This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Bayesian Transformers and Higher-Order Graph Matching for Cell Tracking in Serial Tissue Sections

Mostafa Karami[1], Sahand Hamzehei[1], David Arce[2], Gianna Raimondi[2], Linnaea Ostroff[2], and Sheida Nabavi[1⋆]

[1] School of Computing, Department of Computer Science & Engineering, University of Connecticut , Storrs, CT, USA
{mostafa.karami,sahand.hamzehei,sheida.nabavi}@uconn.edu
[2] Department of Physiology & Neurobiology, University of Connecticut , Storrs, CT, USA
{david.arce,linnaea.ostroff}@uconn.edu
Raimondigia@gmail.com

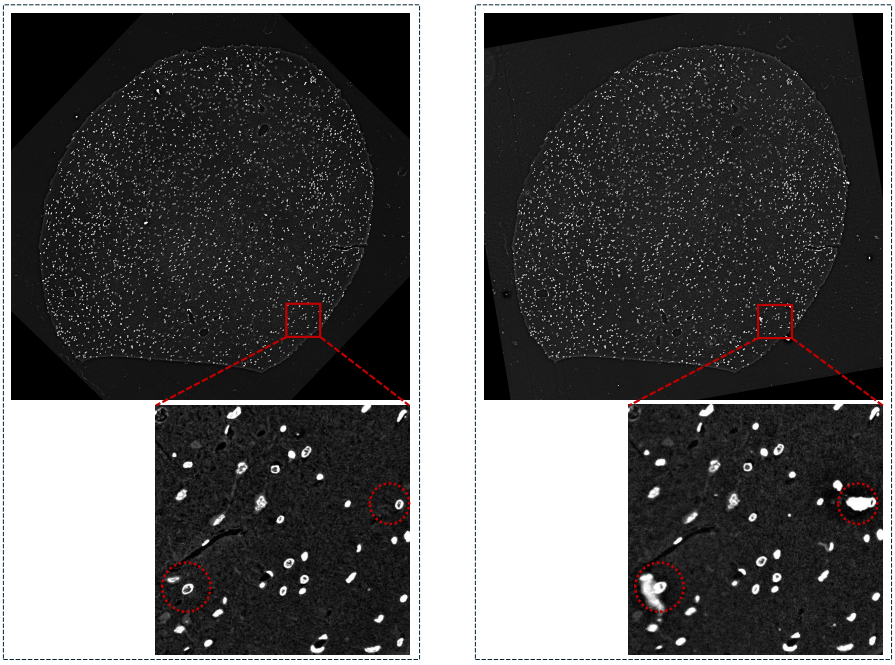
Abstract. Reliable 3D reconstruction of tissue architecture from sequential 2D multiplex images is challenging due to the noise and distortions introduced by ultrathin (50 nm) slicing and complex alignment procedures. Conventional cell tracking methods often fail under such conditions, resulting in inaccurate linkage of cells across sections. To bridge this gap, we propose a Bayesian Transformer framework that incorporates uncertainty-aware feature embeddings and higher-order graph matching with belief propagation. By tracking cells across consecutive sections, our method facilitates the 3D reconstruction of volumetric tissue organization, even in highly noise-prone scenarios. The methodology begins with a standard segmentation step, followed by feature extraction that computes morphological, shape, and texture descriptors, as well as deep CNN embeddings. These rich, uncertainty-sensitive representations reduce errors caused by both registration artifacts and morphological variability. We validate the effectiveness of the proposed approach on a private multiplex dataset of fixed tissue sections and further demonstrate its generalizability on public time-lapse microscopy videos, showcasing adaptability to diverse datasets. Experimental comparisons reveal that our method outperforms baseline tracking techniques, achieving higher accuracy and more consistent cell linkages across multiple serial sections. The code used in this research with sample dataset are publicly available at https://github.com/NabaviLab/bayesian-transformer-cell-tracking.

Keywords: Bayesian Transformers · Higher-Order Graph Matching · Belief Propagation · Cell Tracking · Multiplex Imaging

⋆ Corresponding author: Sheida Nabavi (sheida.nabavi@uconn.edu)

1 Introduction

Reliable 3D reconstruction of tissue from sequential 2D fluorescence images remains a key challenge in bioimaging [26]. Thin slicing and tissue alignment introduce noise and distortions that prevent traditional vision methods [1, 9]. Tracking cell instances across sections is crucial for linking corresponding cells [23], enabling precise mapping of molecular expression and lineage relationships in healthy and diseased tissues [19].



(***a***) (***b***)

Fig. 1: Panels (a) and (b) show two consecutive slices of multiplexed immunofluorescence images. Red circles in slice (b) highlight noisy-background artifacts that obscure the cells and complicate their tracking.

Deep learning has improved 2D cell segmentation [25], but tracking cells across sections, especially under noisy or variable staining, remains challenging (Figure 1). Traditional methods struggle with misalignment, morphological inconsistencies, and background variability [17,32], challenges further intensified in highly multiplexed imaging, where conditions vary across slices and channels [22].

Existing pipelines combine a segmentation model with an independent tracking algorithm, both requiring extensive hyperparameter tuning. While some modern approaches perform well, they rely on large annotated datasets [10,

20, 28], which are costly to obtain for large-scale studies. Performance often fails under varying conditions, as many methods overlook uncertainty in feature measurements and subtle morphology changes between sections [2, 11]. Traditional approaches [16] rely on geometric or intensity-based cost functions, making them prone to errors when shape or texture varies significantly. Without uncertainty modeling or higher-order constraints, mislinking can propagate, distorting the final 3D reconstruction.

In this work, we address the limitations of existing cell tracking pipelines with an unsupervised approach that relies only on features extracted from automatically generated segmentation masks. Our method first computes morphological, shape, and deep descriptors from segmented cells, eliminating the need for extensive manual annotations. We then construct a higher-order graph where nodes represent cell features, and edges link candidate matches across frames. A Bayesian transformer [34] handles variability in cell appearance and imaging artifacts with minimal hyperparameter tuning. Moreover, belief propagation encourages consistent correspondences even under shape changes or misalignments, enabling scalability across datasets while reducing supervision and manual post-processing.

Our main contributions can be summarized as follows: 1. We introduce a Bayesian transformer that refines cell features by modeling uncertainty in morphological, shape-based, and deep descriptors for reducing errors from misalignments or noisy backgrounds in multiplexed tissue sections and live-cell sequences. 2. We formulate cell tracking within a higher-order graph and use belief propagation to enforce consistent matches and reject outliers. This handles major appearance changes, overlapping cells, and divisions across diverse 2D microscopy images, ensuring globally consistent lineage mappings. 3. We present an unsupervised approach generalizing across diverse staining and imaging modalities without ground-truth annotations. 4. We validate on private multiplex imaging data and public benchmarks, showing reliable tracking despite artifacts and variability, with consistent accuracy improvements and reduced manual curation across multiple datasets.

2 Methods

The pipeline of the proposed method is divided into three main stages: 1) Feature extraction, 2) Bayesian transformer embedding, and 3) Higher-order graph matching as shown in Figure 2.

Feature Extraction We begin by extracting meaningful descriptors for each cell using the binary masks generated by Cellpose [27], which identifies individual cells and assigns each one a unique label [9]. Next, we crop each cell's region from the raw image and compute a range of descriptors to form a feature vector $\mathbf{x}_i^t \in \mathbb{R}^d$. Specifically, we gather morphological properties (e.g., centroid, bounding box, area, perimeter, aspect ratio), various shape descriptors (Zernike

moments [29], elliptical Fourier descriptors [15], skeleton-based measures [4]), and texture features (Haralick descriptors [12] and Local Binary Patterns [21]).

We also use a pretrained ResNet-50 [13] (initialized with standard ImageNet [8] weights) to obtain deep CNN embeddings. Specifically, for each segmented cell, we crop its grayscale bounding patch, replicating that into three channels, and then apply the standard resizing and normalization pipeline. Next, the patch is passed through the truncated ResNet-50 to generate a feature vector for each individual cell in each frame. These descriptors are concatenated into a single vector per cell. Any feature columns that are entirely missing or constant across all cells are removed. The resulting feature matrix for frame t is $\mathbf{X}_t = [\mathbf{x}_1^t; \ldots; \mathbf{x}_N^t] \in \mathbb{R}^{N \times d}$, where N is the number of cells in that frame, and this matrix is fed to the Bayesian-transformer embedding and graph-matching stages.

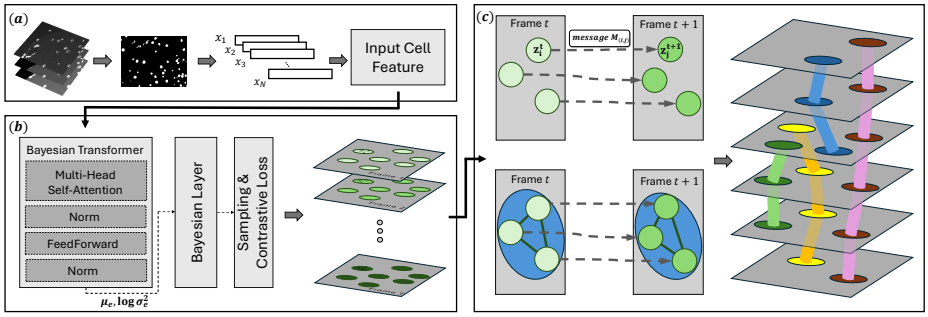


Fig. 2: The overall structure of the proposed method; (a) Feature extraction from segmentation masks; (b) Bayesian transformer-based embedding for capturing parameter uncertainty; (c) Higher-order graph matching with belief propagation and final lineage export.

Bayesian Transformer Embedding We aim to learn a reliable representation of the cell features that also encodes a probabilistic notion of uncertainty. To achieve this, we adopt a Bayesian transformer encoder, which applies Bayesian linear layers in both the multi-head self-attention module (handling the query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) projections, as expressed in Equation 1) and the feed-forward sub-layers, while preserving the usual residual connections and layer normalization.

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q,\ \mathbf{K} = \mathbf{X}\mathbf{W}_K,\ \mathbf{V} = \mathbf{X}\mathbf{W}_V, \qquad \mathbf{W}_{Q,K,V} \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{B \times 1 \times d_{\text{model}}}$ (with $d_{\text{model}} = 64$ and B the mini-batch size) and each projection matrix satisfies $\mathbf{W} = \mu_W + \sigma_W \odot \varepsilon,\ \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For simplicity, we use $\mathbf{W}$ to stand for any one of the three projection matrices $\mathbf{W}_{Q,K,V}$. Here

$\mu_W, \sigma_W \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$ are the element-wise mean and standard deviation of the weights, and $\odot$ denotes the Hadamard (element-wise) product. Splitting into heads reshapes $\mathbf{Q}$ to $(B, h, 1, d_h)$; the same applies to $\mathbf{K}$ and $\mathbf{V}$.

At each frame t we therefore feed a tensor $\mathbf{X} \in \mathbb{R}^{B \times 1 \times d_{\text{model}}}$. For each linear layer we place factorised Gaussian posteriors on both the weights $\mathbf{W}$ and the bias vector $\mathbf{b}$:

$$\mathbf{W} \sim q(\mathbf{W} \,|\, \mu_W, \sigma_W), \qquad \mathbf{b} \sim q(\mathbf{b} \,|\, \mu_b, \sigma_b), \tag{2}$$

where $\mathbf{W}$ is the weight matrix of the linear layer, and q denotes an approximate posterior distribution over $\mathbf{W}$. Also, μ_W and σ_W are trainable parameters with the same dimension as $\mathbf{W}$. We employ the reparameterization trick [14,24], which enables gradient-based learning through random variables by expressing $\mathbf{W}$ as a deterministic function of μ_W, $\log \sigma_W$, and a noise sample. The forward pass samples:

$$\mathbf{W} \;=\; \mu_W + \exp\Big(\tfrac{1}{2} \log \sigma_W^2\Big)\, \epsilon, \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{3}$$

where μ_W and σ_W remain learnable weights, and ϵ is a standard normal sample. By modeling μ_W and σ_W for each layer (rather than a single fixed $\mathbf{W}$), we capture parameter uncertainty during the $Q/K/V$ transformations. We then split these into h heads, each of dimension $d_h = d_{\text{model}}/h$. A scaled dot-product attention is computed for each head:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\big(\tfrac{1}{\sqrt{d_h}} \mathbf{Q}\mathbf{K}^{\mathsf{T}}\big)\mathbf{V}. \tag{4}$$

The transformer encoder produces hidden states that are then mapped to μ_e and $\log \sigma_e^2$ parameters (embedding mean and variance). Specifically, we stack multiple encoder blocks and, at the final stage, pass their output through two Bayesian linear layers, $\mathbf{W}_{\mu_e}$ and $\mathbf{W}_{\log \sigma_e^2}$, which yield these parameters. After the final encoder block produces a hidden state $\mathbf{H} \in \mathbb{R}^{B \times 1 \times d_{\text{embed}}}$, we apply $\mu_e = \mathbf{W}_{\mu_e}\, \mathbf{H} \quad \text{and} \quad \log \sigma_e^2 = \mathbf{W}_{\log \sigma_e^2}\, \mathbf{H}$.

We sample consecutive frames $(t, t+1)$ and randomly select subsets of cells from each to form training examples. Specifically, let $\mathbf{x}_i^t \mapsto (\mu_i^t, \log \sigma_i^t), \quad \mathbf{x}_j^{t+1} \mapsto (\mu_j^{t+1}, \log \sigma_j^{t+1})$. We then reparameterize each to get a sample: $\mathbf{z}_i^t = \mu_i^t + \exp\Big(\frac{1}{2} \log \sigma_i^t\Big)\, \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, I)$, and similarly sample $\mathbf{z}_j^{t+1}$ from $(\mu_j^{t+1}, \log \sigma_j^{t+1})$. To ensure that embeddings of the same cell (across consecutive frames) are pulled together, while different cells are pushed apart, we minimize a margin-based contrastive loss [7]. For every anchor cell i we pair one *positive* j (the same cell in frame t+1) and sample one *negative* k (a different cell in that frame), so k indexes the negative example:

$$\mathcal{L}_{\text{contrast}} \;=\; \max\Big(0,\, \big\|\mathbf{z}_i^t - \mathbf{z}_j^{t+1}\big\|^2 \;-\; \big\|\mathbf{z}_i^t - \mathbf{z}_k^{t+1}\big\|^2 \;+\; m\Big), \tag{5}$$

where $\|\cdot\|^2$ is the squared Euclidean norm in $\mathbb{R}^{d_{\text{embed}}}$ and m is a margin (set to 0.2).

Higher-Order Graph Matching After obtaining embeddings for each cell, we construct a higher-order graph to match cells between consecutive frames. Specifically, let each cell i in frame t have an embedding $\mathbf{c}_i^t \in \mathbb{R}^{d'}$, where d' is the dimension of the Bayesian transformer's output space. We connect these cells across frames by two types of edges: *(i) single-node edges* (linking a single cell i in frame t to a candidate cell j in frame $t+1$), and *(ii) triplet edges* (linking triplets (i_1, i_2, i_3) in frame t to corresponding triplets (j_1, j_2, j_3) in frame $t+1$). Triplets are not required to be disjoint; we enumerate every three-cell combination within a distance threshold so that a single node may belong to multiple triplets. This overlap lets belief propagation share geometric evidence across neighboring triplets. We define the matching cost for each single-node edge (i, j) as:

$$\mathrm{Cost}(i,j) \;=\; \underbrace{\mathrm{Dist}\big(\mathbf{c}_i^t, \mathbf{c}_j^{t+1}\big)}_{\text{first-order}} \;+\; \underbrace{M_{(i,j)}}_{\text{message term}}, \tag{6}$$

where $\mathrm{Dist}(\cdot,\cdot)$ is the Euclidean distance in embedding space, and $M_{(i,j)}$ is the message accumulated through a belief-propagation. While the first-order term handles direct embedding distances, the message $M_{(i,j)}$ aggregates information from triplet edges to ensure higher-order consistency. We store the message for each single-edge (i, j) at iteration ℓ , $M_{(i,j)}^{(\ell)}$, and iteratively update them based on triplet edges that connect (i, j) with (i', j') forming $(i, i') \mapsto (j, j')$. For each iteration $\ell + 1$:

$$\begin{aligned} \mathbf{M}_{(i,j)}^{(\ell+1)} =& (1-\gamma) \sum_{(i',j')\in\mathcal{E}(i,j)} \big[-\mathrm{cost}_{\mathrm{triplet}}(i,i',j,j')\big] \\ & \times \exp\Big(-\tfrac{1}{2}\big[\log\sigma_i^2 + \log\sigma_j^2\big]\Big) \;+\; \gamma\, M_{(i,j)}^{(\ell)}, \end{aligned} \tag{7}$$

where $\mathcal{E}(i, j)$, as the set of all triplet edges that contain the single-node edge (i, j), enumerates triplet edges that include (i, j), $\mathrm{cost}_{\mathrm{triplet}}$ [30] measures geometry differences among triplets, and $\log\sigma_i^2, \log\sigma_j^2$ are the variances from the Bayesian transformer (omitting any layer index for brevity). The damping factor $\gamma \in (0, 1)$ controls the integration of new vs. old messages.

Final Assignment and Tracking After L iterations, the final messages $M_{(i,j)}^{(L)}$ form a global cost matrix for cell matching, from which we determine possible links $(i \to j)$. If a single cell i links to multiple successors, we label it a division event. Unmatched cells are considered to have disappeared or newly appeared. Finally, we compile these matches, divisions, and unmatched events across frames into a consistent lineage table.

3 Experiment Results

Datasets To validate our method, we first used an internally acquired Ultraplex fluorescence microscopic dataset in 10 consecutive frames, which serves as

a challenging testbed prone to slice-to-slice geometric distortions and noise. Second, we used four representative reference datasets from the Cell Tracking Challenge [17,32], namely Fluo-C2DL-MSC, PhC-C2DL-PSC, Fluo-C2DL-Huh7, and Fluo-N2DL-HeLa, capturing a range of cell lines (e.g., HeLa, Huh7, and mesenchymal stem cells) and imaging conditions (phase-contrast, fluorescence), each with different acquisition rates and noise characteristics.

Evaluation Metrics We adopt four standard cell-tracking metrics: false positive (FP) (%), the percentage of incorrect detections; false negative (FN) (%), the percentage of missed detections; ID-Swaps, incorrect identity reassignments; and appearing/disappearing errors (ADE) (%), the percentage of incorrectly identified cell appearing or disappearing events relative to all real appearing and disappearing events. For public benchmarks, we use the tracking accuracy (TRA) metric [18], which compares predicted lineage forests with reference data. TRA ranges from 0 to 1, reflecting the minimal graph edits needed to align predicted and ground-truth lineage graphs; higher scores indicate fewer linking or division errors.

Quantitative Analysis on Ultraplex Microscopic Images Table 1 presents a comparison of cell tracking methods on the Multiplex dataset, which includes challenges such as cell appearing/disappearing events, imaging artifacts, and registration misalignment. Our method achieves the lowest false negative rate, demonstrating its ability to detect newly appearing cells, while Bayesian-based and graph neural network (GNN)-based methods struggle in these scenarios. It also proves more resilient to artifacts, achieving the lowest false positive rate, as it leverages deep feature extraction and higher-order graph matching to better distinguish real cells from noise and imaging distortions. In terms of identity consistency, our approach records only 2 ID swaps, far lower than Hungarian-based (9) and Bayesian-based (6) methods, ensuring robust tracking even with registration misalignment. Additionally, it achieves the lowest ADE, reinforcing its ability to localize cells more precisely compared to other methods. With an overall accuracy of 0.958, it surpasses DLPM-Net and GNN-based tracking, confirming the advantages of integrating higher-order graph matching, Bayesian neural networks (BNN), and deep feature extraction in handling complex cell tracking challenges.

Quantitative Analysis on Public Datasets Table 2 presents a performance comparison of state-of-the-art cell tracking methods on public microscopy datasets using TRA scores. In Fluo-C2DL-MSC and PhC-C2DL-PSC, it surpasses all baselines, highlighting its effectiveness in phase-contrast and fluorescent microscopy, where intensity-based tracking often fails. A notable improvement is seen in Fluo-C2DL-Huh7, confirming its robustness in handling cell appearance variations. The only exception is Fluo-N2DL-HeLa, where Contrastive Tracking slightly leads. This may be due to HeLa cells' dense morphology, which benefits

Table 1: Performance comparison on the Ultraplex images. For FP, FN, ID-Swaps, ADE, lower is better, and for accuracy metrics, higher is better.

Method	FP (%)	FN (%)	ID-Swaps	ADE (%)	Accuracy
Bayesian-based [31]	7.9	13.5	6	11.8	0.812
Hungarian first-order	10.4	15.9	9	14.5	0.625
GNN-based [3]	5.2	10.2	4	8.5	0.886
DLPM-Net [33]	5.0	9.5	3	7.9	0.851
Proposed (Ours)	**4.1**	**8.8**	**2**	**6.3**	**0.958**

Table 2: Performance comparison on public microscopy datasets (TRA scores). The highest value in each row is in **bold**.

Dataset	Contrastive Tracking [35]	CMTT-JTracker [6]	Ultrack [5]	Proposed (Ours)
Fluo-C2DL-MSC	0.741	0.709	0.738	**0.756**
PhC-C2DL-PSC	0.927	0.863	0.932	**0.941**
Fluo-C2DL-Huh7	0.851	0.810	0.830	**0.894**
Fluo-N2DL-HeLa	**0.967**	0.919	0.930	0.925

from contrastive learning's strength in differentiating fine-grained variations. Despite this, our method remains highly competitive across all datasets, reinforcing its generalizability and ability to tackle challenges such as motion variations, cell division, and imaging artifacts more effectively than existing alternatives.

Ablation Study Table 3 presents an ablation study evaluating the effect of deep feature extraction and graph matching order on tracking accuracy. The results reveal that graph matching order has the most significant influence, as higher-order (HO) methods consistently outperform their first-order (FO) counterparts. Additionally, incorporating deep embeddings (DE) leads to an improvement across all methods, highlighting their role in optimizing tracking performance by providing a more robust representation of cell features. While the difference between BNN and discriminative transformer (DT) is relatively smaller compared to the impact of deep embeddings or graph order, both contribute to performance enhancement.

Table 3: Ablation study evaluating the impact of different model components on tracking accuracy with and without deep features.

Method	HO + DE	HO + w/o DE	FO + DE	FO + w/o DE
BNN	0.814	0.795	0.725	0.681
DT	0.847	0.801	0.738	0.714
Proposed (Ours)	**0.958**	0.910	0.863	0.832

4 Conclusion

In this work, we introduced a Bayesian transformer framework for cell tracking in multiplexed tissue sections, addressing the inherent challenges imposed by imaging artifacts, registration misalignments, and sample preparation variability. Our method integrates uncertainty-aware feature embeddings with higher-order graph matching and belief propagation, which reduces error propagation and improves the consistency of cell linkages. Evaluations on both a private multiplex dataset and public cell tracking benchmarks show that the proposed approach outperforms traditional tracking pipelines by achieving fewer missed or false detections, lower ID switches, and higher overall accuracy. Also, our approach requires no manually annotated tracking labels or supervised training examples. Instead, it relies entirely on feature embeddings derived directly from automated segmentations, reducing the need for extensive manual curation and facilitating generalizability across diverse imaging modalities. These findings show that leveraging higher-order consistency constraints, along with uncertainty modeling in deep embeddings, holds potential for enabling more reliable 3D tissue reconstructions in complex biological imaging settings.

Acknowledgments. This work is supported by the National Institutes of Health (NIH) under grant No. RF1MH130472, PI: Ostroff, Co-I: Nabavi.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andrey, P., Maurin, Y.: Free-d: an integrated environment for three-dimensional reconstruction from serial sections. Journal of neuroscience methods **145**(1-2), 233–244 (2005)
2. Bardinet, E., Ourselin, S., Dormont, D., Malandain, G., Tandé, D., Parain, K., Ayache, N., Yelnik, J.: Co-registration of histological, optical and mr data of the human brain. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002. pp. 548–555. Springer (2002)
3. Ben-Haim, T., Raviv, T.R.: Graph neural network for cell tracking in microscopy videos. In: European Conference on Computer Vision. pp. 610–626. Springer (2022)
4. Blum, H.: Biological shape and visual science (part i). Journal of theoretical Biology **38**(2), 205–287 (1973)
5. Bragantini, J., Theodoro, I., Zhao, X., Huijben, T.A., Hirata-Miyasaki, E., VijayKumar, S., Balasubramanian, A., Lao, T., Agrawal, R., Xiao, S., et al.: Ultrack: pushing the limits of cell tracking across biological scales. bioRxiv (2024)
6. Chen, L., Fu, S., Zhang, Z.: Cmtt-jtracker: a fully test-time adaptive framework serving automated cell lineage construction. Briefings in Bioinformatics **25**(6), bbae591 (2024)
7. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 539–546. IEEE (2005)
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. IEEE (2009)
9. Emami, N., Sedaei, Z., Ferdousi, R.: Computerized cell tracking: Current methods, tools and challenges. Visual Informatics **5**(1), 1–13 (2021)
10. Gallusser, B., Weigert, M.: Trackastra: Transformer-based cell tracking for live-cell microscopy. In: European Conference on Computer Vision. pp. 467–484. Springer (2024)
11. Hamzehei, S., Raimondi, G., Karami, M., Ostroff, L., Nabavi, S.: Capture: A clustered adaptive patchwork technique for unified registration enhancement in biological imaging. In: Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. pp. 1–6 (2024)
12. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics (6), 610–621 (1973)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv:1312.6114 (2013)
15. Kuhl, F.P., Giardina, C.R.: Elliptic fourier features of a closed contour. Computer Graphics and Image Processing **18**(3), 236–258 (1982)
16. Magnusson, K.E., Jaldén, J.: A batch algorithm using iterative application of the viterbi algorithm to track cells and construct cell lineages. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 382–385. IEEE (2012)
17. Maška, M., Ulman, V., Delgado-Rodriguez, P., Gómez-de Mariscal, E., Nečasová, T., Guerrero Peña, F.A., Ren, T.I., Meyerowitz, E.M., Scherr, T., Löffler, K., et al.: The cell tracking challenge: 10 years of objective benchmarking. Nature Methods **20**(7), 1010–1020 (2023)

18. Matula, P., Maška, M., Sorokin, D.V., Matula, P., Ortiz-de Solórzano, C., Kozubek, M.: Cell tracking accuracy measurement based on comparison of acyclic oriented graphs. PLOS ONE **10**(12), e0144959 (2015)
19. McDole, K., Guignard, L., Amat, F., Berger, A., Malandain, G., Royer, L.A., Turaga, S.C., Branson, K., Keller, P.J.: In toto imaging and reconstruction of post-implantation mouse development at the single-cell level. Cell **175**(3), 859–876 (2018)
20. O'Connor, O.M., Dunlop, M.J.: Cell-tractr: A transformer-based model for end-to-end segmentation and tracking of cells. bioRxiv pp. 2024–07 (2024)
21. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1), 51–59 (1996)
22. Pérez-Garza, J., Orea, J., Deane, Z., Raimondi, G., Tripp, R., Charles, I., Ostroff, L.: Ultraplex microscopy: versatile highly-multiplexed molecular labeling and imaging across scale and resolution. bioRxiv (2024)
23. Pichat, J., Iglesias, J.E., Yousry, T., Ourselin, S., Modat, M.: A survey of methods for 3d histology reconstruction. Medical Image Analysis **46**, 73–105 (2018)
24. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: International Conference on Machine Learning. pp. 1278–1286. PMLR (2014)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015. pp. 234–241. Springer (2015)
26. Song, Y., Treanor, D., Bulpitt, A.J., Magee, D.R.: 3d reconstruction of multiple stained histology images. Journal of Pathology Informatics **4**(2), 7 (2013)
27. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature Methods **18**(1), 100–106 (2021)
28. Sturm, M., Cerrone, L., Hamprecht, F.A.: Syncellfactory: Generative data augmentation for cell tracking. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 304–313. Springer (2024)
29. Teague, M.R.: Image analysis via the general theory of moments. JOSA **70**(8), 920–930 (1980)
30. Torresani, L., Kolmogorov, V., Rother, C.: Feature correspondence via graph matching: Models and global optimization. In: Computer Vision—ECCV 2008. pp. 596–609. Springer (2008)
31. Ulicna, K., Vallardi, G., Charras, G., Lowe, A.R.: Automated deep lineage tree analysis using a bayesian single cell tracking approach. Frontiers in Computer Science **3**, 734559 (2021)
32. 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**(12), 1141–1152 (2017)
33. Xie, Y., Liu, M., Zhou, S., Wang, Y.: A deep local patch matching network for cell tracking in microscopy image sequences without registration. IEEE/ACM Transactions on Computational Biology and Bioinformatics **19**(6), 3202–3212 (2021)
34. Xue, B., Yu, J., Xu, J., Liu, S., Hu, S., Ye, Z., Geng, M., Liu, X., Meng, H.: Bayesian transformer language models for speech recognition. In: ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7378–7382. IEEE (2021)
35. Zyss, D., Sharma, A., Ribeiro, S.A., Repellin, C.E., Lai, O., Ludlam, M.J., Walter, T., Fehri, A.: Contrastive learning for cell division detection and tracking in live cell imaging data. bioRxiv pp. 2024–08 (2024)