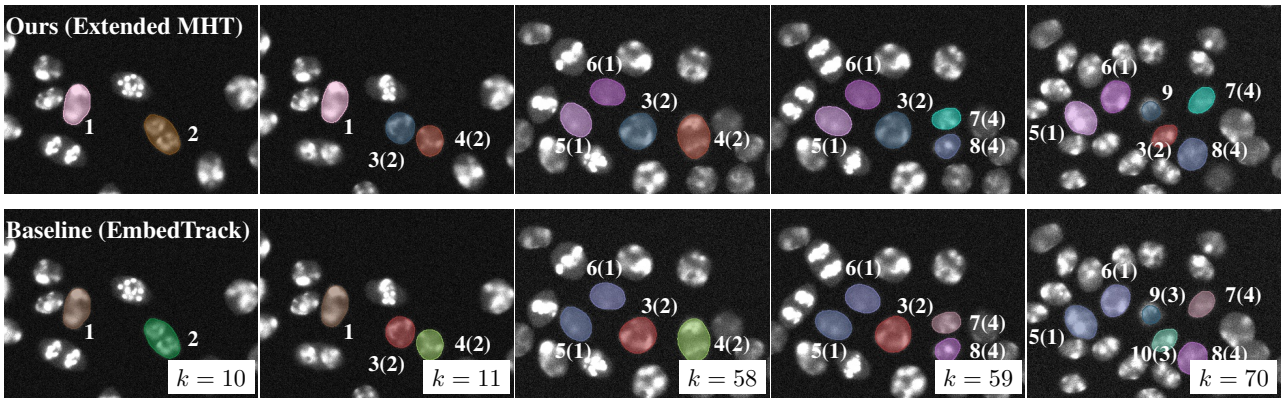

Figure 5. **Failure Case**. The image sequence showcases a scenario in which our method struggles and induces errors (Fluo-N2DH-SIM+). The image sequence is relatively short, with only a few cell cycles. As a result, the modeled mitosis costs based on sequence statistics do not accurately reflect the true lifetime distribution. In frame $k = 59$, cell 3 divides into cells 9 and 10 (see baseline). Our tracker discards this mitosis due to high mitosis costs and instead spawns a new cell 9 (see ours). In contrast, the baseline's local assignment strategy benefits from the distinct appearance of the cells and their limited movement, successfully detecting the correct mitosis event. White numbers indicate cell IDs, and braces show the parent ID when available in the displayed sequence. Cells without label are not of interest and can be considered as tracked correctly.

sociation strategy to the latest trends in cell tracking by showing results achieved by transformer-based methods, *Cell DINO* (Liao et al., 2024) (if available) and *Trackastra* (Gallusser & Weigert, 2024). Note that *Trackastra* is applied to same input detections as our method to enable a fair comparison.

The most meaningful comparison to evaluate the association strategy is against the baseline and *Trackastra*, which use the same input detections. Our method shows a substantial improvement in metrics, particularly for complex datasets, with improvements of up to a factor of $\times 5.7$ (CCA, BF-C2DL-HSC). As the data complexity decreases, the improvement diminishes, aligning with our method's design, as smaller datasets rarely include complete cell life cycles. Our method does not improve metrics on relatively short sequences such as Fluo-N2DH-SIM+ or PhC-C2DH-U373. This behavior is expected because the former includes only a limited number of mitotic events, and the latter contains almost no mitotic events that could benefit from our extended association strategy.

Our method emerges as the new state-of-the-art in 5 out of 9 datasets on the biological metrics benchmark, outperforming all other competitors in the challenge. Since our association strategy relies on the predictive capabilities

of *EmbedTrack*, the method only performs less well when *EmbedTrack* itself has a large performance gap compared to the current state-of-the-art. For example, in BF-C2DL-MuSC, *EmbedTrack* suffers from many over- and under-segmentations. This dependency on detection quality is clearly visible, as our method performs best on BF-C2DL-MuSC when better input detections are used, as evaluated in the next section.

Comparing our method to the latest trends, it is evident that we outperform transformer-based models in complex scenarios with large proliferation trees, such as BF-C2DL-HSC. This may be due to the fact that end-to-end neural networks like these are not designed to model high-level biological information. However, the association performance of transformer-based models is superior on smaller datasets where proliferation and mitosis are limited, such as *Cell DINO* on Fluo-N2DH-GOWT1. These results might also be partially attributed to the better detection quality of *Cell DINO*.

Another noteworthy observation is the discrepancy between the reported biological metrics in Table 2 and the technical metrics reported here[2]. Both our approach and the baseline achieve technical metrics that are close to optimal, with differences largely attributable to noise. This underscores the concern stated in Section 1 that technical metrics often do not reflect biological aspects. Our method addresses this issue by prioritizing long-term consistency by effectively resolving mitosis errors. These errors, while having a minimal impact on technical metrics, significantly influence biologically relevant metrics. In addition to the discussed benchmark metrics, Table 5 shows additional metrics from the *py-ctcmetrics* framework (Kaiser et al., 2024) that we applied to the respective train and validation datasets with disclosed ground truth data. The metrics help practical users to assess our method. Videos of our reported tracking results can be found here[3].

6.2. ISBI Challenge - Linking only

In addition to the evaluation on the CTC benchmark, we applied our method to the seventh ISBI Challenge in the linking-only track[4]. In this challenge, the organizers provided pre-computed and potentially faulty cell detections that needed to be associated. To satisfy this requirement, we replaced the detections from *EmbedTrack* in Equation (1) with the provided ones, while keeping the rest of the system unchanged.

Our first contribution in *EmbedTrack* may be negatively influenced by the data induction in Equation (1). However, our second contribution in the MHT framework significantly improves performance, surpassing all other participants in long sequences with substantial proliferation. Our method demonstrates a notable advantage on the long and densely populated BF-C2DL-HSC/MuSC sequences, with an improvement of approximately $+3\%$. This aligns with the observations in Table 2. However, this advantage does not manifest in shorter sequences without significant proliferation. The full challenge results can be found here[5], where we are named *LUH-GE*.

6.3. Uncertainty Estimation

During test-time augmentation, we apply transformations $\mathcal{T}'$ to shift the image $\mathbf{I}_{k-1}$. This practice helps to increase the variance in uncertain predictions. The impact of various augmentations $\mathcal{T}'$ is illustrated in Figure 1, where shifts of 0, 1, 4, and 8 pixels are applied to an image containing a crowded cell population with approximately 20×20 pixels per cell.

There are two cells visualized: a cell with anappearance change (red) and an easy-to-reidentify cell (blue). When only applying the standard $\mathcal{T}$, we observe small variances in both motion estimations in $\mathbf{\Sigma}^{\mathcal{Z}}_{k-1|k}$, indicating low uncertainty. This confirms the assumption that the default strategy leads to very certain predictions in uncertain environments. However, applying a small shift of 1 pixel ($\mathcal{T}'_1$) leads to a small increase in uncertainty, while a shift of 4 pixels ($\mathcal{T}'_4$) results in significantly growing variances for the red cell. This allows multiple plausible associations during tracking. Lastly, applying $\mathcal{T}'_8$ leads to drastically increasing variance for the uncertain cell, but is still small for the certain blue cell.

Table 4 presents the average standard deviation of motion in pixels to quantify the impact of shifts on different datasets. We compare different augmentation settings with no shift ($\mathcal{T}_0$), our shift ($\mathcal{T}'_{\textbf{Ours}}$) determined by the average cell radian, and shifts with half ($\mathcal{T}'_{\text{Half}}$) and doubled ($\mathcal{T}'_{\text{Double}}$) radian. Furthermore, the resulting CHOTA values are shown. It shows that the radian is a well-suited distance for the shift. As expected, the standard deviation of the estimation increases when applying larger shifts. In most cases, the standard deviation increases significantly when using the doubled shift. This indicates that the standard deviation of a larger amount of cells increases, presumably also from certain ones. On almost all datasets (except Fluo-N2DH-GOWT1), our distance leads to the best results compared to other parameter settings.

[2] www.celltrackingchallenge.net/latest-ctb-results

[3] www.tnt.uni-hannover.de/de/project/MPT/data/BiologicalNeeds/Videos.zip

[4] www.celltrackingchallenge.net/ctc-vii

Table 5. Additional metrics derived from the *py-ctcmetrics* framework (Kaiser et al., 2024). The *Global* block contains metrics that evaluate long-term consistencies. *Local* metrics are only influenced by frame-to-frame associations. The metrics denoted as *Detection* evaluate the detection capabilities and SEG is a segmentation quality measure. We refer to (Kaiser et al., 2024) for detailed metric descriptions.

		BF-C2DL -HSC		**BF-C2DL -MuSC**		**DIC-C2DH -HeLa**		**Fluo-C2DL -MSC**		**Fluo-N2DH -GOWT1**(Bártová et al., 2011)		**Fluo-N2DL -HeLa**(Neumann et al., 2010)		**PhC-C2DH -U373**		**PhC-C2DL -PSC**(Rapoport et al., 2011)		**Fluo-N2DH -SIM+**(Svoboda & Ulman, 2017)	
	Sequence	01	02	01	02	01	02	01	02	01	02	01	02	01	02	01	02	01	02
Global	CHOTA	79.4	74.1	84.6	78.0	96.5	87.1	93.4	80.1	98.2	95.4	93.9	91.3	92.8	91.6	80.0	85.1	97.8	95.0
	HOTA	79.6	89.8	79.7	76.6	96.3	93.5	93.4	76.8	97.7	96.6	94.4	92.6	89.3	87.5	86.1	89.3	97.7	93.7
	IDF1	77.6	87.4	76.5	69.4	97.1	94.5	94.9	73.1	97.6	97.5	93.4	91.5	86.3	81.8	84.1	87.8	98.1	94.0
	MT	95.7	84.8	66.7	42.2	97.2	89.7	86.7	60.0	92.6	90.6	94.3	92.3	85.7	85.7	82.2	85.0	97.8	87.6
	ML	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Local	MOTA	51.4	96.4	77.2	93.5	95.6	91.0	90.2	72.9	99.3	96.3	95.2	93.4	84.1	91.2	88.3	92.0	97.9	93.9
	TRA	95.7	99.6	96.0	98.6	98.3	93.7	91.0	95.1	99.5	97.2	99.0	98.7	98.6	95.6	96.8	97.7	98.7	96.2
	LNK	99.6	99.5	93.4	95.7	98.0	92.3	90.9	94.9	99.1	97.0	98.3	97.6	99.8	87.8	94.3	95.3	98.5	93.4
	IDSW	11	159	123	170	1	1	0	4	4	2	39	185	1	3	910	620	1	31
Detection	DET	95.1	99.6	96.4	99.0	98.4	93.9	91.0	95.1	99.5	97.2	99.1	98.8	98.4	96.8	97.2	98.0	98.7	96.7
	Precision	67.4	96.7	84.4	96.9	97.1	96.8	99.0	81.3	100.0	99.0	96.4	95.2	86.3	97.3	92.6	95.2	99.2	98.6
	Recall	99.9	100.0	98.8	99.6	98.7	94.2	91.1	97.3	99.5	97.3	99.5	99.5	100.0	99.7	98.5	99.1	98.8	97.4
	FAF	2.4	1.2	0.8	0.2	0.4	0.4	0.1	0.9	0.0	0.3	3.7	14.8	1.1	0.5	21.2	11.3	0.3	0.6
	F1	80.5	98.3	91.0	98.2	97.9	95.5	94.9	88.6	99.7	98.2	97.9	97.3	92.7	98.5	95.5	97.1	99.0	98.0
	FP	4192	2172	987	235	33	32	4	42	1	24	320	1280	121	19	5599	2814	22	48
	FN	5	14	67	31	15	60	38	5	10	68	41	137	0	2	1068	537	32	89
	SEG	90.4	86.2	80.2	76.6	89.7	87.8	63.2	68.3	92.6	95.9	86.5	89.5	94.1	86.1	77.5	75.9	89.4	78.6

Table 6. Ablation studies using training data with complete cell life cycles. We set sampling parameter $A_{\max} = 1$, $H_{\max} = 1$, deactivated mitosis costs $c_k^{\mathrm{M},i,h} = 0$, substituted our motion model with Kalman and compare it to *EmbedTrack* (Löffler & Mikut, 2022) and *Trackastra* (Gallusser & Weigert, 2024). Our method performs substantially better on long sequences (upper three datasets) with complex scenarios if all contributions are applied.

	CCA [%]↑	**Ours**	$A_{\max}$	$H_{\max}$	$c_k^{\mathrm{M},i,h}$	Kalman	EmbedTrack	Trackastra
Dataset	BF-C2DL-HSC	**77.32**	**77.32**	72.40	65.54	59.25	12.91	27.85
	BF-C2DL-MuSC	**35.09**	**35.09**	25.41	30.13	24.07	5.24	10.56
	PhC-C2DL-PSC	71.48	**73.76**	70.53	69.48	66.85	48.81	62.15
	Fluo-N2DH-SIM+	**43.28**	43.27	43.27	**43.28**	0.0	43.25	94.74
	Fluo-N2DL-HeLa	89.00	89.42	**91.94**	81.55	61.73	59.05	74.84
	TF [%]↑	Ours	$A_{\max}$	$H_{\max}$	$c_k^{\mathrm{M},i,h}$	Kalman	EmbedTrack	Trackastra
Dataset	BF-C2DL-HSC	**93.10**	91.60	**93.10**	83.99	88.20	74.69	79.21
	BF-C2DL-MuSC	**74.57**	72.19	73.35	72.49	62.45	64.46	64.06
	PhC-C2DL-PSC	**87.40**	**87.40**	87.29	87.33	86.67	86.55	87.23
	Fluo-N2DH-SIM+	**93.58**	93.54	93.33	93.31	91.68	92.90	93.99
	Fluo-N2DL-HeLa	94.05	94.02	**94.55**	93.70	91.71	93.02	94.96
	CHOTA [%]↑	Ours	$A_{\max}$	$H_{\max}$	$c_k^{\mathrm{M},i,h}$	Kalman	EmbedTrack	Trackastra
Dataset	BF-C2DL-HSC	76.75	74.78	**77.30**	48.31	69.31	54.03	56.99
	BF-C2DL-MuSC	**82.24**	80.85	80.38	68.75	32.95	65.15	58.04
	PhC-C2DL-PSC	**83.07**	82.51	82.23	81.85	64.91	74.56	75.86
	Fluo-N2DH-SIM+	**96.37**	**96.37**	96.33	96.36	71.51	96.35	95.74
	Fluo-N2DL-HeLa	**92.61**	92.45	92.36	91.87	79.83	86.08	89.07

6.4. Ablations

The proposed method is a sophisticated system that addresses potential errors by integrating multiple concepts. In the following ablations applied on training data with publicly available ground truth, summarized in Table 6, we explore the strengths, weaknesses, and gain further insights. To assess the impact of our method, we conducted the following experiments: 1) setting the number of sampled hypotheses per association to $A_{\max} = 1$, 2) limiting the total number of hypotheses after pruning to $H_{\max} = 1$, 3) removing our introduced mitosis costs $c_k^{\mathrm{M},i,h} = 0$, and 4) substituting our motion estimation with a Kalman filter. Moreover, we compare us to the vanilla *EmbedTrack* without using our extended association strategy to quantify the overall impact. Since *EmbedTrack* is used as detection framework by your method, performance differences are only caused by the association strategy under equal detection and segmentation preconditions. We evaluated these experiments using training data with complete cell cycles and reported the biological metrics that are least susceptible to noise and the robust CHOTA metric.

The most expressive results can be observed without explicit mitosis costs in setting 3). On the long and complex sequences BF-C2DL-HSC and -MuSC, all metrics collapse significantly when no long-term consistency preserving mitosis costs are incorporated. On the shorter sequence PhC-C2DL-PSC, the effect is also visible but with a lower impact. The mitosis costs do not impact short sequences with short proliferation trees conceptually which is confirmed by the results.

In settings 1) and 2), where we did not evaluate multiple hypotheses, the metrics for the same long and complex image sequences dropped by up to 10 percent points. This drop is reasonable since the framework is forced to preserve the local optimal hypothesis and cannot resolve long-term errors.

Finally, setting 4) shows that the naive Kalman filter is not a suitable motion model to induce long-time consistency. We conclude that motion estimation based on visual cues should be preferred. Similarly, the naive nearest neighbour association strategy of our baseline *EmbedTrack* also leads to large drops in long-term consistency.

The main conclusions of this experiment are, that all of our proposed contributions contribute to a high performing association strategy. We demonstrate that our method is particularly well-suited for applications involving long sequences and complex scenarios.

[5] www.celltrackingchallenge.net/latest-clb-results

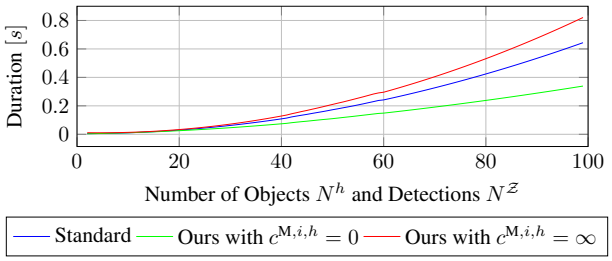

Figure 6. Runtime to solve the association problem between $N^h = N^z$ objects and detections with the Hungarian method. We compare the standard formulation against our mitosis-aware approach with cell splits allowed ($c^{\mathrm{M},i,h} = 0$) and forbidden ($c^{\mathrm{M},i,h} = \infty$). It converges faster allowing cell splits due to the simpler optimization problem.

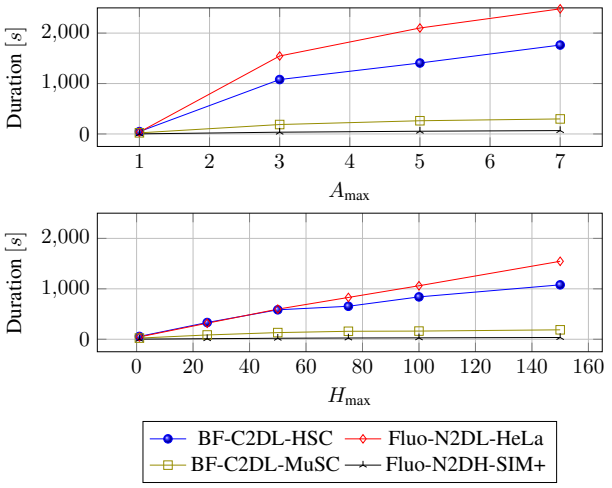

Figure 7. Runtime of our MHT framework with varying values of the sampling parameters $A_{\max}$ and $H_{\max}$.

6.5. Runtime

A significant ratio of the execution time of the MHT framework is spent by the sampling algorithm to draw new association hypotheses Ψ_k^h with sampling algorithms like Murtys (Murty, 1968) or Gibbs (Geman & Geman, 1984). Therefore, the potential impact of our novel mitosis-aware association cost matrix on the algorithms is of interest. To evaluate the impact, we perform the Hungarian algorithm (Kuhn, 1955) (which is the core of Murtys) on instances of $\mathbf{C}^{j,i,h}$ defined by the vanilla formulation from Equation (6) and with our novel formulation with mitosis costs from Equation (12). We sample 2000 different $\mathbf{C}^{j,i,h}$ with random costs and a squared but variable size (*i.e.* , the number of objects and detections is equal). Moreover, we set mitosis costs $c^{\mathrm{M},i,h}$ to either zero or infinity to simulate that mitosis is always allowed or strictly forbidden. Finally, we perform the Hungarian algorithm on all problem instances and aggregate the execution time. The results are presented in Figure 6.

When adding infinite costs $c^{\mathrm{M},i,h} = \infty$ to simulate the unlikely setting that mitosis is strictly forbidden, the execution time increases slightly as expected. While the underlying optimization problem stays the same, more elements need to be parsed by the algorithm. More interestingly, the more likely setting with $c^{\mathrm{M},i,h} = 0$ leads to drastic improvements in efficiency. This can be explained by the simpler optimization problem in which a heuristic initial solution is more often the final optimal solution. This reduces the calculation time in the Hungarian Method and allows to decrease the number of samples in approximations like Gibbs. Besides improving accuracy, this experiment shows that our contribution generally increases the efficiency of MHT frameworks.

To provide an intuition of the execution time, Figure 7 presents the execution time of our MHT framework on different datasets. We use the standard configuration but vary either the association sampling limit $A_{\max}$ or the hypotheses storage limit $H_{\max}$. It shows that the runtime grows approximately linearly concerning the investigated parameters. The computations were performed on a desktop PC with an Intel i9-9900K CPU running at 16×3.60GHz. It is important to note that we only consider the runtime of the MHT tracker without *EmbedTrack*.

6.6. Discussion

The experiments conducted on our proposed method reveal its strengths and weaknesses. Our framework reevaluates uncertain situations and incorporates long temporal and globally consistent lineage information. In practice, these advantages are taken into account in specific settings and scenarios.

Figure 4 and Figure 5 show scenarios where our method contributes to better results or leads to errors, respectively. In Figure 4, cell instances are very small and densely populated. The cells cannot be distinguished based on their appearance, and even mitosis lacks visual cues. When such cells exhibit strong displacements between consecutive frames—*e.g.* , cells 1, 2, and 3—trackers that rely on local visual cues, such as *EmbedTrack* or *Trackastra*, are prone to errors. This can result in implausible events, such as the mitosis observed in frames 1005–1007 with *EmbedTrack*. Due to the long temporal context, our mitosis-aware MHT system is able to detect those uncertain situations and resolve them.

In contrast, Figure 5 shows a shorter, less populated sequence with larger cells that are visually more distinguishable. Mitosis is also more apparent due to visible nuclei. The strengths of our method are not leveraged here, as the mitosis cost statistics (Equation (13)) are poorly estimated in this short sequence where cells also move in and out of the field of view. In fact, the mitosis costs lead to suppressed mitosis detections, as shown in Figure 5, frame $k = 70$ (upper row), where the cell with label 3 splits into two daughter

cells. Due to unaligned mitosis costs, our system assigns mitosis a lower probability and instead spawns a new cell with label 9, which is incorrectly assumed to have entered from outside the field of view.

The dataset statistics in Table 3, together with the benchmark results in Table 2, highlight scenarios where our framework enhances tracking accuracy. The most impressive improvements can be seen in the long datasets BF-C2DL-HSC and -MuSC. These datasets contain large numbers of cells that are relatively small and densely populated compared to the others. Improvement is also observed for the similarly populated but much shorter PhC-C2DL-PSC dataset, though the gain is smaller. It's worth noting that the relative cell motion per frame compared to the cell size is high in situations where our method typically performs well. This could be an indicator of the uncertainty induced by simple visual association methods. Short sequences with fewer and larger cells, like in Fluo-C2DL-MSC, do not benefit to the same extent. This clearly shows that the effectiveness of our method varies depending on the data characteristics of the application.

7. Conclusion

This paper presents a novel cell tracking framework that combines the strong local performance of neural tracking-by-regression approaches with the global optimal assignment strategy of MHT trackers. This fusion is achieved by predicting the estimation uncertainty of the motion regression framework using test-time augmentation and expanding the MHT assignment problem formulation to incorporate mitosis costs. We demonstrate that our approach outperforms the current state-of-the-art on various competitive datasets, without the need for additional data or re-training. Our ablation studies also offer insights into scenarios where long-term consistency is crucial and highlight when heuristic tracking-by-regression methods remain effective. We hope that this work raises awareness about the importance of long-term consistency within the cell tracking community.

References

Anjum, S. and Gurari, D. CTMC: Cell tracking with mitosis detection dataset challenge. In Conference on Computer Vision and Pattern Recognition Workshops, 2020. doi: 10.1109/CVPRW50498.2020.00499.

Antony, P. P. M. A., Trefois, C., Stojanovic, A., Baumuratov, A. S., and Kozak, K. Light microscopy applications in systems biology: opportunities and challenges. Cell Communication and Signaling, 2013. doi: 10.1186/1478-811X-11-24.

Arbelle, A. and Raviv, T. R. Microscopy cell segmentation via convolutional lstm networks. In International Symposium on Biomedical Imaging, 2019. doi: 10.1109/ISBI.2019.8759447.

Bao, R., Al-Shakarji, N. M., Bunyak, F., and Palaniappan, K. Dmnet: Dual-stream marker guided deep network for dense cell segmentation and lineage tracking. In International Conference on Computer Vision Workshops, 2021. doi: 10.1109/ICCVW54120.2021.00375.

Belyaev, I., Praetorius, J.-P., Medyukhina, A., and Figge, M. T. Enhanced segmentation of label-free cells for automated migration and interaction tracking. Cytometry Part A, 2021. doi: 10.1002/cyto.a.24466.

Ben-Haim, T. and Raviv, T. R. Graph neural network for cell tracking in microscopy videos. In European Conference on Computer Vision, pp. 610–626. Springer, 2022.

Bergmann, P., Meinhardt, T., and Leal-Taixe, L. Tracking without bells and whistles. In International Conference on Computer Vision, 2019. doi: 10.1109/ICCV.2019.00103.

Bernardin, K. and Stiefelhagen, R. Evaluating multiple object tracking performance: the clear mot metrics. Image and Video Processing, 2008, 2008. doi: 10.1155/2008/246309.

Bártová, E., Šustáčková, G., Stixová, L., Kozubek, S., Legartová, S., and Foltánková, V. Recruitment of oct4 protein to uv-damaged chromatin in embryonic stem cells. PLOS ONE, 6(12):1–13, 12 2011. doi: 10.1371/journal.pone.0027281.

Chen, Y., Song, Y., Zhang, C., Zhang, F., O'Donnell, L., Chrzanowski, W., and Cai, W. Celltrack r-cnn: A novel end-to-end deep neural network for cell segmentation and tracking in microscopy images. In Symposium on Biomedical Imaging, 2021. doi: 10.1109/ISBI48211.2021.9434057.

Crouse, D. F., Willett, P., Pattipati, K., and Svensson, L. A look at gaussian mixture reduction algorithms. In International Conference on Information Fusion. IEEE, 2011.

Gallusser, B. and Weigert, M. Trackastra: Transformer-based cell tracking for live-cell microscopy. In European Conference on Computer Vision, 2024.

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al. A survey of uncertainty in deep neural networks. Artificial Intelligence Review, 2023. doi: 10.1007/s10462-023-10562-9.

Geman, S. and Geman, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence, pp. 721–741, 1984.

Granström, K., Svensson, L., Xia, Y., Williams, J., and García-Fernández, A. F. Poisson multi-bernoulli mixture trackers: Continuity through random finite sets of trajectories. In International Conference on Information Fusion, 2018. doi: 10.23919/ICIF.2018.8455849.

Gupta, D. K., de Bruijn, N., Panteli, A., and Gavves, E. Tracking-assisted segmentation of biological cells. Conference on Neural Information Processing Systems Workshops, 2019. doi: 10.48550/arXiv.1910.08735.

Hornakova, A., Kaiser, T., Rolinek, M., Rosenhahn, B., Swoboda, P., Henschel, R., and equal contribution), . Making higher order mot scalable: An efficient approximate solver for lifted disjoint paths. In International Conference on Computer Vision, 2021. doi: 10.1109/ICCV48922.2021.00627.

Hossain, M. I., Gostar, A. K., Bab-Hadiashar, A., and Hoseinnezhad, R. Visual mitosis detection and cell tracking using labeled multi-bernoulli filter. In International Conference on Information Fusion, 2018. doi: 10.23919/ICIF.2018.8455486.

Hu, T., Xu, S., Wei, L., Zhang, X., and Wang, X. CellTracker: an automated toolbox for single-cell segmentation and tracking of time-lapse microscopy images. Bioinformatics, 37, 2021. doi: 10.1093/bioinformatics/btaa1106.

Kaiser, T., Reinders, C., and Rosenhahn, B. Compensation learning in semantic segmentation. In Computer Vision and Pattern Recognition Workshops (CVPRW), June 2023.

Kaiser, T., Vladimir, U., and Rosenhahn, B. Chota: A higher order accuracy metric for cell tracking. In European Conference on Computer Vision Workshops (ECCVW), October 2024.

Kalman, R. E. A new approach to linear filtering and prediction problems. Basic Engineering, 82, 1960.

Kuhn, H. W. The hungarian method for the assignment problem. Naval research logistics quarterly, 2(1-2):83–97, 1955.

Kukulage, D. S., Samarasinghe, K. T., Don, N. N. M., Shivamadhu, M. C., Shishikura, K., Schiff, W., Ramezani, F. M., Padmavathi, R., Matthews, M. L., and Ahn, Y.-H. Protein phosphatase pp2cα s-glutathionylation regulates cell migration. Journal of Biological Chemistry, 300(10), 2024.

Lalit, M., Tomancak, P., and Jug, F. Embedseg: Embedding-based instance segmentation for biomedical microscopy data. Medical Image Analysis, 81, 2022. doi: 10.1016/j.media.2022.102523.

Liao, W., Luo, L., Wang, C., and Zhang, C. Cell DINO: End-to-End Cell Segmentation and Tracking with Transformer . In 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3491–3494. IEEE Computer Society, December 2024. doi: 10.1109/BIBM62325.2024.10821971.

Löffler, K., Scherr, T., and Mikut, R. A graph-based cell tracking algorithm with few manually tunable parameters and automated segmentation error correction. PLOS ONE, 16, 2021. doi: 10.1371/journal.pone.0249257.

Löffler, K. and Mikut, R. Embedtrack—simultaneous cell segmentation and tracking through learning offsets and clustering bandwidths. IEEE Access, 10, 2022. doi: 10.1109/ACCESS.2022.3192880.

Magnusson, K. E. G., Jaldén, J., Gilbert, P. M., and Blau, H. M. Global linking of cell tracks using the viterbi algorithm. Medical Imaging, 34, 2015. doi: 10.1109/TMI.2014.2370951.

Malin-Mayor, C., Hirsch, P., Guignard, L., McDole, K., Wan, Y., Lemon, W. C., Kainmueller, D., Keller, P. J., Preibisch, S., and Funke, J. Automated reconstruction of whole-embryo cell lineages by learning from sparse annotations. Nature biotechnology, 41(1):44–49, 2023.

Mancusi, G., Panariello, A., Porrello, A., Fabbri, M., Calderara, S., and Cucchiara, R. Trackflow: Multi-object tracking with normalizing flows. In International Conference on Computer Vision, 2023. doi: 10.1109/ICCV51070.2023.00874.

Maška, M., Ulman, V., Svoboda, D., Matula, P., Matula, P., Ederra, C., Urbiola, A., España, T., Venkatesan, S., Balak, D. M., Karas, P., Bolcková, T., Štreitová, M., Carthel, C., Coraluppi, S., Harder, N., Rohr, K., Magnusson, K. E. G., Jaldén, J., Blau, H. M., Dzyubachyk, O., Křížek, P., Hagen, G. M., Pastor-Escuredo, D., Jimenez-Carretero, D., Ledesma-Carbayo, M. J., Muñoz-Barrutia, A., Meijering, E., Kozubek, M., and Ortiz-de Solorzano, C. A benchmark for comparison of cell tracking algorithms. Bioinformatics, 30, 2014. doi: 10.1093/bioinformatics/btu080.

Maška, M., Ulman, V., Delgado-Rodriguez, P., Gómez de Mariscal, E., Necasova, T., Guerrero Peña, F. A., Ing Ren, T., Meyerowitz, E., Scherr, T., Löffler, K., Mikut, R., Guo, T., Wang, Y., Allebach, J., Bao, R., Al-Shakarji, N., Rahmon, G., Toubal, I. E., Palaniappan, K., and Ortiz-de Solorzano, C. The cell tracking challenge: 10 years of

objective benchmarking. Nature Methods, 20, 2023. doi: 10.1038/s41592-023-01879-y.

Matula, P., Maška, M., Sorokin, D. V., Matula, P., Ortiz-de Solorzano, C., and Kozubek, M. Cell tracking accuracy measurement based on comparison of acyclic oriented graphs. PLOS ONE, 10, 2015. doi: 10.1371/journal.pone.0144959.

Mori, S., Chong, C.-Y., Tse, E., and Wishner, R. Tracking and classifying multiple targets without a priori identification. IEEE Transactions on Automatic Control, 31(5): 401–409, 1986. doi: 10.1109/TAC.1986.1104306.

Murty, K. G. An algorithm for ranking all the assignments in order of increasing cost. Operations Research, 16, 1968. doi: 10.1287/opre.16.3.682.

Neumann, B., Walter, T., Hériché, J.-K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U., et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature, 464(7289):721–727, 2010.

Nguyen, T. T. D., Vo, B.-N., Vo, B.-T., Kim, D. Y., and Choi, Y. S. Tracking cells and their lineages via labeled random finite sets. Signal Processing, 69, 2021. doi: 10.1109/TSP.2021.3111705.

O'Connor, O. M. and Dunlop, M. J. Cell-tractr: A transformer-based model for end-to-end segmentation and tracking of cells. bioRxiv, 2024. doi: 10.1101/2024.07.11.603075.

Paul, R. D., Seiffarth, J., Scharr, H., and Nöh, K. Robust approximate characterization of single-cell heterogeneity in microbial growth. In International Symposium on Biomedical Imaging, 2024.

Payer, C., Štern, D., Feiner, M., Bischof, H., and Urschler, M. Segmenting and tracking cell instances with cosine embeddings and recurrent hourglass networks. Medical Image Analysis, 57, 2019. doi: 10.1016/j.media.2019.06.015.

Rapoport, D. H., Becker, T., Madany Mamlouk, A., Schicktanz, S., and Kruse, C. A novel validation algorithm allows for automated cell tracking and the extraction of biologically meaningful parameters. PLOS ONE, 6(11): 1–16, 11 2011. doi: 10.1371/journal.pone.0027315.

Reid, D. An algorithm for tracking multiple targets. Automatic Control, 1979. doi: 10.1109/TAC.1979.1102177.

Rezatofighi, S. H., Gould, S., Vo, B. T., Vo, B.-N., Mele, K., and Hartley, R. Multi-target tracking with time-varying clutter rate and detection profile: Application to time-lapse cell microscopy sequences. Medical Imaging, 34, 2015. doi: 10.1109/TMI.2015.2390647.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24574-4_28.

Scherr, T., Löffler, K., Böhland, M., and Mikut, R. Cell segmentation and tracking using cnn-based distance predictions and a graph-based matching strategy. PLOS ONE, 15, 2020. doi: 10.1371/journal.pone.0243219.

Schiegg, M., Hanslovsky, P., Kausler, B. X., Hufnagel, L., and Hamprecht, F. A. Conservation tracking. In International Conference on Computer Vision, 2013. doi: 10.1109/ICCV.2013.364.

Sixta, T., Cao, J., Seebach, J., Schnittler, H., and Flach, B. Coupling cell detection and tracking by temporal feedback. Machine Vision and Applications, 31, 2020. doi: 10.1007/s00138-020-01072-7.

Solano-Carrillo, E., Sattler, F., Alex, A., Klein, A., Costa, B. P., Rodriguez, A. B., and Stoppe, J. Utrack: Multi-object tracking with uncertain detections, 2024.

Stegmaier, J. and Mikut, R. Fuzzy-based propagation of prior knowledge to improve large-scale image analysis pipelines. PLOS ONE, 12, 2017. doi: 10.1371/journal.pone.0187535.

Stegmaier, J., Alshut, R., Reischl, M., and Mikut, R. Information fusion of image analysis, video object tracking, and data mining of biological images using the open source matlab toolbox gait-cad. Biomedical Engineering, 57, 2012. doi: doi:10.1515/bmt-2012-4073.

Svoboda, D. and Ulman, V. Mitogen: A framework for generating 3d synthetic time-lapse sequences of cell populations in fluorescence microscopy. IEEE Transactions on Medical Imaging, 36(1):310–321, 2017. doi: 10.1109/TMI.2016.2606545.

Theorell, A., Seiffarth, J., Grünberger, A., and Nöh, K. When a single lineage is not enough: uncertainty-aware tracking for spatio-temporal live-cell image analysis. Bioinformatics, 35(7):1221–1228, 2019.

Ulman, V., Maška, M., Magnusson, K. E., Ronneberger, O., Haubold, C., Harder, N., Matula, P., Matula, P., Svoboda, D., Radojevic, M., et al. An objective comparison of cell-tracking algorithms. Nature Methods, 14, 2017. doi: 10.1038/nmeth.4473.

Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., and Vercauteren, T. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing, 338, 2019. doi: 10.1016/j.neucom.2019.01.103.

Wang, J. and Lukasiewicz, T. Rethinking bayesian deep learning methods for semi-supervised volumetric medical image segmentation. In Conference on Computer Vision and Pattern Recognition, 2022. doi: 10.1109/CVPR52688.2022.00028.

Wehrbein, T., Rudolph, M., Rosenhahn, B., and Wandt, B. Utilizing uncertainty in 2d pose detectors for probabilistic 3d human mesh recovery. In IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), February 2025.

Yates, C. A., Ford, M. J., and Mort, R. L. A multi-stage representation of cell proliferation as a markov process. Bulletin of mathematical biology, 79:2905–2928, 2017.

Yoon, B., Kim, H., Jung, S. W., and Park, J. Single-cell lineage tracing approaches to track kidney cell development and maintenance. Kidney International, 105 (6):1186–1199, 2024. ISSN 0085-2538. doi: https://doi.org/10.1016/j.kint.2024.01.045.

Zhou, L., Tang, T., Hao, P., He, Z., Ho, K., Gu, S., Hou, W., Hao, Z., Sun, H., Zhan, K., Jia, P., Lang, X., and Liang, X. Ua-track: Uncertainty-aware end-to-end 3d multi-object tracking, 2024.

Zhou, Z., Wang, F., Xi, W., Chen, H., Gao, P., and He, C. Joint multi-frame detection and segmentation for multi-cell tracking. In Conference on Image and Graphics. Springer, 2019. doi: 10.1007/978-3-030-34110-7_36.