amU
Aix Marseille Université

NNT : 2024AIXM0001

THÈSE DE DOCTORAT

Soutenue à AMU — Aix-Marseille Université
le 12 mai 2025 par

Jules VANARET

Exploring tissue patterning and stochastic cellular dynamics in photon-limited imaging of light-scattering organoids:
combined deep learning and bayesian strategies

Discipline
Informatique

École doctorale ED184
Mathématiques et Informatique

Laboratoires
I2M (UMR 7373)
IBDM (UMR 7288)
Institut Fresnel (UMR 7249)
Turing Centre for Living Systems

Composition du jury

David ROUSSEAU Professeur, Univ. d'Angers, IRHS	Rapporteur
Nicolas LOMÉNIE Professeur, Univ. Paris Cité, LIPADE	Rapporteur
Hervé ISAMBERT DR CNRS, Institut Curie	Examinateur & Président du jury
Daniel SAGE IR, EPFL, Laboratoire d'imagerie biomédicale	Examinateur
Frédéric RICHARD Professeur, Aix-Marseille Univ., I2M	Directeur de thèse
Philippe ROUDOT CR CNRS, Aix-Marseille Univ., Institut Fresnel	Co-directeur de thèse
Sham TLILI CR CNRS, Aix-Marseille Univ., IBDM	Co-encadrante de thèse
Pierre-François LENNE DR CNRS, Aix-Marseille Univ., IBDM	Membre invité

INSTITUT DE MATHÉMATIQUES DE MARSEILLE
IBDM
INSTITUT FRESNEL
CENTURI TURING CENTRE FOR LIVING SYSTEMS

Affidavit

Je soussigné, Jules Vanaret, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Frédéric Richard, Philippe Roudot et Sham Tlili, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte AMU relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 14 janvier 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse:

1. VANARET, Jules, DUPUIS, Victoria, LENNE, Pierre-François, et al. A detector-independent quality score for cell segmentation without ground truth in 3D live fluorescence microscopy. IEEE Journal of Selected Topics in Quantum Electronics, 2023, vol. 29, no 4: Biophotonics, p. 1-12.
2. GROS, Alice, VANARET, Jules, DUNSING-EICHENAUER, Valentin, et al. A quantitative pipeline for whole-mount deep imaging and multiscale analysis of gastruloids. bioRxiv, 2024, p. 2024.08. 13.607832.

Participation aux conférences et écoles d'été au cours de la période de thèse:

1. Physics of Living Matter 2022, Marseille
2. Gretsi 2023, Grenoble
3. Summer School "Advanced Methods of Biomedical Image Analysis" 2023, Brno (Tchéquie)
4. Eusipco 2024, Lyon
5. Rencontres du GDR IMABIO 2024, Grenoble
6. EPFL LATSIS "Symposium on smart microscopy" 2024, Lausanne (Suisse)

Résumé et mots clés

Cette thèse explore la modélisation biophysique et computationnelle de l'auto-organisation dans les gastruloïdes, des modèles tridimensionnels in vitro du développement embryonnaire. Elle étudie comment les décisions cellulaires émergent d'interactions mécano-génétiques à l'échelle de la cellule individuelle, en s'appuyant sur des techniques avancées de microscopie, d'analyse d'images et de suivi pour quantifier les comportements cellulaires et les signaux du microenvironnement.

La première partie présente un pipeline méthodologique visant à quantifier les propriétés mécaniques et génétiques dans les tissus profonds des gastruloïdes. Une stratégie de segmentation basée sur l'apprentissage profond, adaptée aux formes hétérogènes des noyaux cellulaires, est alors développée. Celle-ci intègre l'amélioration du contraste et la fusion à double vue afin d'accroître la profondeur et la précision de l'imagerie.
En s'appuyant sur ces avancées en segmentation, l'étude explore le paysage mécano-génétique des gastruloïdes par des analyses quantitatives intégrant l'expression génique et la déformation cellulaire. Un pipeline computationnel dédié, basé sur Python, est introduit afin de permettre une analyse multi-échelle ainsi qu'une visualisation conviviale des résultats.

Dans la deuxième partie, les défis liés au suivi cellulaire dans les gastruloïdes sont évalués, en montrant comment la stochasticité et l'incertitude de mesure compliquent la reconstruction des lignées cellulaires. Différentes méthodologies de suivi utilisées en biologie du développement ou spécifiquement pour le suivi cellulaire sont passées en revue, menant au développement d'un cadre de filtrage bayésien permettant de quantifier l'incertitude de suivi à plusieurs échelles.
Enfin, ces méthodes probabilistes sont appliquées pour valider les algorithmes de segmentation et de suivi, illustrant comment l'approche bayésienne permet de guider le réglage non supervisé des paramètres ainsi que le classement des algorithmes de segmentation. À travers de nombreuses simulations et applications expérimentales, cette étude propose des méthodes rigoureuses facilitant l'analyse de la dynamique cellulaire dans des jeux de données complexes issus des gastruloïdes, avec des implications pour la biologie du développement et l'ingénierie tissulaire.

Mots clés: organoïde embryonnaire, suivi de trajectoires cellulaires, imagerie en temps réel, hétérogénéité cellulaire, apprentissage profond, filtrage bayésien

Abstract and keywords

This thesis explores the biophysical and computational modeling of self-organization in gastruloids, three-dimensional in vitro models of embryonic development. It investigates how cellular decisions emerge from mechanogenetic interactions at the single-cell level, leveraging advanced microscopy, image analysis, and tracking methodologies to quantify cellular behaviors and micro-environmental cues.

The first part presents a methodological pipeline to quantify mechanical and genetic properties in deep gastruloid tissues. A deep-learning-based segmentation strategy tailored for heterogeneous nuclei shapes is then developed, integrating contrast enhancement and dual-view fusion to improve imaging depth and accuracy.
Building on these segmentation advances, the study probes the mechanogenetic landscape of gastruloids, integrating quantitative analyses of gene expression and cellular deformation. A dedicated Python-based computational pipeline is introduced, enabling multiscale analysis and user-friendly visualization.

In the second part, the challenges of cell tracking in gastruloids are assessed, presenting how stochasticity and measurement uncertainty complicate lineage reconstruction. Various tracking methodologies applied in the context of cell tracking or developmental biology are reviewed, leading to the development of a Bayesian filtering framework for the multiscale quantification of tracking uncertainty.
The final sections apply these probabilistic methods to validate segmentation and tracking algorithms, demonstrating how the Bayesian filtering algorithm can guide unsupervised parameter tuning and segmentation algorithm ranking. Through extensive simulations and experimental applications, the study provides rigorous methods to aid cellular dynamics analysis in complex gastruloid datasets, with implications for developmental biology and tissue engineering.

Keywords: embryonic organoid, cell tracking, live imaging, cellular heterogeneity, deep-learning, bayesian filtering

Remerciements

En premier lieu, je souhaite remercier chaleureusement les membres du jury pour leur relecture attentive et leurs précieux conseils. Leur regard critique et constructif a été essentiel pour améliorer cette thèse. Je suis particulièrement reconnaissant envers David Rousseau et Nicolas Loménie, qui ont immédiatement accepté le travail titanesque qu'est celui de rapporteur. Je remercie également Hervé Isambert d'avoir accepté de présider cette soutenance, ainsi que Daniel Sage pour sa présence.
Je suis reconnaissant pour votre bienveillance et vos questions pertinentes qui ont enrichi la discussion lors de la soutenance.

Je suis également infiniment reconnaissant envers l'ensemble des gens qui m'ont encadré pour que je mène à bien cette thèse.
Je tiens à remercier les membres de mon comité de suivi de thèse, Pierre Ronceray et Arnauld Sergé, pour leurs conseils avisés et leur accompagnement tout au long de ce parcours.
Je suis particulièrement reconnaissant envers Pierre-François pour m'avoir accueilli dans cette équipe si attachante qui est la sienne. Merci pour tes conseils, tes livraisons hebdomadaires du meilleur pain de la région et pour m'avoir présenté, je cite, "le plus beau campus du monde".
Mille mercis à mes directeurs de thèse (soit environ 333 mercis par directeur). Merci d'abord à Frédéric Richard pour avoir été la figure de stabilité qui m'a permis de retrouver le cap à chaque fois que j'en avais besoin. Merci pour ton calme, ton recul, et la confiance que tu as eu à l'égard de mon travail. Un grand merci à Sham Tlili pour ses idées novatrices, ses passions communicatives, que ce soit pour le culte de la visco-élasto-plasticité ou pour les massifs montagneux autour de Marseille. Merci pour toutes les heures que tu as passées en salles de culture et d'imagerie pour que j'ai les meilleures données possibles. Merci aussi pour toutes ces chouettes balades et moments familiaux. Je remercie également Philippe Roudot pour la liberté qu'il m'a laissé dans mon travail, et dans la confiance totale qu'il a eu en mes idées, et qui m'a permis de me tromper beaucoup pour mieux me trouver. Merci pour tes bonnes idées, et ta capacité à activement remettre en question tes méthodes pour coller au mieux aux modes de fonctionnement de tes étudiants. Merci aussi pour tous ces bons conseils de restaurants et ces discussions autour de la cuisine.
Merci à tous les trois pour votre complémentarité, votre bienveillance et vos relectures méticuleuses du manuscrit.

Je souhaite également remercier chaleureusement mes collègues et les équipes avec lesquelles j'ai eu le plaisir de collaborer. Je suis reconnaissant envers Jean-François Rupprecht et Léo Guignard pour nos collaborations intéressantes, qui ont donné un sens supplémentaire à mon travail.
Un merci spécial à Daniel Sapède de la plateforme de microscopie de l'IBDM, "l'homme qui murmurait à l'oreille des microscopes", qui a de maintes fois dépanner des appareils qui ont joué un rôle dans ma thèse.
Je suis également reconnaissant envers l'équipe Signal et Image de l'I2M pour leur accueil. Un merci particulier à Piyush pour son enthousiasme et pour nos discussions passionnantes, tant sur le plan scientifique que culturel. Je remercie l'équipe Lenne, et plus particulièrement Pierre, éternel ronchon grâce à qui j'ai découvert tant de bons produits, et Claire pour

m'avoir motivé à reprendre la course à pied et pour m'avoir formé en escalade avec beaucoup de bienveillance. Merci aussi à Raphaël, Valentin, Leda, Vanessa, Kenji, Alessandro, Jialing, Lukas, Victoria, Dalia, Ornella, Marie, et tous les stagiaires qui sont passés dans le labo pendant ma thèse, pour leur bonne humeur, les bons moments partagés, et leur soutien.

Je tiens à exprimer ma gratitude envers le personnel administratif de l'Institut Centuri, notamment Pauline, Rémi et Mélina pour l'organisation de retraites et d'afterworks mémorables, et Marlène et Jasmina pour leur efficacité et leur patience dans l'organisation des missions. Mention spéciale à Marlène qui a toujours été d'une patience infinie pour m'aider à naviguer à travers les embûches de Notilus. Je suis conscient de la chance que nous avons d'avoir un staff aussi dévoué.
Je remercie également le personnel administratif de l'I2M et de l'IBDM, notamment Sébastien Poulain, pour m'avoir ouvert toutes les portes (littéralement).
Je remercie aussi le monsieur à la caisse du Crous pour avoir toujours illuminé nos midis par sa bonne humeur et ses expressions chantantes.

Merci à tous les profs qui ont attisé ma passion pour la physique, l'informatique, ou la science en général. Merci à M. Laurent, Mme. Bronquard, Mme. et M. Aubert, M. Gaidet, Julien Ségala, François Giraud, Béatrice Ducret, Henning Samtleben, Ruben Staub, Christophe Bailly, Alain Dougnac, Nicolas Hourcade, et bien d'autres.

Je tiens à remercier tous ces amis avec qui j'ai partagé des moments qui me sont chers. Merci aux membres du groupe de vélo, Colin, Robin, Sophie, Maxime, Baptiste et Baptiste (le sg de la vn), quel plaisir de bombarder sur nos belles routes à vos côtés pour dévaliser des boulangeries.
Je remercie également tous les gens de ma période ENS, Arthur et Francis les OGs de la Tacoloc qui sont venus me remonter le moral lorsque j'écrivais les dernières pages de ce manuscrit, Jules (qui a toujours été le Jules que j'aimerais devenir quand je serai grand), Raphaël, Louis, Frédéric, Clément, Guillaume, Jeanne, Manon, Gauthier, Dop, Benoit, François et j'en oublie forcément...
Merci également aux "maniaques dans la salle", Damien et Etienne, ainsi qu'à Léa, et à Snorri notre mascotte ronronnante, qui dormait sur mes genoux quand j'écrivais les premières lignes de ce manuscrit.
Merci aux "andouilles" de Centuri, Marc-Eric et Andonis, pour avoir fait de leur appartement le QG de la bande, ainsi qu'à Martin pour être l'éternel premier année qui nous fait bénéficier de sa précieuse bonne humeur. Quel plaisir j'ai eu de vous retrouver autour d'un plat de qualité variable au Crous.
Je pense également à mes amis de longue date, Salwa, Gwendoline, qui m'ont toujours abreuvé de memes de qualité sans que je ne leur rende au centième, et Jean-Étienne, "loin des yeux près du coeur", pour l'impact qu'il a eu sur ma vie.

Enfin, je remercie ma famille, notamment pour m'avoir toujours laissé en paix quand j'avais besoin de travailler. Je pense à mes grands-parents, Jacques, Françoise et Renée, qui ne pourront pas lire ces mots, et tout particulièrement Françoise, qui aurait été ravie d'apprendre que j'ai commencé le piano. À Jean, qui a déjà placé ce manuscrit à côté de celui de mon frère.

Je remercie mes parents et mes frères pour leur soutien constant et pour m'avoir donné des bases solides sur lesquelles je me construis encore aujourd'hui.

Et enfin merci à Alice, pour tout.

Table of Contents

Table of Contents 9

List of Figures 13

List of Tables 17

General Introduction 19
- 0.1 Stem cells functions emerge from stochastic processes . . . 19
- 0.2 Morphogenesis is robust despite cell stochasticity . . . 20
- 0.3 Gastruloids: a model system to study self-organization in vitro . . . 22
- 0.4 Key research questions require cell scale comprehension of mechanogenetic interplay . . . 24
  - 0.4.1 Coarse-grained biophysical models of gastruloids morphogenesis exist but do not reach cell-scale resolution . . . 24
  - 0.4.2 Single cell tracking is necessary to bridge tissue-level organization and single-cell decision-making . . . 28
- 0.5 Two-photon microscopy allows deep-imaging of gastruloids . . . 29
  - 0.5.1 Basics of fluorescence microscopy . . . 29
  - 0.5.2 Comparison of advanced fluorescence microscopy techniques . . . 29
  - 0.5.3 Two-photon microscopy dominates for deep imaging . . . 30
  - 0.5.4 Imaging setups for fixed and live gastruloid imaging . . . 32
- 0.6 Image analysis challenges in gastruloid imaging . . . 32

Objectives of this thesis 37

Thesis overview 37

I Fixed datasets 39

1 Detecting heterogeneous nuclei shape in packed 3D gastruloids requires tailored neural-network training and validation strategy 41
- 1.1 Chapter overview . . . 41
- 1.2 Motivations and Challenges . . . 42
- 1.3 Methods . . . 43
  - 1.3.1 Preprocessing . . . 43
  - 1.3.2 Segmentation using a deep-learning algorithm . . . 45

1.4 Results . . . 50
1.4.1 Dual-view fusion allows imaging nuclei *in toto* in fixed samples . . 50
1.4.2 Thanks to a tailored preprocessing pipeline, a deep-learning algorithm allows nuclei segmentation in deep gastruloids . . . 50
1.5 Discussion . . . 51

2 Probing cellular micro-environment: integrating nuclei segmentation with mechanics and gene expression analysis 55
2.1 Chapter overview . . . 55
2.2 Motivations and Challenges . . . 56
2.3 Methods . . . 56
2.3.1 Comparing fields of mechano-genetic properties at different scales 56
2.3.2 Measuring gene expression from biased fluorescence intensity . . . 58
2.3.3 Quantifying cell deformation . . . 62
2.4 Results . . . 64
2.4.1 Nuclei deformations bear spatial signatures of internal mechanical stresses . . . 64
2.4.2 Gene expression can be quantified in deep gastruloids despite imaging artifacts by leveraging a model of fluorescence scattering . . . 67
2.5 The *Tapenade* software suite brings interactive and user-friendly tools to empower researchers . . . 75
2.5.1 The Python package . . . 75
2.5.2 The Napari plugins . . . 75
2.5.3 User-friendliness of the tools . . . 79
2.6 Discussion . . . 79
2.6.1 Multiscale imaging and analysis pipeline . . . 79
2.6.2 Exploring gastruloid development in 3D . . . 80
2.6.3 A Python integrated pipeline for non-specialists, with Napari-enabled 3D exploration . . . 82

II Live datasets 85

3 Cell tracking strategies in developmental biology: challenges associated to stochastic dynamics and uncertain measurements 87
3.1 Chapter overview . . . 87
3.2 Challenges in cell tracking strategies for developmental biology . . . 87
3.2.1 Aleatoric and epistemic uncertainties in gastruloid cell observation 88
3.2.2 Features required by an effective tracking strategy for cell lineage reconstruction in gastruloids . . . 88
3.3 Multiple particle tracking without epistemic uncertainty: case-study on canonical Brownian dynamics . . . 89
3.3.1 Maximum-likelihood solution for multiple Brownian particles tracking without measurement uncertainty . . . 89
3.4 The intrinsic limitations of diffusive dynamics, framerate, and density . . . 91
3.5 The challenge of solving the assignment problem with lineage constraints 92
3.5.1 Early strategies based on frame-to-frame assignment . . . 92

3.5.2 Integer linear programming for frame-to-frame assignments 93
3.5.3 Generalizing integer linear programming for global enforcement of biological lineage constraints and multiple-candidate assignment . 96
3.5.4 Designing robust linking weights 99
3.6 The challenge of modeling cell dynamics 99
3.6.1 Heuristic cost definitions 99
3.6.2 Learning from manually annotated tracks 101
3.6.3 The case of deep methods 101
3.6.4 Direct probabilistic modeling via state-space models 102
3.6.5 Summary 103
3.7 The challenge of direct visualization in cell tracking strategies 103
3.7.1 All tracking methods require parameter tuning based on some form of visualization 103
3.7.2 State of the art 103
3.7.3 Our custom visualization tool allows interactive and modular exploration in 3D and 2D views 105
3.7.4 The role of manual editing of tracks 105
3.7.5 Discussion 107

4 Bayesian filtering provides a probabilistic framework for unsupervised uncertainty quantification 109
4.1 Chapter overview 109
4.2 Bayesian tracking to solve complex and heterogeneous dynamics 109
4.2.1 Motivation 109
4.2.2 Bayesian filtering with hidden Markov models 110
4.2.3 Kalman filtering applied to a single particle 112
4.2.4 Relevance to the simulation model of 5.4.3.1 113
4.2.5 Bayesian filtering in the multiple-hypothesis context 114
4.3 Trackability: probing local tracking ambiguity to measure global performance 115
4.3.1 Trackability as an estimate for multiple hypothesis likelihood . . . 116
4.3.2 Trackability is a versatile measure 117
4.4 Limitations of this framework for direct cell lineage reconstruction in gastruloids 118
4.4.1 Robustness to dynamical heterogeneity 118
4.4.2 Robustness to segmentation uncertainty and cell lineage constraints enforcement 118
4.4.3 Calibrating the required priors 120
4.5 Conclusion 120

5 Application: Trackability allows unsupervised ranking of segmentation algorithm with minimal priors 123
5.1 Chapter overview 123
5.2 Motivations 124
5.3 State of the art 126
5.4 Methods 126
5.4.1 Bayesian filtering based on hidden Markov models 126
5.4.2 Search radius definition 127

5.4.3 Simulation of cellular motion dynamics . . . 127
5.4.4 Simulation of point-like segmentation errors . . . 132
5.5 Results . . . 133
5.5.1 Validation using particle simulations . . . 133
5.5.2 Application to experimental data . . . 135
5.6 Discussion . . . 140

6 Application: Trackability allows unsupervised calibration of Bayesian tracking algorithms 143
6.1 Chapter overview . . . 143
6.2 Motivations . . . 143
6.3 Methods . . . 144
6.3.1 Particle simulations . . . 144
6.3.2 Penalized trackability as a robust score for hyperparameter optimization . . . 144
6.4 Results . . . 146
6.4.1 Penalized trackability guides optimal process noise in Brownian dynamics simulations . . . 146
6.4.2 Penalized trackability shows promising results in real gastruloid datasets . . . 147
6.5 Discussion . . . 149

Conclusion 151

Perspectives 153

Appendices 157

References 167

List of Figures

1 Gastruloids are in vitro models composed of embryonic stem cells that self-organize while undergoing differentiation. 23
2 Early polarization and elongation of gastruloids can be captured by a physical model where spatial gradients in tissue properties drive morphological changes. 26
3 Biophysical models of morphogenesis in gastruloids do not reach cell-scale resolution. 27
4 Illustration of the differences between confocal, lightsheet and two-photon microscopy in terms of optical setup and 3D excitation profile. 31
5 Overview of raw fixed and live typical datasets. 33
6 Challenges of image analysis. 36

1.1 Outline of the training and inference procedures of the Stardist3D algorithm. 47
1.2 Training procedure of our custom StarDist3D. 48
1.3 2D annotations can be used to estimate a close lower bound to the performances of StarDist3D. 50
1.4 The spatial registration and fusion procedure. 51
1.5 Dual-view fusion and local contrast enhancement proved crucial to improve performance in the deepest planes. 52
1.6 Perfomance of our custom StarDist3D model. 52
1.7 Illustrative examples of the performances of our custom StarDist3D model. 54

2.1 Illustration of the intensity model and of the normalization procedure based on a ubiquitous signal. 59
2.2 The wavelength-dependent normalization method. 60
2.3 Masked gaussian smoothing reveals patterning of properties related to cell packing from cellular to tissue scale. 65
2.4 Nuclei align along a stereotyped boundary between low- and high-density domains. 66
2.5 Reproducibility of nuclei alignment along density boundary. 67
2.6 Intensity normalization retrieves homogeneous Hoechst spatial distribution despite intensity artifacts 68
2.7 Wavelength-dependence of intensity artifacts is mitigated using a physical model of scattering. 69
2.8 Intensity normalization is required to measure spatial patterning of gene expression. 71
2.9 Intensity normalization reveals spatial correlation features in key genes. . 72
2.10 Gastruloid exhibit radial expression patterns for T-Bra and FoxA2. 74

2.11 Interactive Napari plugins for bioimage handling, manual registration, preprocessing, data exploration and analysis. 76
2.12 The Tapenade software suite was designed with user-friendliness in mind. 79
2.13 Overview of the experimental and computational pipeline for *in toto* imaging of organoids at cellular resolution and multiscale analyses. 81

3.1 Illustration of the tracking complexity as a function of the adimensional number ℓ_{dyn}/ℓ_{nn} 92
3.2 Napari visualization tool overview. 106

4.1 Hierarchical schematics of the hidden Markov model used in Kalman filtering. 111
4.2 Trackability as a Monte Carlo estimate of assignment stability in a Bayesian filtering framework. 117
4.3 Roudot et al. [75] showed that trackability is a multiscale mesure of tracking ambiguity and quality. 119

5.1 The density and heterogeneity of cellular objects challenge the comparison and validation of leading-edge segmentation approaches based on deep neural networks. 125
5.2 Breakdown of each component of the simulation, and their effect on the simulated dynamics. 129
5.3 A simple particle simulation framework allows mimicking typical nuclear dynamics observed in gastruloids. 130
5.4 Illustration of the different types of errors introduced in the simulations. . 133
5.5 Distribution of frame-to-frame displacement and nearest neighbor distance in our simulation 135
5.6 Our evaluation score responds strongly to simulated merging, splitting and false positive errors and is sensitive to false negatives in scenarios with mixed error types. 135
5.7 Visual example of typical Stardist3D and Cellpose response on 3D live carcinoma cells data. 137
5.8 Our evaluation score can accurately rank all models studied on the carcinoma cells dataset. 137
5.9 Visual example of typical Stardist3D and Cellpose response on challenging 3D live organoid cells data. 139
5.10 Our evaluation score can accurately rank three of the four models studied on the organoid dataset. 140

6.1 In Brownian simulations, trackability can be used to optimize for process noise with ground-truths. 147
6.2 Optimization of process noise with trackability is qualitatively validated on real gastruloid datasets. 148
6.3 Visualization of difference between ground-truth and candidate segmentation at 50% IoU. 160
6.4 Comparison of immunostained T-Bra and endogenous T-Bra reporter signals in gastruloids fixed at 72 h. 162

6.5 Hierarchical schematics of the hidden Markov model used in the PMMS approach. 164

List of Tables

1 Comparison of key features between confocal microscopy, light sheet microscopy, and two-photon microscopy for deep organoid and gastruloid imaging. 30

1.1 Description of the datasets used for training our custom StarDist3D model to detect nuclei. 46

2.1 Preprocessing steps with corresponding typical computation times 78

General introduction

Understanding how complex organisms develop from a single fertilized cell is a central question in developmental biology. This field explores the intricate processes of cell proliferation, differentiation, and self-organization into functional tissues and organs, which are influenced not only by genetic instructions but also by dynamic mechanical interactions within the cellular environment [1]. In embryos, morphogenesis results from the interplay between genetic patterning and biomechanical forces, determining tissue architecture. A key challenge is how cells robustly self-organize despite intrinsic noise and external perturbations. In this thesis, I have developed methodologies aimed at advancing the quantification of individual dynamic cellular behavior within complex 3D tissues during morphogenesis.

0.1 Stem cells functions emerge from stochastic processes

Stem cells possess the unique ability to proliferate and differentiate into various cell types in response to diverse chemical and mechanical signals within biological tissues [2]. Found throughout the body, from developing embryos to adult tissues, they play a crucial role in building, maintaining, and repairing organs. In humans, stem cells are essential for replenishing constantly renewing tissues such as blood and skin.

Despite their essential role in maintaining organismal health, stem cell behavior exhibits inherent stochasticity at multiple levels [3]. This stochasticity arises from both variability in gene expression [4] and randomness in cellular movements [5], highlighting that stem cell biology operates as a complex probabilistic system rather than a purely deterministic one. Understanding this stochastic nature is fundamental to both biology and medicine, as it influences embryonic development, tissue regeneration, and disease progression [6].

At the level of gene expression, stochasticity arises from the intrinsic noise inherent to transcriptional and translational processes. Gene expression in stem cells is governed by complex regulatory networks subject to fluctuations [7]. Single-cell transcriptomic analyses have revealed that stem cells within a seemingly homogeneous population exhibit significant variability in gene expression profiles [8, 9]. This heterogeneity is partly due to the low copy number of key regulatory molecules, which amplifies transcriptional noise [3]. Such variability can result in transient fluctuations in lineage-specific gene expression, pushing certain cells to differentiate while allowing others to retain pluripotency. Im-

portantly, this stochastic activation and repression of genes do not occur in isolation but are influenced by extrinsic factors, including signaling from the microenvironment of the cells [10].

Beyond gene expression, stochasticity also governs cell motion and shape dynamics. Time-lapse imaging studies have demonstrated that stem cell migration, division orientation, and shape fluctuations are not uniform but rather exhibit probabilistic patterns [11, 12]. These behaviors are driven by the stochastic assembly and disassembly of cytoskeletal components[13], dynamic cell-cell and cell-matrix interactions [14], and fluctuating signaling gradients within the niche. Similarly, mesenchymal stem cells exhibit unpredictable migratory trajectories [15], which, rather than being purely random, reflect a balance between stochastic fluctuations and guided movement toward biochemical and biomechanical cues.

0.2 Morphogenesis is robust despite cell stochasticity

Morphogenesis is the process by which tissues and organs acquire their characteristic shape. It is a fundamental step of embryo development, guiding the transformation of undifferentiated stem cells into organized structures [16]. Despite the inherent stochasticity of cellular behavior and gene expression described in the previous section, morphogenetic processes are remarkably robust. A key factor in this robustness is mechano-genetic coupling, where genetic and mechanical cues jointly regulate cell fate and tissue organization [6].

Traditionally, embryo morphogenesis has been understood through genetic regulation, with molecular signaling pathways directing cell differentiation, proliferation, and cellular positionning and shape. Reaction-diffusion models and gene regulatory networks have provided theoretical frameworks for understanding how genetic factors establish tissue patterning. Classic studies have demonstrated that positional information emerges from diffusible morphogens, which are signaling molecules that can be both produced and received by cells, and form gradients that instruct cells on their fate based on their location within a developing structure [17]. Reaction-diffusion models provide a coarse-grained model of how the spatial distribution of morphogens evolve by treating morphogens and tissues in a continuum fashion. The dynamics of a morphogen concentration $c(\mathbf{x}, t)$ are typically governed by the reaction-diffusion equation

$$\frac{\partial c}{\partial t} = D\nabla^2 c + R(c), \tag{1}$$

where D is the diffusion coefficient, governing the transport property of the morphogen through the tissue, and $R(c)$ represents reaction terms that model morphogen production, degradation, or nonlinear interactions. Early models, such as Turing's seminal work on pattern formation, proposed that self-organized spatial heterogeneity could arise from an activator-inhibitor system, where $R(c)$ takes the form of coupled equations for short-range activation and long-range inhibition [18]. These mechanisms provide a fundamental framework for understanding periodic structures in embryonic development, including limb bud formation and digit patterning [19, 20].

In addition, tissue patterning can arise through reaction-diffusion mechanisms where cell population dynamics replace traditional morphogen-based equations. In this framework, diffusion emerges from cell migration and rearrangements within tissues, while reactive terms capture competitive interactions between distinct cell types. Experimental and theoretical studies demonstrate that differential adhesion strengths, motility rates, and mechanical interactions between cell populations can self-organize periodic patterns through Turing-like instabilities [21, 22]. For example, during vertebrate embryogenesis, mesenchymal cell populations with varying adhesion properties sort into spatially organized domains through tension-driven redistribution, effectively behaving as activator-inhibitor pairs in a cellular reaction-diffusion system. This cell-centric mechanism bypasses explicit morphogen gradients while maintaining key Turing system features: local self-enhancement through cell clustering and long-range inhibition via mechanical displacement or density-dependent motility constraints [22]. Such models successfully explain pattern formation in tissues like the zebrafish lateral line system and mammalian blastocyst inner cell mass organization, where cell-type proportions and migration dynamics dictate spatial arrangements [22].

Beyond genetic regulation, the spatio-temporal patterning of mechanical forces is essential for organogenesis. Cells are typically part of cohesive tissues, a physical environment where mechanical stresses arise from cell adhesion, cytoskeletal tension, and interactions with the extracellular matrix. Emergent mechanical stresses are now recognized as central players in tissue morphogenesis, driving changes in cell shape, size, proliferation, and movement. Indeed, as an example, mechanical stresses can influence the orientation of cell division, and conversely, oriented cell divisions can generate mechanical stresses [23]. At the tissue level, processes such as epithelial folding and tissue elongation result from mechanical instabilities and force-driven organization [24, 25], for which continuum mechanics provides a framework to describe these behaviors quantitatively.

In biological tissues, mechanical equilibrium is typically governed by the conservation of momentum in the overdamped regime, where inertial effects are negligible due to low Reynolds numbers. A minimal expression of this formalism is

$$\nabla \cdot (\boldsymbol{\sigma} + \boldsymbol{\sigma}_a) + \mathbf{f} = \mathbf{0}, \tag{2}$$

where $\boldsymbol{\sigma}$ is the Cauchy stress tensor, $\boldsymbol{\sigma}_a$ is an active stress term due to the spontaneous organization of the tissue, and $\mathbf{f}$ represents external body forces, such as frictional interactions with the environment (such as extraembryonic tissues). In a continuum-mechanical framework, active stress is often modeled as an additional stress component that arises from cellular activity, distinct from passive stresses such as viscous or elastic stresses. A common form for $\mathbf{f}$ is a frictional force $\mathbf{f} = -\zeta \mathbf{v}$, where ζ is a coefficient of friction with an external substrate, and $\mathbf{v}$ is the velocity. To capture the viscoelastic nature of tissue dynamics, the stress-strain relationship can be modeled usually using a Maxwell-type material, where stress relaxes over time according to

$$\dot{\boldsymbol{\sigma}} + \frac{E}{\eta}\boldsymbol{\sigma} = E\dot{\boldsymbol{\varepsilon}}, \tag{3}$$

where E is the tissue elasticity (which results from the interplay between cell cortical ten-

sion, adhesion and potentially from cytoskelatal and nuclei properties), η is the emergent tissue viscosity (which results from plastic events at the cellular scale such as cell division or rearrangements), and $\dot{\boldsymbol{\varepsilon}}$ is the strain rate tensor $\dot{\boldsymbol{\varepsilon}} = \frac{1}{2}\left(\nabla \mathbf{v} + \nabla \mathbf{v}^{\mathrm{T}}\right)$ with velocity $\mathbf{v}$. In the purely viscous limit, $\dot{\boldsymbol{\sigma}}$ vanishes and the stress is directly proportional to the strain rate with $\sigma = \eta \dot{\boldsymbol{\varepsilon}}$.

Active stress in cellular systems can be represented through distinct mechanisms that capture various aspects of cellular mechanics, integrating myosin activity, cytoskeletal organization, and adhesion dynamics. As a first mechanism, contractile stress can be modeled as an isotropic stress $\boldsymbol{\sigma}_a = \alpha c_m$, where it is directly proportional to myosin concentration (c_m) in cells. A second mechanism could involve anisotropic stresses arising from cell shape anisotropies, where actin stress fibers or microtubules orient along a primary cellular axis. These anisotropies are encoded in the nematic tensor Q, reflecting cell deformation, yielding $\boldsymbol{\sigma}_a = \beta Q$ [26]; the sign of β determines whether the stress is contractile ($\beta > 0$) or extensile ($\beta < 0$). Finally, adhesion-related stresses can be described by $\boldsymbol{\sigma}_a = -\kappa \left[(\nabla\phi) \circ (\nabla\phi) - \frac{1}{2}(\nabla\phi)^2 \mathbf{I}\right]$, where gradients in adhesion molecule concentration (ϕ) generate anisotropic stresses analogous to surface tension effects in fluid interfaces [27].

The terms involved in the active stress σ_a such as myosin concentration c_m or adhesion molecule concentration ϕ can be themselves integrated in the naive reaction-diffusion model of equation (1).

Using continuum models has been a fruitful approach to understanding tissue patterning and morphogenesis on a phenomenological level. Nonetheless, to accurately predict how effective parameters like tissue viscosity η or the effective diffusion coefficient of cells D depend on single-cell properties, such as cell motility, it is necessary to quantify individual stochastic cell behavior within tissues [28, 29].

0.3 Gastruloids: a model system to study self-organization in vitro

While in vivo studies of embryonic development in model systems like the Drosophila fly or the Zebrafish embryos offer crucial insights into the relationship between single-cell behavior and collective tissue behavior, extending these studies to mammalian development is inherently complex. This complexity arises from the implantation of the embryo and extra-embryonic tissues into maternal tissues, a process that presents both ethical and technical challenges [30].

In contrast, in vitro models offer controlled settings where specific variables (such as tissue boundary conditions or external chemical factors) can be manipulated, allowing researchers to dissect the interplay between genetic regulation and biophysical forces in a more systematic manner. However, traditional two-dimensional (2D) cell culture models fail to fully capture the spatial and mechanical constraints that shape cell behavior in three-dimensional (3D) environments [31, 32].

Stem cells aggregates, called embryonic organoids, have emerged as powerful mod-

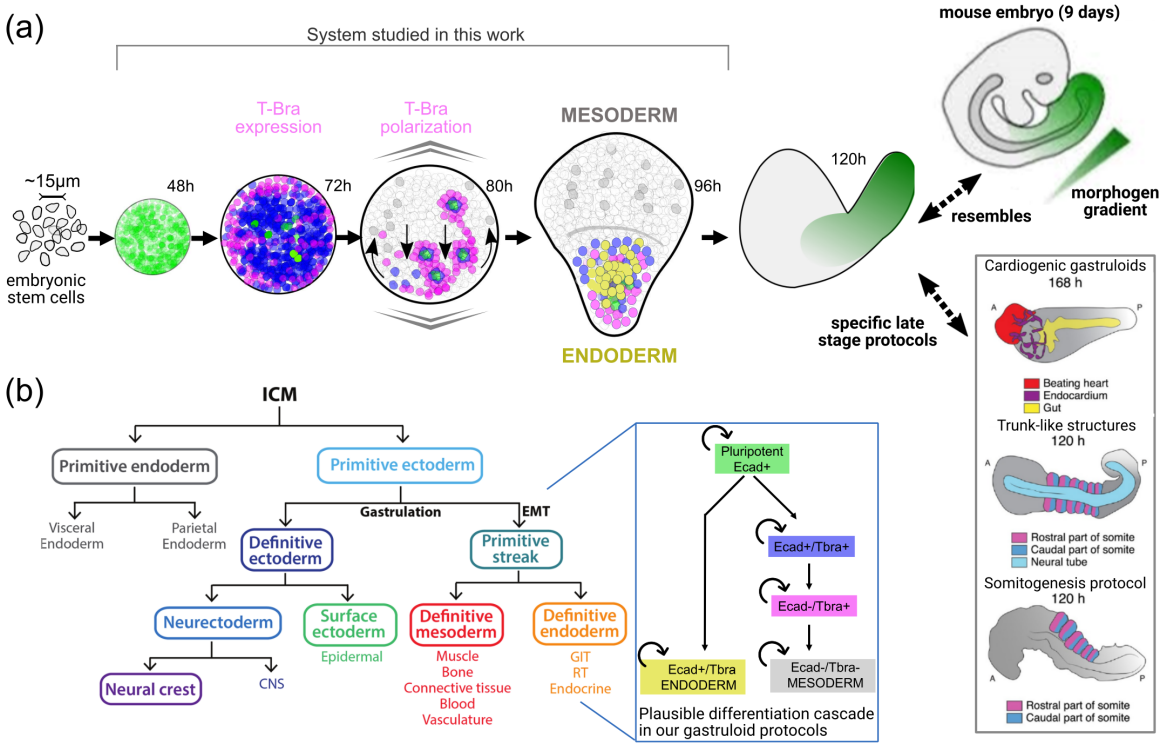

Figure 1: **Gastruloids are in vitro models composed of embryonic stem cells that self-organize while undergoing differentiation. (a)** Similarities with in vivo examples of gastrulation make them interesting to study development. By varying the culture protocol, early stage gastruloids can be led towards specific advanced biological tissues and structures. Some elements were modified from [36, 38]. **(b)** Differentiation of fully pluripotent stem cells (ICM) into the primary germ layers during mammalian embryo morphogenesis. The protocols we study show similarities to the creation of mesoderm and endoderm through an epithelial-to-mesenchymal (EMT) transition.

els for studying tissue development, disease, and therapeutic responses due to their ability to recapitulate key aspects of organogenesis in a controlled, three-dimensional environment [33]. Among these, gastruloids are self-organizing embryonic organoids derived from embryonic stem cells (notably, they can be derived from mouse or human embryonic stem cells) [34]. They offer a unique system to investigate early developmental processes. They display striking similarities to their in vivo counterparts, including the emergence of structures very similar both genetically and morphologically to the neural tube and somites [35] (see Fig. 1(a)). Recent advancements in gastruloid culture protocols have extended their differentiation potential, enabling the formation of tissues characteristic of both anterior mesoderm (e.g., cardiac and vascular precursors) and endoderm (e.g., gut-like structures) [27, 36, 37] (see Fig. 1(a)). These features make gastruloids an attractive model for studying self-organization, morphogenesis, and the interplay between genetic and mechanical cues during embryonic development.

Researchers in my experimental host team, the team Physics of cell dynamics and tissue morphogenesis at the Marseille Developmental Biology Institute, investigate gastruloids morphogenesis.

Gastruloids studied in my host team are generated by aggregating mouse embryonic stem cells (mESCs) cultured in the lab. The process begins with the detachment of mESCs from their culture substrate, allowing them to form aggregates. These aggregates are then cultured under conditions that promote their self-organization in 3D. Over time, the cells within these aggregates undergo differentiation, patterning, and morphogenesis, giving rise to tissues that resemble those found in early embryos, including mesoderm, endoderm, and ectoderm Fig. 1. The first major morphogenetic events in gastruloids occur between 48 and 96 hours post-aggregation [36, 39], and is illustrated on Fig. 1(a). Initially, these structures consist of relatively homogeneous pluripotent stem cells. By 72 hours, distinct cellular subpopulations emerge, including endoderm cells, nascent mesoderm (corresponding to the expression of the gene Brachyruy (T-Bra)), and cardiac and vascular mesoderm, each expressing specific molecular markers such as E-cadherin and N-cadherin. Between 72 and 96 hours, these populations undergo spatial reorganization, accompanied by an epithelial-to-mesenchymal transition (EMT) that generates surface tension gradients and drives a stereotypical tissue flow [27, 36] (see Fig. 1(b)). This transition is critical for gastruloid elongation along the anteroposterior axis, mirroring early embryonic patterning events [40, 41]. The EMT occuring in gastruloids bears analogies with in-vitro systems recently developed associating cancer spheroids with cancer-associated fibroblasts, which exhibit cooperative migration through E-Cadherin and N-Cadherin heterotypic interactions [42]. More generally, recent studies reference gastruloids as a relevant model system for examining the onset of pediatric cancers [43–45].

A fundamental question is how cells spontaneously and robustly self-organize during gastruloids self-organization and morphogenesis. In natural embryos, morphogenesis is guided by a complex interplay between genetic patterning and biomechanical forces, which collectively shape tissue architecture. The same principles are at play in gastruloids (Fig. 1(a)), yet their higher sample-to-sample variability compared to in vivo embryos makes them particularly valuable for studying the stochasticity and robustness of developmental processes. Unlike embryos, which can be aligned to create averaged morphological and genetic atlases [46–48], gastruloids exhibit a high degree of variability that challenges classical registration approaches. This variability provides a broader window into the dynamics of self-organization across different developmental trajectories.

0.4 Key research questions require cell scale comprehension of mechanogenetic interplay

0.4.1 Coarse-grained biophysical models of gastruloids morphogenesis exist but do not reach cell-scale resolution

Until now, efforts to quantify gastruloids morphogenesis have largely remained at the tissue scale due to technical limitations in live imaging and data processing. While advanced microscopy techniques provide valuable insights into large-scale self-organization, they often fall short of resolving individual cellular behaviors, hindering a mechanistic understanding

of how collective tissue dynamics emerge from single-cell decisions.

For instance, Moos et al. [49] demonstrated how multi-view light-sheet microscopy enables quantitative single-cell analysis in large specimens but achieves penetration depths of only 100–150 μm for dense gastruloids, restricting its applicability in deep tissue contexts. Similarly Merle et al. [50] used confocal microscopy to investigate self-organization in mammalian pseudo-embryos, offered high spatial resolution but failed to achieve the precise cell-scale tracking accomplished in Drosophila due to the increased complexity and depth of mammalian systems. Oriola et al. [51] applied confocal imaging to analyze aggregate coalescence in two-dimensional systems, focusing on tissue-scale gene expression rather than resolving cell-specific contributions. While these studies provide essential insights into multicellular organization, they largely infer cellular behavior from large scale tissue properties without considering single-cell aspects.

In our experimental host team, recent imaging advancements have attempted to bridge the gap in understanding gastruloid development, albeit with some limitations. Sham Tlili has successfully employed two-photon imaging to create the first 3D live movies of gastruloids undergoing symmetry breaking at cellular resolution [27, 36]. By utilizing coarse-grained analysis of tissue flows (using optical flow [52] to measure coarse-grained velocity maps) and of a nascent mesoderm gene (T-Bra) expression pattern, these works have proposed a model for gastruloid early polarization and elongation.

In brief, this model, illustrated on Fig. 2,

1. describes how a tissue-scale spatial gradient of cell differentiation from the nascent mesoderm state to the mesoderm state emerges in the gastruloid through a reaction-diffusion mechanism (which is not yet fully understood or characterized)

2. describes how this differentiation gradient generates cell-cell adhesion gradients and tissue surface tension gradients within the tissue

3. proposes a model where active stress σ_a generated by these surface tension gradients in the tissue is coupled to the tissue's effective viscous response to this stress.

This model describes two distinct mechanisms for cell transport: first, tissue-advective flows, which are predictable using the model described above, and second, individual-cell diffusion within advected tissues, which may contribute to the reaction-diffusion mechanism driving differentiation gradient formation.

This model is able to predict the tissue collective flows quantified during early gastruloids symmetry breaking and elongation. However, the coarse-grained velocity maps obtained with optical flow are not precise enough to resolve how individual cells move and interact with their environment. Thus, the nature of the reaction-diffusion mechanism underlying the emergence of cell differentiation gradients in gastruloids remains to be determined 3. To further explore this mechanism, we need to characterize individual cellular movements through single-cell tracking.

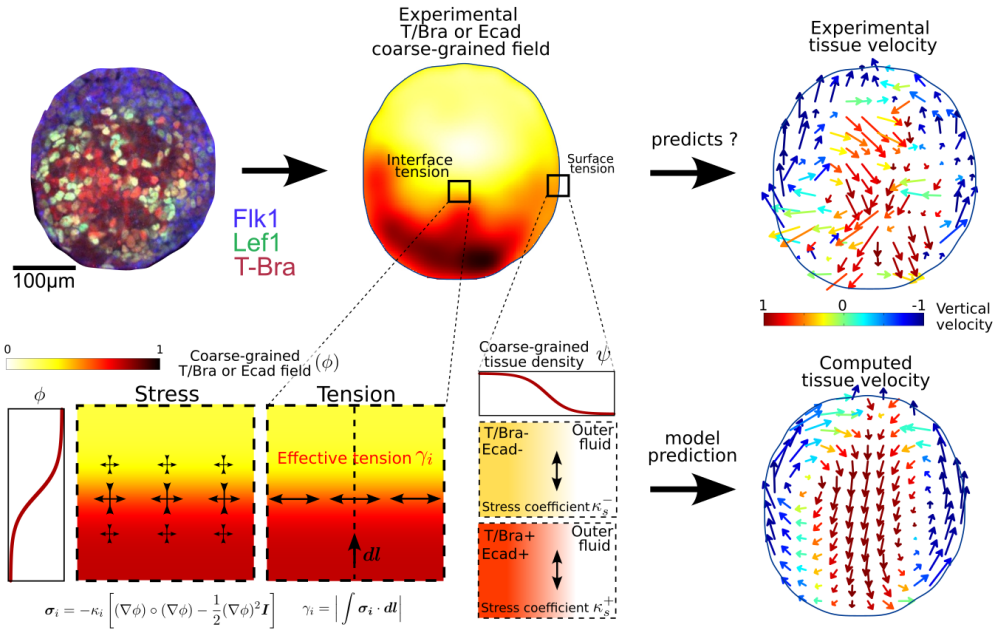

Figure 2: **Early polarization and elongation of gastruloids can be captured by a physical model where spatial gradients in tissue properties drive morphological changes.** The model proposed by [27] predicts collective tissue flows from spatial gradients in cell differentiation, generated through a reaction-diffusion process. Differentiation gradients create spatially varying mechanical properties (similar to Marangoni flows), where active stresses lead to predictable tissue-scale advection. Individual cell diffusion within advected tissues may also influence differentiation patterns. The resulting tissue-scale gradients induce active flows analogous to Marangoni flows, driven by local variations of cellular properties. This approach enables coarse-grained prediction of tissue morphodynamics directly from experimentally measurable fields. Some elements were modified from [27].

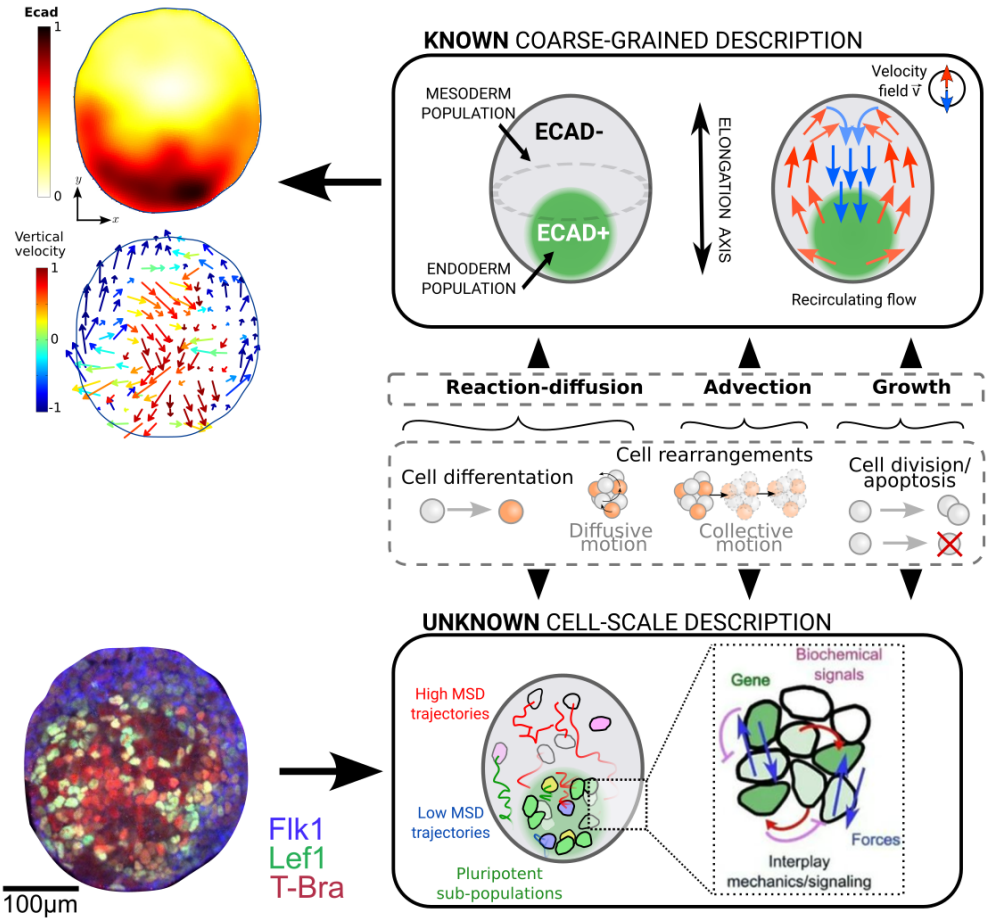

Figure 3: **Biophysical models of morphogenesis in gastruloids do not reach cell-scale resolution.** Recent studies [27] have shed light on how coarse-grained T-Bra patterning gives rise to collective flow in early gastruloids. A cell scale description of the mechanisms governing the interplay between mechanics and genetics remain unattainable without single cell tracking data. Some elements were modified from [27].

0.4.2 Single cell tracking is necessary to bridge tissue-level organization and single-cell decision-making

While many studies have focused on tissue-scale phenomena, understanding morphogenesis at the cellular level is crucial for deciphering how microenvironmental factors influence stem cell fate. The physical properties of the local niche, including matrix stiffness, cell-cell adhesion, and fluid dynamics, shape stem cell behavior by modulating intracellular signaling pathways. Cells integrate these cues to navigate stochastic fluctuations in gene expression and determine their differentiation trajectories. Advances in live-cell imaging, single-cell transcriptomics, and mechanobiology have begun to shed light on how cells interpret and respond to their surroundings [53]. However, further work is needed to connect these insights across scales. Bridging the gap between tissue-level organization and single-cell decision-making is essential for a comprehensive understanding of how morphogenesis remains robust despite underlying cellular stochasticity.

Recent advances in imaging and computational biology have begun to address this challenge. McDole et al. [47] demonstrated that statistical vector flow techniques can reconstruct lineage trajectories by stitching together incomplete cell tracks from multiple embryos. However, such indirect approaches rely on statistical inference rather than direct observation, introducing uncertainties in lineage reconstruction. A breakthrough would be achieved through robust single-cell tracking, allowing continuous monitoring of cellular decisions across developmental time. This remains technically challenging in gastruloids, where lineage reconstruction is complicated by the high density and dynamic rearrangements of cells. Nevertheless, overcoming these challenges is essential for linking tissue-scale morphogenesis to the underlying cellular mechanisms that drive it.

By integrating high-resolution imaging with computational lineage tracking, future studies could directly map how stem cells navigate their microenvironment, balancing stochastic fluctuations with deterministic regulatory cues [25, 54]. Approaches such as CausalXtract [55], previously used to identify causal interactions in complex tumor-on-chip models through time-lapse imaging, could reveal direct causal relationships between cellular behaviors and mechanical signals within gastruloids. Additionally, methods for direct dynamical model inference from noisy data [56, 57], which successfully inferred dynamical laws governing cell migration and collective behaviors from noisy experimental trajectories, could reconstruct underlying physical mechanisms from lineage-tracing data. Such approaches will be pivotal in bridging the gap between tissue-level self-organization and single-cell fate determination, ultimately advancing our understanding of how morphogenesis remains robust despite cellular heterogeneity.

0.5 Two-photon microscopy allows deep-imaging of gastruloids

0.5.1 Basics of fluorescence microscopy

Fluorescence microscopy is a fundamental technique for deep tissue imaging, offering significant advantages over brightfield microscopy, particularly in thick and scattering biological specimens. Unlike brightfield imaging, which relies on light absorption and scattering, fluorescence microscopy employs fluorophores that emit light upon excitation, enabling molecular specificity and enhanced contrast. This technique allows visualization of specific structures with high spatial resolution, making it indispensable for biological imaging.

The key mechanism of fluorescence microscopy involves illuminating the sample with excitation light, which is absorbed by fluorophores, leading to emission at a longer wavelength. By filtering the emitted fluorescence, high-contrast images of labeled structures are obtained. This approach minimizes background noise and enhances image clarity, particularly in complex tissues.

Despite its advantages, fluorescence microscopy has limitations, including photobleaching, where fluorophores degrade under prolonged exposure, and phototoxicity, which can damage live tissues. Additionally, sample preparation often requires complex staining or genetic modifications, which may introduce variability. Advanced fluorescence microscopy techniques, such as confocal, light sheet, and two-photon microscopy, address some of these limitations by improving optical sectioning, penetration depth, and reducing photodamage.

0.5.2 Comparison of advanced fluorescence microscopy techniques

Fluorescence microscopy is a powerful tool for imaging biological tissues, yet it faces significant limitations when applied to deep tissue imaging.

- The combination of light scattering, absorption, and optical aberrations, which occurs as photons interact with cellular structures and extracellular components, leading to signal degradation and reduced resolution at increasing depths. This effect limits the ability to obtain sharp images beyond a few hundred micrometers in thick specimens.
- Phototoxicity, where prolonged exposure to excitation light induces cellular stress and damage. High-intensity illumination, particularly in live specimens, disrupts physiological processes, potentially altering cell behavior and viability
- Photobleaching, which is the the irreversible degradation of fluorophores due to repeated excitation, reduces signal intensity over time, limiting the duration of long-term imaging experiments and the ability to track dynamic processes

Confocal, light sheet, and two-photon microscopy each offer unique advantages for fluorescence imaging of biological specimens. Fig. 4 illustrates each microscopy modality.

Confocal microscopy utilizes point-scanning illumination and pinhole detection to achieve optical sectioning, providing high-resolution imaging but with limited depth penetration and increased photobleaching. Light sheet microscopy, in contrast, illuminates the sample with a thin sheet of light, enabling rapid volumetric imaging with minimal photodamage, making it particularly suited for developmental biology. Two-photon microscopy, which relies on nonlinear excitation, allows deep tissue imaging by minimizing scattering and out-of-focus photobleaching, making it ideal for thick specimens.

A comparative summary is presented in Table 1, proposing pros and cons of each method, and giving sensible order of magnitudes from recent studies on gastruloids or deep organoids (Moos et al. [49], Merle et al. [50], Oriola et al. [51], and Gsell et al. [27]).

Microscopy	Confocal	Light Sheet	Two-Photon
Illumination	Point-scanning (parallel)	Planar	Point-scanning
Speed	Fast	Very fast	Moderate
Photodamage	Moderate-High	Low	Low
Resolution (lateral)	200nm	200nm	500nm
Resolution (axial)	500-700nm	1-2µm	800nm-1µm
Sample prep	Any mounting	Special mounting	Any mounting
Photobleaching	High	Low	Low
Scattering	High	Medium	Low

Table 1: **Comparison of key features between confocal microscopy, light sheet microscopy, and two-photon microscopy for deep organoid and gastruloid imaging.** Values are meant to provide an order of magnitude, informed by [27, 49–51]. They are approximate and depend on specific implementations.

0.5.3 Two-photon microscopy dominates for deep imaging

Recent advances in light-sheet and confocal imaging enable high-resolution, single-cell digital atlases of developing organoids, with techniques like deep-learning-based segmentation and tracking [49, 60–62]. Light-sheet microscopy (LSM) has been widely used for live imaging in model organisms such as Drosophila, zebrafish, and mouse embryos, particularly in seminal atlas studies [46–48]. This technique provides high-speed imaging with pixel sizes on the order of 300 nm and frame rates exceeding 1 volume per second. However, while LSM is well-suited for rapid imaging of relatively transparent specimens, it requires excellent tissue clearing to maintain a good signal-to-noise ratio (SNR) at depth due to light scattering.

The applicability of LSM to mouse embryos arises from their structural organization: despite their three-dimensional nature, early-stage mouse embryos exhibit a more two-dimensional arrangement of tissues, particularly in structures such as the notochord, where cells remain relatively close to the embryo's boundary [47]. In contrast, gastruloids, which are mouse embryonic stem cell-derived organoids mimicking key genetic and structural aspects of early embryonic development [27, 34, 36], present a different challenge. These dense, highly scattering tissues significantly limit the penetration depth of LSM, making it

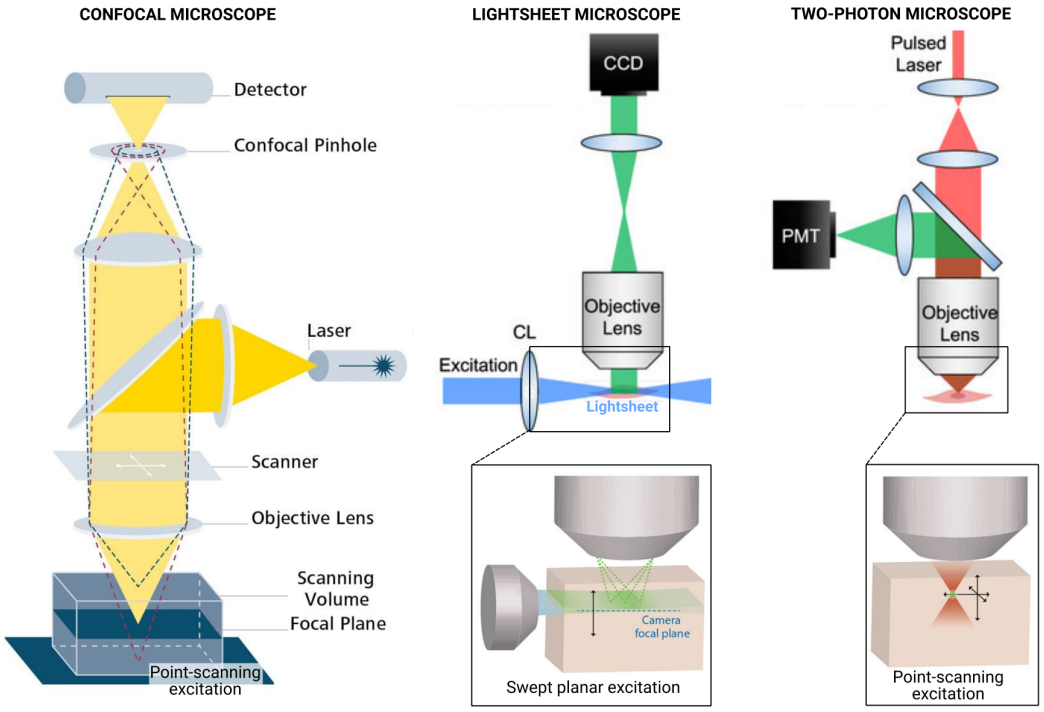

Figure 4: **Illustration of the differences between confocal, lightsheet and two-photon microscopy in terms of optical setup and 3D excitation profile.** Modified from [58, 59].

difficult to image beyond 200 µm, often insufficient for whole-mount imaging.

For imaging such dense tissues, two-photon microscopy (2PM) offers significant advantages. By utilizing longer excitation wavelengths, 2PM achieves deeper tissue penetration with minimal photodamage and reduced scattering [63, 64]. Unlike LSM, which suffers from intensity gradients and blurring in optically complex samples [65], 2PM mitigates these effects due to the use of longer excitation wavelengths that reduce scattering and absorption, and its intrinsic confinement of excitation to a small focal volume, enabling efficient collection of scattered emission photons through non-descanned detectors. However, 2PM is inherently slower due to its point-scanning acquisition scheme and typically exhibits lower lateral resolution than LSM. 2PM allows using a single laser to excite two channels (two fluorophores with distinct emission spectra) due to the broad excitation spectra. This is particularly interesting for live imaging, during which sample illumination has to be minimized as much as possible to prevent phototoxicity and bleaching. However, this advantage is accompanied by increased crosstalk between the channels.

Adaptive optics (AO) significantly enhance two-photon microscopy (2PM) by addressing limitations in imaging depth and frame rate. Incorporating AO with three-photon microscopy has notably improved imaging depth within scattering tissues like the brain, enabling visualization of neuronal structures deeper than conventional methods allow [66]. Additionally, techniques such as two-photon light-sheet microscopy [67] and multi-beam two-photon microscopy [68] can be used to allow higher imaging rates crucial for capturing dynamic biological processes. However, AO techniques remain technically complex, requiring specialized equipment, skilled maintenance, and substantial resources, limiting

their use in our experiments.

0.5.4 Imaging setups for fixed and live gastruloid imaging

Gastruloids are cultured from mouse embryonic stem cells. These cells are first placed in non-adherent plates, allowing them to form 3D clusters. After 48 hours, the addition of a signaling molecule called Chiron prompts differentiation, and by 72 hours, the gastruloids start to develop a specific pattern of gene expression, resembling early tissue organization. The culture conditions and protocol is presented in [36].

For imaging, the gastruloids are mounted between two glass slides with spacers, using various solutions to ensure proper transparency for imaging. Fixed samples were treated with solutions like glycerol or antifade mountants to preserve the tissue for long-term imaging. Live samples, on the other hand, were imaged in a special chamber kept at 37°C with a 5% CO2 atmosphere to maintain the cells in a physiological state.

A laser is used to excite fluorescent markers that highlighted specific proteins in the cells. Different fluorophores were chosen to label distinct proteins or cell structures, such as cell nuclei and the intercellular space. For live imaging, a laser at 920 nm was used to excite specific markers (notably, GFP derivatives exist in this wavelength range), and the fluorescence emitted by the samples was detected through special filters to produce detailed images.

Some example datasets are illustrated on Fig. 5.

0.6 Image analysis challenges in gastruloid imaging

Challenges stem from experimental limitations combined with aleatoric and epistemologic stochasticity The analysis of 3D and 3D+time datasets of gastruloids acquired through two-photon microscopy is fundamentally challenging due to the interplay between experimental limitations and different sources of stochasticity. Aleatoric uncertainty (inherent stochasticity of the system), which arises from the inherent stochasticity of gene expression, cell motion, and shape dynamics, is further aggravated by physical and optical constraints such as scattering, signal degradation, and limited acquisition speed. These experimental limitations not only amplify biological variability but also introduce epistemic uncertainty (uncertainty introduced by the observation process) by affecting the accuracy of measurements and computational analyses. Segmentation errors, inaccuracies in lineage reconstruction, and the difficulty of extracting precise gene expression readouts from fluorescence intensities all stem from these limitations, reflecting uncertainty due to incomplete or imperfect observations. As a result, both the precision of gene expression quantification and the robustness of nuclei segmentation and tracking are compromised. Furthermore, the difficulty of visually exploring and annotating datasets adds another layer of epistemic uncertainty, complicating the generation of reliable ground-truths for segmentation and tracking validation.

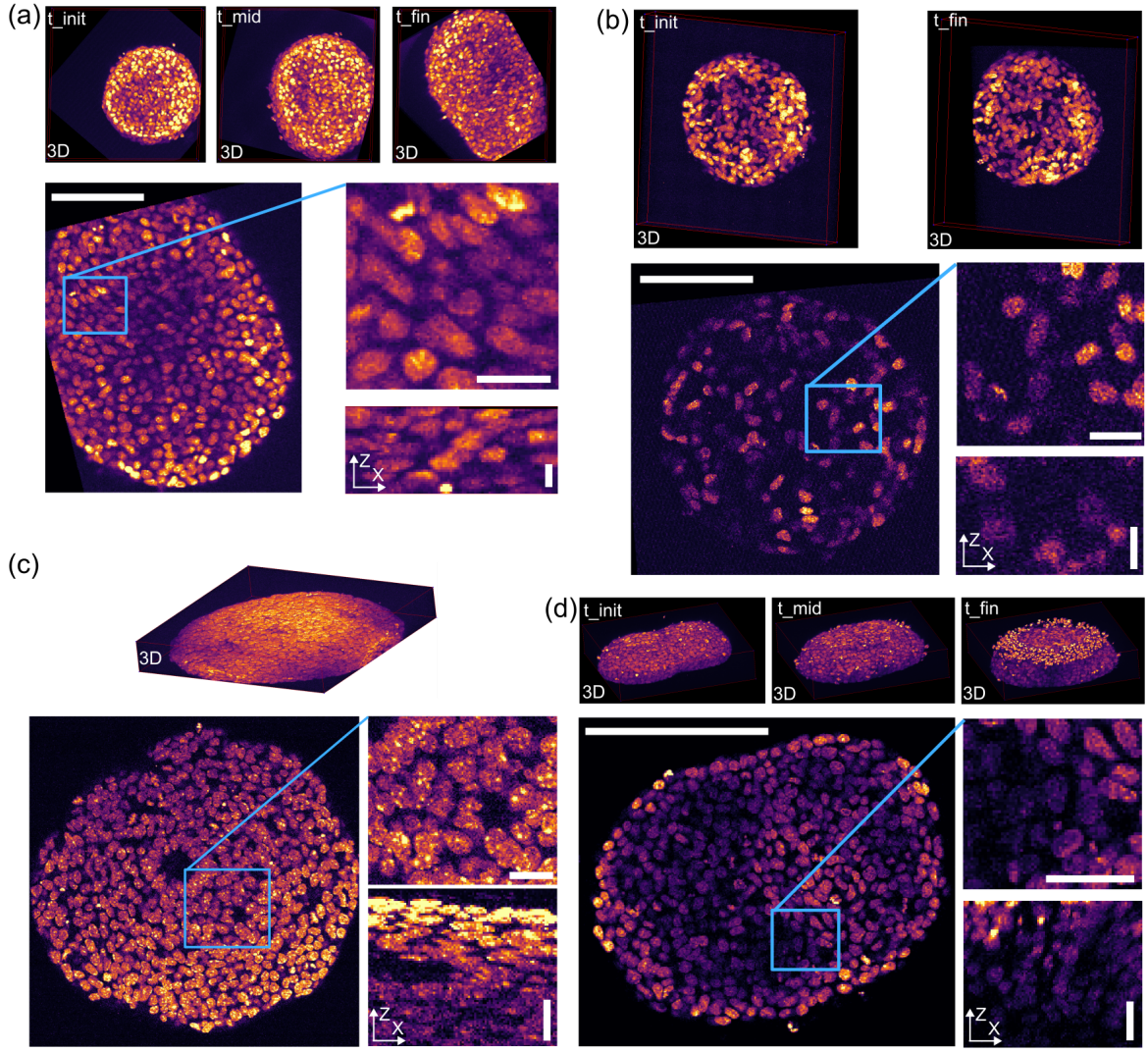

Figure 5: **Overview of raw fixed and live typical datasets. (a)** Live, recirculation experiment $d_z = 2\,\mu m$, $d_{xy} = 0,6\,\mu m$, time interval $280\,s$. **(b)** Live, sparse ($\sim$ 30% fluorescent cells) $d_z = 1\,\mu m$, $d_{xy} = 0,6\,\mu m$, time interval $300\,s$. **(c)** Fixed, $d_z = 2\,\mu m$, $d_{xy} = 0,6\,\mu m$. **(d)** Live, fusion experiment, $d_z = 1\,\mu m$, $d_{xy} = 0,6\,\mu m$, time interval $300\,s$. Scale bars: $100\,\mu m$ (wide views), $20\,\mu m$ (close-ups).

Optical artifacts and measurement noise in gene expression readout Extracting reliable gene expression readouts from two-photon microscopy datasets is hindered by a variety of factors. Measurement noise, photon shot noise, and scattering introduce spatially dependent intensity variations, which degrade the signal-to-noise ratio (SNR) and limit the accuracy of fluorescence intensity quantification (see Fig. 6(a)). Additionally, volumetric intensity artifact which are caused by excitation power attenuation and depth-dependent signal degradation create non-uniform expression readouts across the imaging volume (see decay in XZ slices on Fig. 5(c,d)). For fixed tissue datasets, immunostaining penetration is particularly problematic in densely packed tissues, as antibodies may not uniformly reach all regions, further reducing the reliability of fluorescence intensity as a direct proxy for gene expression levels (see Fig. 5(c)).

Nuclei segmentation challenges in a dense and dynamic context Automated segmentation of nuclei is complicated by both biological and imaging factors. Nuclei in gastruloids are highly deformable, with a large nucleus-to-cell volume ratio that makes their contours difficult to delineate (see Fig. 6(a-b)). Furthermore, nuclear fluorescence intensity and shape change dynamically throughout the cell cycle, particularly during mitosis, when nuclei undergo rapid morphological transformations (see Fig. 6(c-d)). These factors contribute to shape heterogeneity and the presence of poorly resolved regions, making segmentation difficult even for expert human annotators.
The limitations imposed by phototoxicity and photobleaching further exacerbate segmentation difficulties. The need to limit laser power to preserve sample viability constrains image acquisition, leading to low SNR in small, dimly fluorescent nuclei. The combination of low SNR, heterogeneous nuclear morphology, and tightly packed cells creates a particularly challenging environment for segmentation algorithms, which struggle to delineate individual nuclei with high accuracy.

Complexity of cell tracking and lineage reconstruction The tracking of individual nuclei over time, required for lineage reconstruction, faces additional challenges due to the dynamic nature of the system. Gastruloids contain multiple cell populations with distinct mechanical properties, leading to heterogeneous movement patterns (see Fig. 6(c)). Additionally, rapid cell divisions frequently disrupt nuclear continuity, as daughter nuclei often migrate away from each other shortly after mitosis, increasing the difficulty of linking tracks across time points (see Fig. 6(d)). The short contact times between neighboring cells, frequent neighbor exchanges, and tissue mixing further increase tracking complexity, requiring sophisticated tracking algorithms capable of handling high levels of cell rearrangement.
Limitations on temporal resolution introduce additional difficulties. The photon budget constraints imposed by phototoxicity and bleaching limit the achievable framerate, leading to large displacements of cells between consecutive frames. The resulting high mean free path of nuclei between time points makes tracking particularly difficult, as fast-moving cells may appear to "jump" unpredictably in image sequences (see Fig. 6(c)). This type of rapid movement is also highly involved in tumorigenesis as EMT allows cancer cells to invade surrounding tissues and enter the bloodstream or lymphatic system [42].

Visualization and annotation challenges Beyond automated analysis, the visualization and manual annotation of these datasets pose additional difficulties. The high nuclear packing density, small cell sizes, and low SNR make it difficult to explore datasets visually and track individual nuclei over time using standard 2D projections. Given the inherent complexity of gastruloid motion, manually following a single nucleus through time is challenging, as rapid cell rearrangements make it difficult to maintain continuity between frames.

These issues significantly impact the ability to generate reliable ground-truth datasets for segmentation and tracking. Manually annotating nuclei in dense, low-SNR images is both time-consuming and error-prone, limiting the scalability of ground-truth generation for supervised learning approaches (see Fig. 6(d)). For instance, using a typical annotation tool like Napari [69], we record an average time of $\sim$15 seconds per annotated nucleus in 3D, factoring in a diverse pool of nuclei morphologies. Bright, well-delineated nuclei can be annotated much quicker, but nuclei in low SNR regions or in a high-packing neighborhood require careful inspection in perpendicular views or in time (if available) to prevent mistakes. Developing robust computational frameworks that integrate uncertainty quantification, active learning, or self-supervised strategies will be critical for overcoming these challenges and improving automated segmentation and tracking in gastruloid datasets.

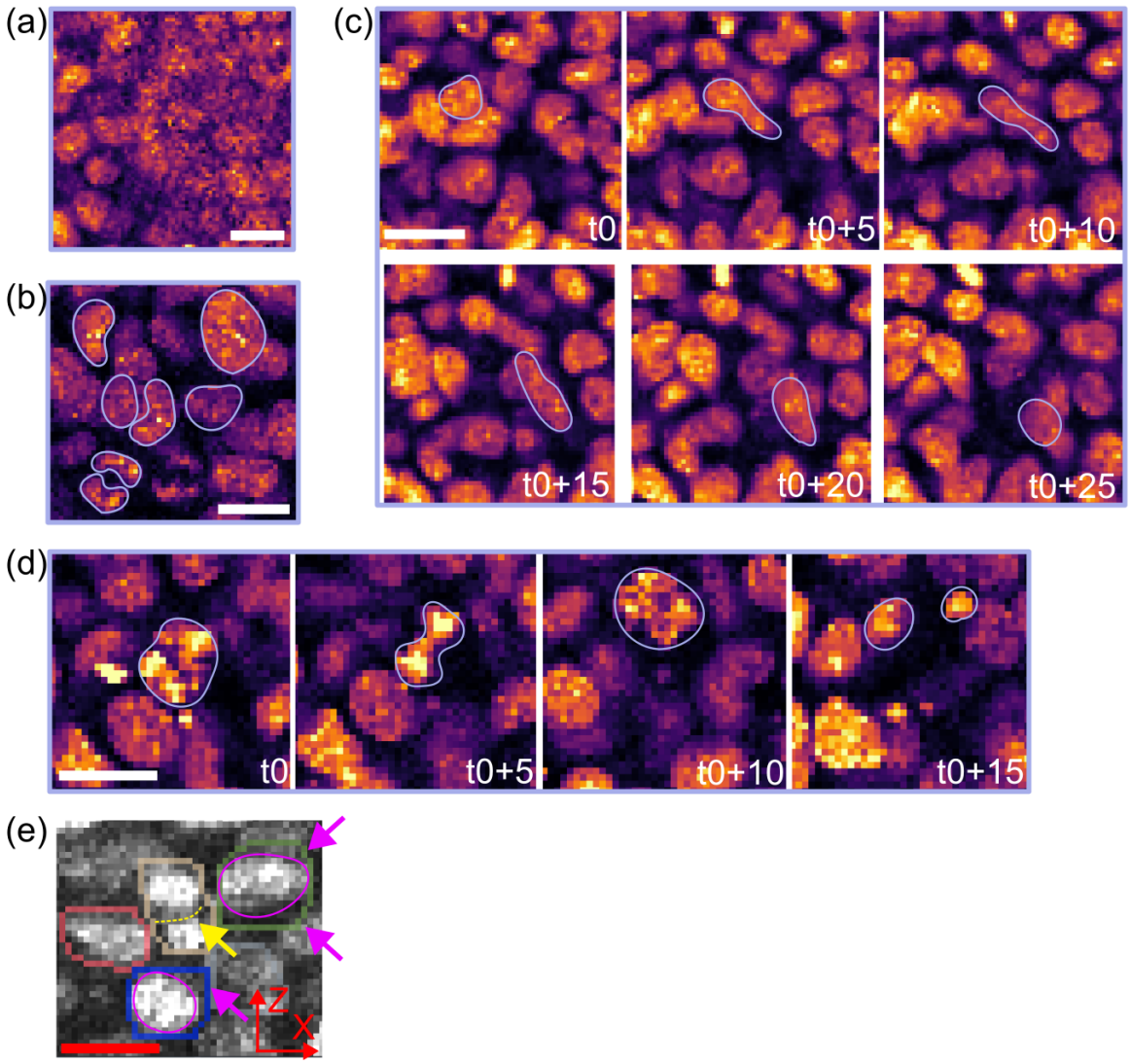

Figure 6: **Challenges of image analysis.** **(a)** Low SNR regions cannot be resolved. **(b)** Shape heterogeneity complicates segmentation. **(c)** Nuclear plasticity leads to squeezing events which are characterized by elongated nuclei and rapid movement. **(d)** Cell divisions generate rapid movement and shape discontinuity. **(e)** Manual annotation (here a segmentation "ground-truth") is prone to error. In high object packing regions, annotators tend to merge nuclei in Z because they cannot see the separation when annotating in XY (yellow arrow). Annotators also tend to overshoot shape along Z, so that segmented nuclei appear cylindrical (pink arrows). Time in minutes. Scale bars $10\,\mu m$

Objectives of this thesis

The global objective of this thesis is to advance the quantification of cell-scale lineage dynamics in gastruloids, with the ultimate goal of understanding the microenvironmental perspective of individual cells. To achieve this, the following objectives will be considered:

- **Establishing technical milestones** necessary for improving the quantification of single-cell lineages, including enhancements in imaging protocols, along segmentation and tracking methodologies.
- **Developing computational tools** that address the challenges inherent to the datasets, particularly in handling stochasticity and enabling efficient data exploration.
- **Investigating mechanical and dynamic properties** within the datasets to provide new insights into the interplay between tissue mechanics and cellular behavior.
- **Designing accessible and interactive tools** tailored to the long-term needs of the research team, ensuring usability for non-expert users and facilitating widespread adoption.

Thesis overview

This thesis is structured into two main parts.

In the first part, we focus on analyzing mechanical and genetic properties in fixed datasets. Chapter 1 introduces an adaptation of a state-of-the-art deep learning-based instance segmentation algorithm, specifically tailored for detecting nuclei in densely packed gastruloid tissues, despite challenges posed by optical artifacts. In Chapter 2, we integrate this custom segmentation method into a comprehensive processing pipeline capable of analyzing mechanical and gene expression patterns across scales, from individual nuclei to entire tissues. Additionally, we demonstrate how this pipeline has been incorporated into a user-friendly Python package, featuring graphical plugins compatible with the visualization software Napari.

The second part of the thesis addresses critical technical milestones and introduces methodological advances aimed at enabling large-scale cell lineage reconstruction in gastruloids. In Chapter 3, we provide a review of existing cell tracking and lineage reconstruction methods, contextualizing their relevance for deep gastruloid tissues. We identify significant challenges associated with imperfect segmentation, heterogeneous and rapid cell movements, and the automatic or visualization-based calibration of existing methods.

We further highlight that Bayesian tracking techniques are particularly suitable for managing stochastic systems with complex dynamics, due to their proficiency in integrating physical priors and quantifying uncertainties.

Building upon these insights, Chapter 4 presents a basic Bayesian tracking algorithm which can handle rapid and heterogeneous cell movements. We also extend recent developments,which introduced a Bayesian metric termed the trackability score to quantify assignment uncertainties. Chapter 5 explores the application of this trackability score for the unsupervised ranking of segmentation algorithms based primarily on temporal coherence. A custom particle simulation framework, alongside benchmark and gastruloid datasets, validates the efficacy of this approach.

Finally, in Chapter 6, we illustrate how incorporating simple priors related to cell lifetimes can enhance the trackability score, enabling automated and unsupervised calibration of Bayesian tracking algorithms. We again validate our method through simulations and demonstrate promising results with real gastruloid datasets.

Part I

Fixed datasets

Chapter 1

Detecting heterogeneous nuclei shape in packed 3D gastruloids requires tailored neural-network training and validation strategy

1.1 Chapter overview

This chapter addresses the specific challenges associated with detecting nuclei shapes in densely packed, three-dimensional gastruloid tissues. We present an adaptation of a deep-learning segmentation algorithm (StarDist3D), tailored to overcome the limitations posed by optical artifacts and high object density. By developing a dedicated preprocessing pipeline involving dual-view fusion and local contrast enhancement, we significantly enhance segmentation accuracy, setting a solid foundation for subsequent analysis pipelines discussed in later chapters.

This chapter is part of the publication [70]:
GROS, Alice*, VANARET, Jules*, DUNSING-EICHENAUER, Valentin*, et al. A quantitative pipeline for whole-mount deep imaging and multiscale analysis of gastruloids. bioRxiv, 2024, p. 2024.08. 13.607832. (*= co-first authorship)

My contributions:

1. jointly designed a preprocessing pipeline and a training protocol to optimize a state-of-the-art deep-learning model for 3D segmentation in challenging deep tissue images using a limited amount of manually annotated gastruloid data
2. created a method to validate 3D segmentation performance based on comparison between 3D predictions and 2D annotations, which greatly improves the speed at which a validation dataset can be created
3. showed that the trained StarDist3D model achieved 85% F1 score at 0.5 IoU, which, together with the preprocessing protocol established, allows segmenting nuclei reliably in full gastruloids despite high nuclear packing and low SNR

1.2 Motivations and Challenges

Fluorescence imaging allows direct visualization of individual cell nuclei. Nuclear shape is a well-established proxy for cell shape, as the nucleus is mechanically coupled to the cytoskeleton and extracellular matrix [71], allowing it to reflect cellular responses to mechanical stresses and strains. Studies have demonstrated that nuclear morphology changes in response to forces such as actomyosin tension, cell spreading, and external compressive stresses, highlighting its role as a mechanosensor [72]. Automated nuclear segmentation is essential for quantifying nuclear shape changes, enabling the inference of tissue-scale deformation patterns and the mapping of stress and strain distributions critical for understanding tissue mechanics and morphogenesis. Moreover, segmentation is a critical step of several multiple cell tracking algorithms [73–75].

Traditional computer-vision methods, such as thresholding-based techniques and watershed algorithms, have been used widely due to their simplicity and efficiency in segmenting clear, isolated nuclei and clustered nuclei with roundish shapes [76, 77]. However, these approaches struggle in dense organoid tissues where nuclei are closely packed, display irregular shapes, and contrast varies heavily within the volume due to light scattering, absorption and optical aberration in tissues. This leads to segmentation errors in the form of false positives/negatives objects, merges (under-segmentation), and splits (over-segmentation).

In recent years, advancements in 3D nuclei segmentation within dense biological tissues have enabled researchers to analyze cellular structures with increasing precision. Developments in deep-learning have opened new avenues for segmentation, including unsupervised and semi-supervised approaches that minimize the need for large annotated datasets. Self-supervised methods like CellSeg3D [78] or Ref. [79] have shown promising results, training from raw images by leveraging contrastive losses or surrogates training tasks, but remain ineffective in deep tissue imaging, where low SNR obscures nuclear boundaries. Other self-supervised approaches rely on separating objects from background using temporal coherence [80], but require that movement across the image is first approximated using optical flow, and thus only applies to smooth flows. Methods such as Segment Anything Model (SAM) and more recently [81], MicroSAM [82] propose universal segmentation models trained on massive open datasets. They offer interactive and automated ways to guide segmentation using point-based annotations or bounding boxes. These techniques allow users to specify object locations with minimal effort, but they are primarily designed for larger, well-defined objects and often fail when applied to small, densely packed nuclei with blurry edges.

Machine learning-based methods, including U-Net and other convolutional neural network (CNN) architectures, have improved segmentation accuracy by learning complex patterns within dense tissue images. These models perform particularly well at identifying nuclei contours even when the contrast is low and boundaries are ill-defined. However, their performance often drops when segmenting closely packed objects, and they require large, manually annotated datasets for training, which limits their usability in dense, challenging to segment organoid tissues. A breakthrough in addressing this challenge is the development of StarDist3D, which integrates deep learning with a star-convex polygon

representation to efficiently encode individual nuclei shapes, even in densely packed samples. Unlike conventional segmentation methods that rely on the spatial arrangement of individual pixels or voxels to reconstruct object boundaries, which often results in ambiguities, StarDist3D encodes complete shape information within a single central voxel. This approach enables more precise and coherent segmentation, particularly in crowded environments with overlapping or irregularly shaped structures. By predicting star-shaped distance distributions from nuclei centers, StarDist3D effectively separates overlapping structures and accurately segments non-convex nuclei.

However, segmentation errors are bound to persist in the deepest planes where the SNR is the lowest and nuclei clusters are impossible to separate even by a human expert. Exhaustive manual annotation of complex cases is virtually impossible which, together with the fact that manual 3D segmentation of packed objects is inherently slow, practically prevents obtaining a large training dataset that is sufficient to "saturate" the network learning capacity. To alleviate these issues, we designed a preprocessing pipeline to maximize contrast across the sample volume and to mix annotated datasets coming from different human annotators and imaging modalities. As annotating validation datasets is obviously also problematic, we proposed a method to obtain a close lower bound on the segmentation performance using only 2D annotations.

1.3 Methods

1.3.1 Preprocessing

1.3.1.1 Spatial registration and fusion using a block-matching algorithm

In experiments with fixed organoids, multiple organoids can be imaged from the same slide, first from one side and then from the opposite side. After pairing each organoid to its counterpart in the reverse view, the two views need to be fused.

Registration by block-matching Popular user-friendly solutions for multi-view registration and fusion include BigStitcher [83, 84], as well as intensity-based frameworks such as Elastix [85, 86] (both available as Fiji [87] plugins). In our case, each dual-view image stack pair was registered using a rigid 3D transformation calculated via a content-based block-matching algorithm [88], previously applied for multi-view and time-lapse registration in light-sheet imaging [47, 89]. We opted for this implementation because it is distributed as a pure-Python package, accepts a user-defined initial transform (which can be crucial when the overlap between views is small and images texture lacks distinctive features), allows the resulting transform to be applied in batch to additional pairs, and natively handles multichannel stacks.

From the two images of each pair, one was selected as the reference view, while the other, termed the floating view, was registered to the reference (Fig.,1.4a). A Gaussian pyramid is built for both views by applying Gaussian smoothing to the images at different kernel

scales: at each scale, cubic blocks from the floating image are matched to their most similar neighbors in the reference view using normalized cross-correlation. A robust estimator removes inconsistent correspondences and fits a single six-parameter rigid transform (three rotations, three translations), refined from coarse to fine scales.

After registration, each dual-view pair was fused into one image stack by giving adaptive mixing weights to pixels belonging to volume common to the two views. The weight relative proportion of a pixel belonging to a view decreased according to a sigmoid profile in deeper planes with respect to the imaging objective to compensate for the decrease in SNR. Because imaging included multiple channels, we performed the registration on the ubiquitous Hoechst-stained nuclei channel and applied the resulting transformation matrix across all channels. Registering a typical 512 × 512 × 300 voxel pair completes in the order of a minute on a standard scientific laptop (8-core CPU, 64 GB RAM).

Initializing the rigid transformation using manually defined parameters Due to the computational cost of running the block-matching procedure in large 3D images, the neighborhood search space was restricted, and thus, an initial transformation was required to approximate the actual alignment. In cases where both views are opposite, the initial transformation corresponds approximately to a 180° rotation around the X-axis with a small translation. However, minor additional rotations due to slide flipping were sometimes observed (around 10° around the X-, Y-, and Z-axes, respectively), which occasionally prevented the registration algorithm from converging. For these cases, we created a custom *Napari* plugin *napari-manual-registration* (Fig.2.11) to apply interactive pre-registration by manually adjusting rotation and translation, and to visualize the alignment in 3D.

Assessing registration quality using FRC-QE The Fourier Ring Correlation Quality Estimate (FRC-QE) is a robust, automated metric for assessing image quality in three-dimensional fluorescence microscopy datasets, particularly for cleared organoids [90]. It is based on Fourier ring correlation, which measures spatial frequency correlations between adjacent slices in an image stack. This approach enables the evaluation of depth-dependent image quality across the z-axis, providing insights into the efficiency of optical clearing protocols and the resolution of microscopy images.

1.3.1.2 Contrast enhancement

We used percentile-based contrast enhancement with intensity clipping. This procedure transforms the intensity $I(\vec{x})$ of a given voxel at position $\vec{x} = (z, y, x)$ according to the following formula

$$I'(\vec{x}) = \frac{I(\vec{x}) - P_{min}(\vec{x})}{P_{max}(\vec{x}) - P_{min}(\vec{x})}, \tag{1.1}$$

where $P_{min}(\vec{x})$ and $P_{max}(\vec{x})$ denote the lower and upper percentile intensity values, respectively. The transformed intensity values $I'(\vec{x})$ are subsequently clipped to the range $[0, 1]$ to ensure they lie within a standardized interval.

Global Contrast Enhancement In the case of global contrast enhancement, we typically choose $P_{min}(\vec{x})$ and $P_{max}(\vec{x})$ as the first and 99th percentiles of the intensity values computed across the entire image. While this approach effectively improves the overall image contrast, it was observed (Fig. 1.5(a) that it fails to adequately restore intensities in the deeper regions of certain samples. This limitation reduces segmentation performance in these regions.

Local Contrast Enhancement To overcome the limitations of global contrast enhancement, we used a localized approach in which $P_{min}(\vec{x})$ and $P_{max}(\vec{x})$ were computed within a sliding window centered at $\vec{x}$. The window size w was selected to approximate the average size of nuclei in the dataset, thereby tailoring the contrast adjustment to the spatial characteristics of the sample. This localized enhancement method effectively addresses regional variations in contrast and yields improved uniformity of intensity across the image. As with the global approach, the resulting locally enhanced intensities were clipped to the range $[0, 1]$.

1.3.2 Segmentation using a deep-learning algorithm

Quantifying cell-scale gene expression and nuclei morphological properties requires accurate instance segmentation of individual nuclei. In early-stage gastruloid datasets acquired with two-photon microscopy, accurate segmentation of nuclei is hindered by high nuclei packing, shape heterogeneity, and spatially varying SNR due to optical aberrations and scattering, especially deeper in the sample. We thus trained a custom StarDist3D [91] model to segment nuclei, as it excels with separating star-convex objects in close contact.

1.3.2.1 The StarDist segmentation algorithm

To segment nuclei, we trained a custom StarDist3D [91] model. StarDist3D is a deep-learning-based algorithm designed for the segmentation of 3D objects, particularly those with star-convex shapes, such as cell nuclei. We illustrate the algorithm on Fig. 1.1. The method leverages a convolutional neural network (typically a U-Net [92] or Resnet [93] architecture) to predict two key outputs for each voxel in a 3D image: an object "centerness" probability and a set of radial distances. The object probability map, on the other hand, indicates the likelihood that a given voxel corresponds to the center of an object. The radial distances represent the lengths from the central point of an object to its boundary along a predefined set of directions in 3D space (e.g., uniformly distributed rays). These distances encode the geometry of the object under the assumption that it is star-convex, meaning all boundary points can be linked to a single central voxel by lines which do not exit the object. This assumption allows representing shapes that can significantly deviate from pure convexity (e.g the nucleus at time $t0 + 10$ in Fig.6(c)).

During training, ground truth annotations are transformed into maps of radial distances and object probability, which serve as supervised targets for the network. The model learns

density	**dimension**	**num. nuclei**	**pixel size** (μm/pix)
dense	3D	1018	Z: 1, Y/X: 0.25
dense	3D+time	100	Z: 1, Y/X: 0.62
sparse	3D+time	3296	Z: 1, Y/X: 0.62

Table 1.1: **Description of the datasets used for training our custom StarDist3D model to detect nuclei.** All datasets were acquired with two-photon microscopy.

to map input 3D images to these outputs by minimizing a loss function that combines errors in predicting radial distances and object probabilities. At inference time, the trained network predicts these outputs for unseen images. A post-processing step reconstructs individual object boundaries by interpreting the predicted map of radial distances and applying non-maximum suppression to resolve overlapping predictions. This ensures that distinct objects are accurately delineated even in densely packed regions. The non-maximum suppression step requires a threshold for the map of object probability, which reduces the amount of pixels that are considered and removes low probability maxima, and a second threshold which sets the IoU above which two segment candidates are considered as conflicting (in which case the segment associated to the lowest object probability is discarded).

The algorithm's reliance on the star-convex shape assumption enables efficient representation and segmentation of objects with compact geometries in packed environments. However, the assumption of star-convex geometry also limits its applicability to structures that deviate significantly from such morphologies, as is the case with highly irregular shapes, e.g with multiple elongated components. Despite this limitation, StarDist3D has proven highly effective for segmenting densely packed, blob-like structures in 3D microscopy datasets, offering a robust solution for many biological imaging tasks where traditional segmentation methods struggle [91].

1.3.2.2 Expanding the training dataset: multi-annotator and synthetic approaches

Mixing several annotated datasets from different modalities To increase robustness, the network was trained on three datasets acquired with two-photon microscopy in different experimental conditions, namely on (1) live-imaging of H2B-GFP gastruloids, (2) live-imaging of sparsely labeled mosaic gastruloids (composed of a mix of 75% of non-fluorescently labeled cells and 25% of cells expressing H2B-GFP under the Brachyury gene promoter [36]), and (3) 96 h fixed samples stained with Hoechst. This combined dataset presented variations in labeled nuclei density, nuclei texture, and in voxel dimensions (Fig.1.2a, see table 1.1). In total, we annotated 4414 nuclei in 3D.

We enhanced the contrast of all datasets crops using our local contrast enhancement method. Crops of $64 \times 64 \times 64$ voxels were extracted (Fig. 1.2b).

Data augmentation Data augmentation is a critical technique for training 3D CNNs as it enhances model robustness and generalization by artificially increasing the diversity of

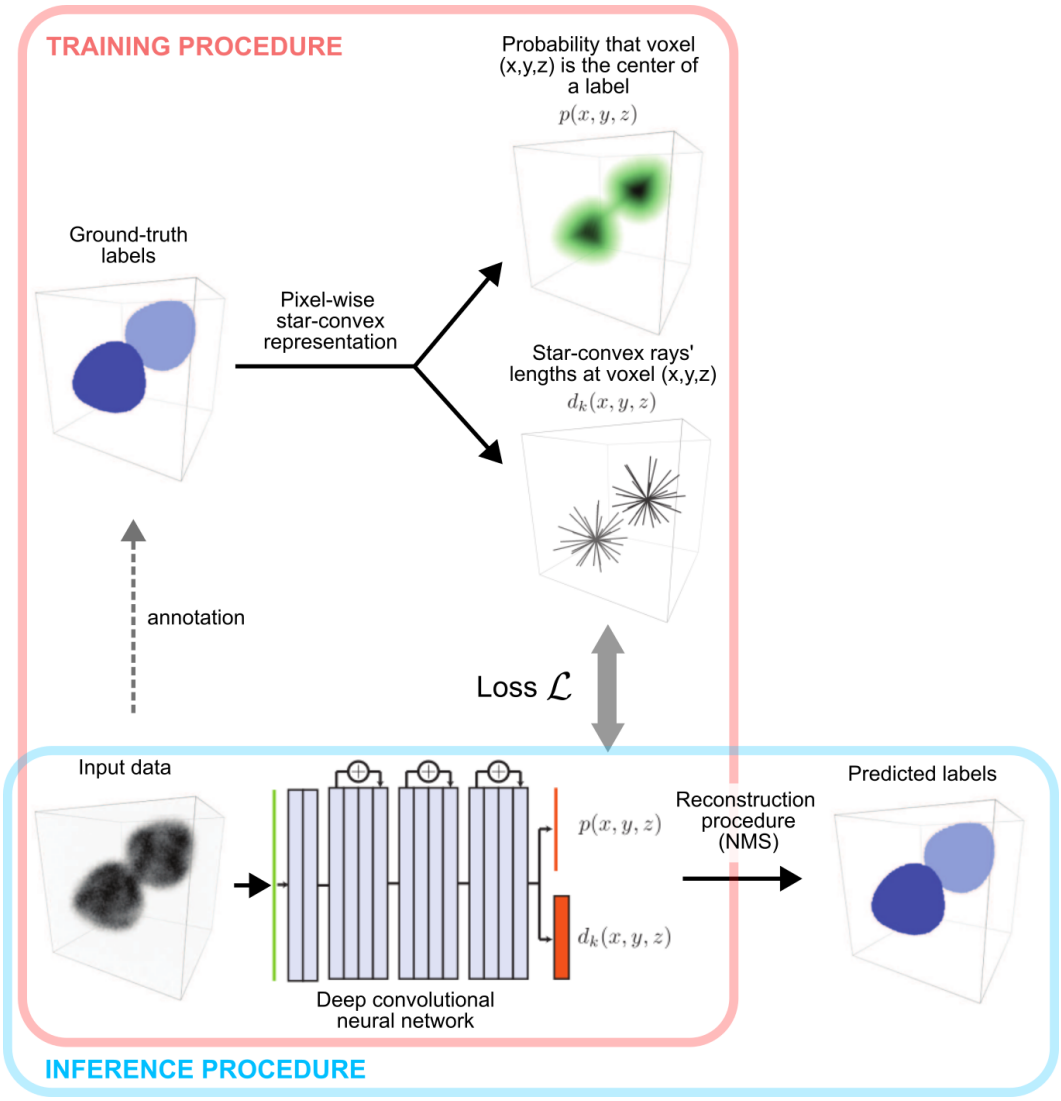

Figure 1.1: **Outline of the training and inference procedures of the Stardist3D algorithm.** During training, ground-truth labels (e.g obtained via manual annotation) are transformed into continuous maps of pixel "centerness" and star-convex-representation rays' lengths, which are compared to the prediction of a deep convolutional network taking raw images as input. The comparison yields a loss term $\mathcal{L}$ which is used to update the network's weights. During inference, the two predicted maps can be reconstructed into labels using non-maximum suppression (NMS).

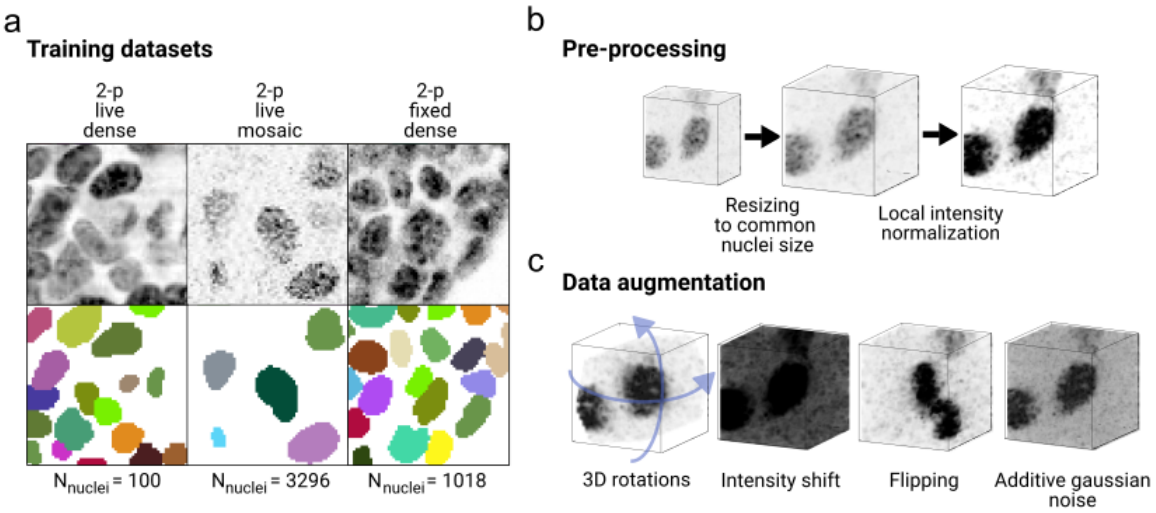

Figure 1.2: **Training procedure of our custom StarDist3D. (a)** Three different datasets were annotated and used for training. Images show exemplary annotated patches (top row: intensity images, bottom row: label images). **(b)** Before training, images were resampled to a common voxel size of $0.62 \times 0.62 \times 0.62\,\mu m^3$. Local contrast enhancement was used to harmonize the intensity distribution within the images. **(c)** During training, data augmentation was used to maximize training input.

training data [94]. By applying transformations such as rotations, scaling, translations, and other affine operations directly in 3D space, the augmented data better represents real-world variations in object orientation, size, and position. This approach is particularly effective in overcoming challenges posed by limited datasets, which are common in 3D applications [95]. Furthermore, these transformations preserve the semantic meaning of the data while exposing the model to a broader range of spatial configurations, which promotes invariance to global orientations and scale differences. This reduces overfitting and enhances predictive accuracy by improving the network's ability to generalize across unseen scenarios.

During training, we applied classical augmentation transforms, i.e 3D rotations, flipping, intensity shifts, and additive Gaussian noise (of variance 0.05), as illustrated on Fig. 1.2c, using the library Augmend [96], provided within the Stardist3D suite. The intensity and relative occurrence of each transform was chosen empirically by starting from default parameters and visually validating the plausibility of tuned parameters over many transformed images. We note that additive Gaussian noise does not fully represent the noise distribution observed in the datasets (closer to a mix of Poisson and Gaussian noise). However, other noise types were not available in the library used to perform augmentations. Note that the parameters and relative frequency of occurrence of each augmentation transform could be more systematically using the validation dataset (e.g in a k-fold cross validation setting).

Applying scaling transformations to crops during training is often regarded as the best way to achieve robustness in the case of a wide object size distribution. This is typically the case when several raw datasets acquired in different experimental conditions are combined, particularly due to different pixel sizes across datasets. However, we found in our case that

the size distribution related to a single dataset is narrow enough that robustness is better achieved by first rescaling all training datasets to a common isotropic nuclei size [95, 97].

1.3.2.3 Inference

During inference, a new dataset simply needs to be rescaled to reach that common nuclei size. We thus rescaled all crops to an isotropic voxel scale of $0.62\,\mu m/pix$ in all directions using linear interpolation.
We report inference times of around 4 seconds and 3 minutes on a computer with a Nvidia RTX A6000 GPU and a standard scientific laptop with an 11th Gen Intel Core i7-1185G7 CPU respectively for a typical 3D volume of shape $300\times512\times512$. For the reconstruction procedure performed from the neural network's outputs (see Fig. 1.1), we report times of 15 and 25 seconds on a computer with an Intel Xeon Gold 6242R CPU and a standard scientific laptop with an 11th Gen Intel Core i7-1185G7 CPU respectively for the same 3D volume comprised of $\sim$6000 nuclei.

1.3.2.4 Assessing 3D performance using simplified 2D validation datasets

To assess the performance of the training, we manually annotated 2D planes at different depths throughout a sample (see Fig.1.3a) and compared the result of StarDist3D to the ground truth.

We show for four different samples on Fig.1.3b that the 2D score obtained by sampling XY planes is very close but lower than the 3D score, as for some nuclei, the 2D and 3D annotations do not exactly coincide, which can artificially inflate the count of false positives/negatives. Thus, by computing the IoU score of 3D objects based on 2D images, we obtain a lower bound of the estimated score.

As manually annotating many dense organoid datasets for validation is a slow and cumbersome process, we designed a validation method based on comparing the 3D prediction of StarDist3D with 2D annotations in several XY planes along the z-axis. We validate this approach on Fig.1.3(a-b) using four different samples in which we chose XY planes spaced evenly along the z-axis, and on which nuclei were annotated in 3D. The f1 score obtained by comparing the 3D prediction and the 3D ground-truth is well approximated by the f1 score obtained by comparing the 2D annotations and the 2D sliced annotated segments, with at most a 5% difference between the two scores. To check for eventual biases due to the original anisotropy of the image, we also plot the f1 scores using 2D annotations in YZ and XZ (yellow and pink bars). As the score computed that way remain consistent with the 3D and XY scores, this confirms that the quality of a single XY slice of a predicted segmented nucleus is a good proxy for quality of its 3D reconstruction.

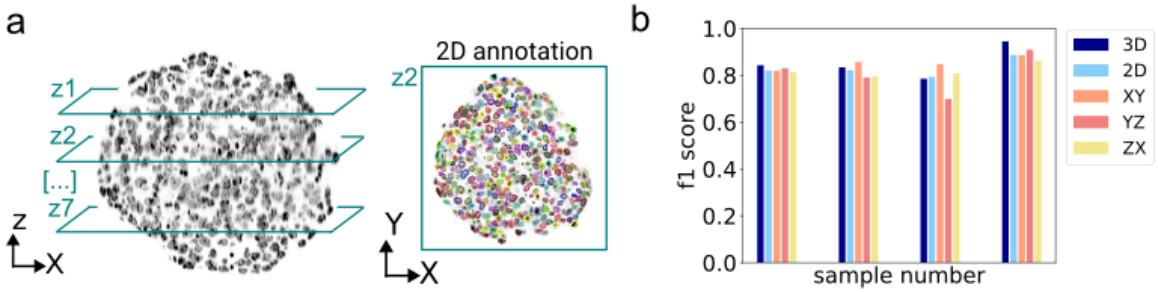

Figure 1.3: **2D annotations can be used to estimate a close lower bound to the performances of StarDist3D. (a)** We chose nuclei belonging to 2D planes along the z-axis and annotated them in 3D. **(b)** The interest of the protocol is validated by comparing the F1 score of a 3D prediction compared to either the 3D or 2D annotations.

1.4 Results

1.4.1 Dual-view fusion allows imaging nuclei *in toto* in fixed samples

Two-photon microscopy allows imaging up-to 150 μm in depth (Fig.1.4b), after which nuclei cannot be resolved, which manifests by a sharp dip in the FRC-QE metric (Fig.1.4c). As this is not sufficient for *in toto* imaging, we fused dual-view pairs using a block-matching registration algorithm. This allows recovering the full samples (e.g 300 μm in diameter on Fig.1.4). The fused stacks show an approximately four-fold intensity reduction at the mid-plane relative to outer edges but maintained consistent FRC-QE quality metrics across depth, as shown on Fig.1.4c.

As information seems to be roughly preserved across depth, this motivates the use of a contrast enhancement method to homogeneize the overall appearance of the signal in deeper planes with respect to the appearance of the brightest planes.

1.4.2 Thanks to a tailored preprocessing pipeline, a deep-learning algorithm allows nuclei segmentation in deep gastruloids

Dual-view fusion combined with local contrast enhancement are essential to segment nuclei consistently at all depths Both during training and inference, we used local contrast enhancement to harmonize the intensity profiles within the images. This step was crucial to homogenize the intensity distribution across and inside datasets during training, and improved segmentation quality by reducing the number of false detections at depths plagued by low SNR. Comparing the recall and precision to the ones obtained with global contrast enhancement, i.e., image normalization based on the percentiles of the whole image, our local contrast enhancement scheme mitigates both false positives and false negatives in deeper z planes (Fig.1.5a-b). Local contrast enhancement appears to pro-

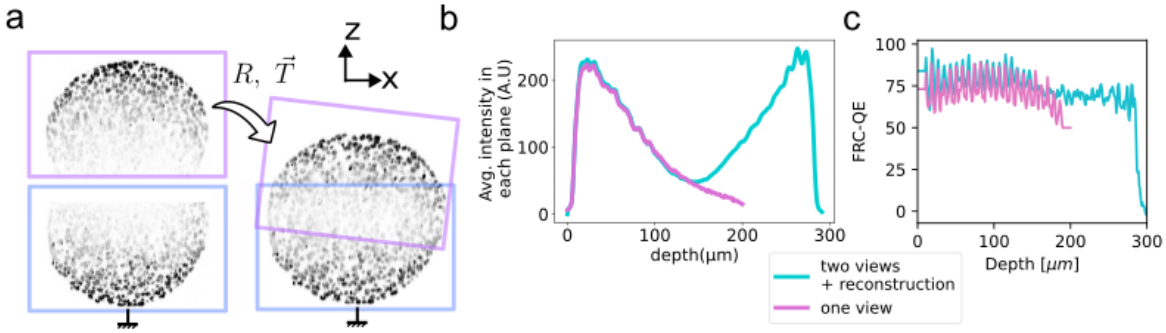

Figure 1.4: **The spatial registration and fusion procedure. (a)** To allow *in-toto* imaging, samples are imaged from both sides and the two views are registered using rigid 3D transformations (represented by a rotation matrix R and a translation vector $\vec{T}$) with a block-matching algorithm. Individual cells are visible in the mid-plane as shown for a gastruloid of typical diameter of 250 μm. **(b)** Average intensity per plane and FRC-QE for one- and dual-view images.

vide a boost of up to 20% for the deepest planes, and 10% for the extremal planes. It is still necessary to image from two views to recover all the nuclei, as shown Fig 1.5c, even using the local contrast enhancement method.

Our custom StarDist3D model reliably segments nuclei in depth The results of the performance estimation method based on simplified 2D annotations is shown on Fig. 1.6. The model has consistent performances across depth above 200 μm (Fig. 1.6a), with a significant drop in shape reconstruction performance above 50% IoU threshold. To characterize the detection performance at 50% IoU threshold, we computed the recall and the precision (respectively quantifying how many relevant items are retrieved and how many retrieved items are relevant), and the F1 score as the harmonic average between the two. We show in Fig.1.6b that our trained model recovers a constant precision and recall across the full depth, and the F1 score is superior to 0.8 up to IoU threshold 0.5. Overall, our custom model achieved an F1 score of $85 \pm 3\%$ at 50% IoU. Using a variety of annotated datasets, dual-view imaging and a local contrast enhancement algorithm, we trained a powerful StarDist3D model that shows robust segmentation performances regardless of depth, sample size, and experimental conditions.

1.5 Discussion

Segmenting cell nuclei in deep 3D organoids presents significant challenges, particularly due to the combination of low SNR and densely packed structures, which can make individual nuclei difficult to resolve. We created a custom training and inference strategy with StarDist3D, a supervised deep-learning method, to address these difficulties, demonstrating its ability to segment objects under these challenging conditions (see e.g Fig. 1.7(a)). However, limitations in training data, annotation variability, and inherent algorithmic biases still pose notable challenges.

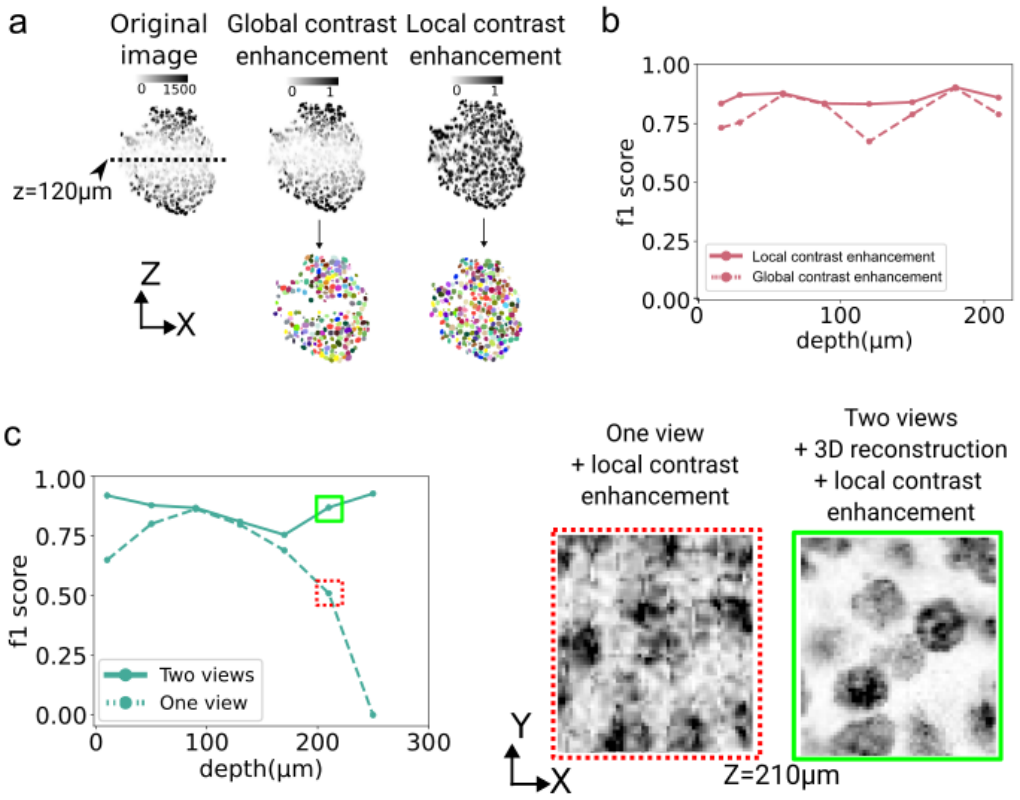

Figure 1.5: **Dual-view fusion and local contrast enhancement proved crucial to improve performance in the deepest planes. (a-b)** Segmentation results with global and local contrast enhancement. The latter allows segmenting all cells in the sample mid-plane, which greatly improves performance. **(c)** Local contrast enhancement alone is not sufficient to grant satisfactory performance in deeper planes.

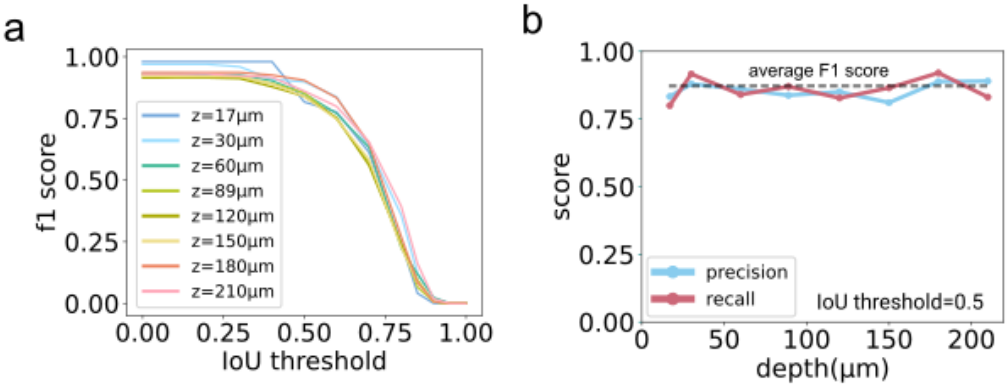

Figure 1.6: **Perfomance of our custom StarDist3D model. (a)** Our model achieves consistant shape reconstruction performances across depth. **(b)** The model reaches an average F1 score of 85% in planes above 200 μm.

One major difficulty in training a deep-learning model for this task is that low SNR and extreme object crowding lead to cases where even human experts struggle to distinguish individual nuclei without additional temporal context. As a result, manual segmentation is particularly arduous and time-consuming, which limited the amount of training data at out disposal.

Moreover, ground truth annotations contain ambiguities and inconsistencies, particularly when different annotators are involved. Prior work has shown that manual annotations often introduce biases that can propagate into deep-learning models [98]. For example, we found several instances of systematic errors in human annotations, such as the tendency to over-annotate in the XY plane near the extremities of objects along the Z-axis (see Fig. 6(e)), which results in elongated or tubular appearances of actually roundish structures. This behavior is also reported to appear in major benchmark datasets in [99].

These limitations are reflected in Stardist segmentations, which we qualitatively illustrate on Fig. 1.7. Limitations lead to errors such as nuclei merges (primarily in dense clusters or when small nuclei are adjacent to larger ones, see Fig. 1.7(b)) and splits (often affecting large or elongated nuclei, Fig. 1.7(c)). In the case of most ill-segmented elongated nuclei (e.g the nucleus in the middle of Fig. 1.7(c)), note that the star-convex paradigm of Stardist3D is fundamentally competent in representing such shapes, but the relatively low occurrence of such morphologies in the training datasets limits the performance of the model. Additionally, due to Stardist's star-convex approximation, segmented shapes tend to be slightly smoothed and often miss actual boundaries by $\sim$1 voxels. Nuclei undergo significant morphological changes during mitosis. In particular, chromosomes condense into a 3D ring (or torus) during the metaphase [100] (Fig. 1.7(f)), which is often oversegmented (split), see Fig. 1.7(c, e).
Given these limitations, our downstream analysis of nuclear morphology will focus on "low-order" descriptors (e.g., those based on ellipsoidal approximations) and rely on local averaging to mitigate the influence of segmentation errors and outliers.

Several approaches could enhance our segmentation strategy with StarDist3D. One promising avenue is leveraging the availability of additional input channels in the current version of the software. Particularly, designing experiments with membrane markers could improve boundary delineation and separation of clustered nuclei by providing exclusive signal between nuclei boundaries. Sparse annotation-based training, as proposed in [101], could allow model refinement with fewer annotations by setting up annotation sessions in which annotators are purposefully focusing only on hard cases. However, this approach is limited by the fact that some objects remain fundamentally unannotable even to human experts due to poor SNR, high object packing, and lack of temporal context in the case of fixed datasets.
For live datasets, the second part of this thesis will explore algorithms that leverage images temporal coherence to refine or validate segmentations.

Alternative segmentation strategies may also provide improvements. Methods such as InstanSeg [102] and EmbedSeg [103] offer promising advancements, but the former is currently limited to 2D applications, while the latter remains untested in this context. More relevant is u-segment3D [104], which reconstitutes 3D segmentation by combining multiple 2D segmentations from all directions. Given that pretrained Stardist and Cellpose models

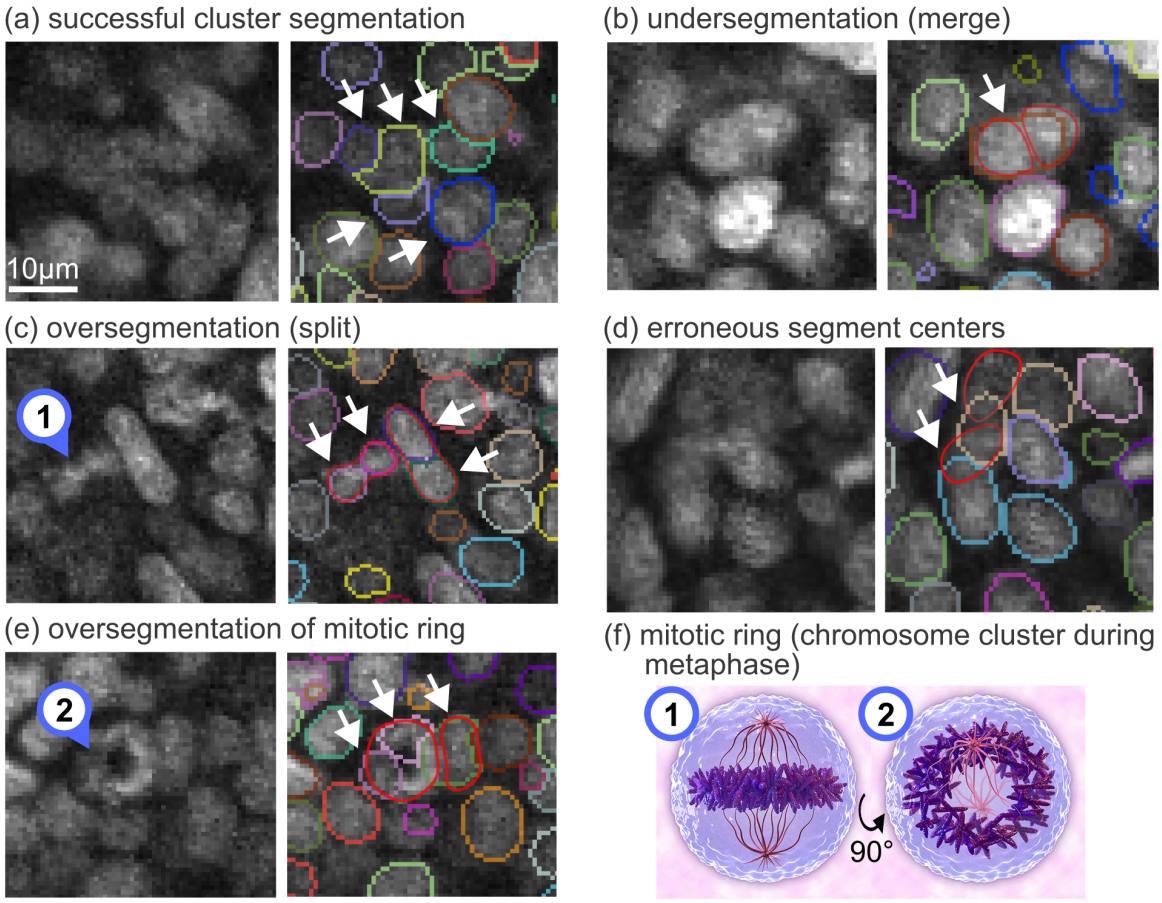

Figure 1.7: **Illustrative examples of the performances of our custom StarDist3D model. (a)** A challenging nuclei cluster is successfully segmented. **(b)** Undersegmentation (two nuclei segmented as one). Red lines indicate real contours. **(c)** Two adjacent cases of oversegmentation (a nucleus, often with an elongated shape) is segmented as two or more. **(d)** Erroneous estimation of nuclei centers in a cluster, leading to a chain of errors. **(e-f)** During mitosis, nuclei appear as tori (f), which are often oversegmented (see (c)). In this case, the ring is segmented in two, and the segmentation of an adjacent nucleus overlaps with the rest of the ring. (f) was modified from [100].

are continuously being trained and improved on new massive 2D datasets to be released to the community, methods exploiting 2D models could open the possibility to tap into the richness and impressive volume of 2D datasets.

Chapter 2

Probing cellular micro-environment: integrating nuclei segmentation with mechanics and gene expression analysis

2.1 Chapter overview

This chapter focuses on analyzing mechanical and genetic properties across multiple scales, from individual nuclei to entire tissues, using fixed datasets. We integrate the custom segmentation tools developed in Chapter 1 into a comprehensive processing pipeline and present practical applications demonstrating the pipeline's effectiveness. Additionally, we introduce "Tapenade," a user-friendly Python package integrated with Napari plugins, enhancing accessibility and enabling easy visualization and analysis of the methods developed.

This chapter is part of the publication [70]:
GROS, Alice*, VANARET, Jules*, DUNSING-EICHENAUER, Valentin*, et al. A quantitative pipeline for whole-mount deep imaging and multiscale analysis of gastruloids. bioRxiv, 2024, p. 2024.08. 13.607832. (*= co-first authorship)

My contributions:

1. designed a method to remove intensity artifacts caused by scattering or optical aberrations, enabling accurate measurement of fluorescence intensity as a readout for gene expression even in deeper planes
2. showed that characteristic spatial patterning of key genes related to cell differentiation into early tissues can be measured using this method
3. showed that Brachyury positive domains have increased density, and that nuclei elongate along the interface between low- and high-density domains

2.2 Motivations and Challenges

The quantification of the mechanical and genetic microenvironment of cells in biological tissues has been extensively studied [105, 106], leveraging advanced imaging and analytical techniques. Fluorescence microscopy has become a cornerstone for visualizing gene expression patterns in tissues. It correlates fluorescence intensity with gene expression, but its accuracy diminishes with depth due to optical diffraction and scattering [107]. This limitation is often addressed by studying shallow samples or ignoring depth-related biases, leaving deep tissue datasets underexplored.

In three-dimensional models such as organoids, the nucleus often occupies an unusually large volume relative to the cell size [108]. This makes nuclear deformation a reliable proxy for cellular deformation, which reflects mechanical constraints within the microenvironment [109]. Mechanical parameters such as stiffness and strain have been quantified using both endogenous cell-generated forces and external mechanical inputs. For instance, cells probe their surrounding stiffness dynamically, with tissue mechanics varying spatially and temporally during processes like morphogenesis [110].

A key interest in *in toto* imaging is that it allows quantifying expression gradients of genetic markers and proteins of interest across entire organoids. In the case of gastruloids, gradients of differentiation emerge during Wnt activation, prior or concomitant with symmetry breaking [36, 111–113]. Due to the previously mentioned limitations of light-sheet microscopy for *in toto* coverage and quantitative intensity analyses for gastruloids beyond $100 - 200\,\mu m$ diameter, investigations of such gradients have thus far been restricted to analyses of low resolution 2D images [111, 112]. Because of this limitation, it is still unknown when gradients of differentiation are established during gastruloid development.

2.3 Methods

2.3.1 Comparing fields of mechano-genetic properties at different scales

To probe volumetric signals at different length scales σ ranging from the average nuclei size to the width of the organoid, we convolve signals with a Gaussian kernel of size σ while carefully managing boundary artifacts. We propose two ways to compute convolutions depending on whether the signal is dense (e.g an image representing a signal intensity) or sparse (e.g nuclei volumes defined only locally on nuclei centroids).

2.3.1.1 Dense data

We apply convolution with a normalized Gaussian kernel to dense signals. For a given voxel at position $\mathbf{x} = (z, y, x)$, its value $I(\mathbf{x})$, whether scalar or vector, is transformed according

to:

$$I^{\sigma}(\mathbf{x}) = \sum_{\mathbf{x}' \in [-a,a]^3} k_{\sigma}(\mathbf{x}') I(\mathbf{x} - \mathbf{x}'), \tag{2.1}$$

with $k_{\sigma}(\mathbf{x}') = (2\pi\sigma^2)^{-3/2} \exp\left(-\left\|\mathbf{x}'\right\|^2 / 2\sigma^2\right)$, and a the convolution window size. In our implementation, we choose $a = 3\sigma$.

When a binary mask (e.g defining the inside of the organoid) is available, we apply a masked convolution with a correcting normalization factor. Let M be a binary masking function, e.g $M(\mathbf{x}) = 1$ for voxels at positions $\mathbf{x}$ inside the organoid, and 0 everywhere else, we compute

$$I^{\sigma,M}(\mathbf{x}) = \frac{\sum_{\mathbf{x}' \in [-a,a]^3} k_{\sigma}(\mathbf{x}') I(\mathbf{x} - \mathbf{x}') M(\mathbf{x} - \mathbf{x}')}{\sum_{\mathbf{x}' \in [-a,a]^3} k_{\sigma}(\mathbf{x}') M(\mathbf{x} - \mathbf{x}')}. \tag{2.2}$$

This additional normalization factor prevents the artifactual decay of the convolved signal close to the mask boundaries by "excluding" voxels outside the mask from the convolution. Note that the masked formulation allows the user to compute values of a quantity outside its domain of expression. For instance, if both a mask M_{inst} of nuclear instances and a mask $M_{in/out}$ that marks the inside and outside of the sample are provided, a quantity like the nuclear Hoechst expression can be computed in the full region defined by $M_{in/out}$, even though it only appears in the nuclei. This requires setting $M = M_{inst}$ and applying equation (2.2) at positions $\mathbf{x}$ where $M_{in/out}(\mathbf{x}) = 1$.

Dense smoothing is illustrated e.g. on Fig. 2.3, on which the nuclear volume fraction field is computed by smoothing the mask M_{inst} of nuclear instances.

2.3.1.2 Sparse data

For sparse data, i.e quantities defined locally at n specific positions $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, we compute the convolution and masked convolution in a similar manner, but with a mask M that is non-zero only at the positions at which the sparse signal is defined (e.g nuclei centroids). This formulation allows the user to choose between studying the sparse signal (i) at its original positions, i.e by computing the above convolution at positions $\mathbf{x} \in \mathbf{X}$, (ii) at resampled positions, e.g on a uniform grid for convenience, or (iii) as a dense continuous field across the whole domain of the gastruloid.
We provide a second implementation of the masked convolution specifically for sparse data, which is more convenient when the smoothed sparse signals should also be smooth (e.g when it should be defined at its initial positions or on a regular grid). This implementation uses the Kd-Tree data structure implemented in *Scipy* [114] to efficiently find neighbor relationships from the positions at which the sparse signal is defined, which is used to optimize the computation of the convolution (2.2).

Sparse smoothing is illustrated e.g. on Fig. 2.4(a), on which true strain tensors (which is a quantity defined for each segmented nucleus, and thus only "exists" at their centroids) are sparsely smoothed, which results in a sparse field.

2.3.2 Measuring gene expression from biased fluorescence intensity

2.3.2.1 Multiscale model of intensity spatial variations

Although our dual-view imaging pipeline delivers *in toto* coverage, images of the ubiquitous Hoechst nuclei stain still exhibited intensity gradients along the z-axis and between outer and inner cell layers (see e.g. Fig. 2.8). We attributed these gradients to optically induced heterogeneities and developed a correction method that re-normalizes the channels of interest based on the local nuclei intensity.

The main sources of artifacts are (local) variations in optical paths and cell density, that usually manifest as decreased intensities towards deep tissue regions. Artifactual gradients can also appear during the registration of two non-colinear views, which further complexifies the expected distribution of artifacts and resulting intensity gradients.

We model the intensity signal $I(\mathbf{x})$ from a given channel as

$$I(\mathbf{x}) = A(\mathbf{x})S(\mathbf{x})M_{inst}(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{2.3}$$

where $A(\mathbf{x})$ is the field representing the artifactual intensity variations, $S(\mathbf{x})$ is the gene expression field, which provides a continuous representation of the cell-scale variations of the biological signal, $M_{inst}(\mathbf{x})$ is a binary mask of the signal instances (e.g nuclei or membranes), i.e $M_{inst}(\mathbf{x}) = 1$ if $\mathbf{x}$ is inside an instance and 0 otherwise, and $\varepsilon(\mathbf{x})$ is an additive noise term which models (i) acquisition noise (spatially correlated on lengthscales less than 2-3 voxels) and (ii) variations in the cellular signal due to cellular structural heterogeneity (spatially correlated on lengthscales typically small than the cell size). This model is illustrated on Fig. 2.1(a).

2.3.2.2 Modeling intensity decay with depth

We hypothesize that the Hoechst signal $S_h(\mathbf{x})$ is globally homogeneous throughout the volume, i.e $S_h(\mathbf{x}) = S_{h,0}$, and that the field of optical artifacts A has large scale variations (typically superior to a few cell diameters). In particular, direct observations of homogeneous signals like Hoechst with different emission wavelengths, as displayed of Fig2.2(a-b), indicate that optical artifacts are mostly dominated by exponential decay with depth with a wavelength-dependent decay length. A second order contribution comes from scattering effects related to heterogeneities in photon's optical paths, cellular density, or to geometrical considerations, e.g proximity to the sample's boundaries. For every wavelength λ and every signal, we thus model A to first order as

$$A(\mathbf{x}, \lambda) = e^{-z/d} + a(\mathbf{x}, \lambda), \tag{2.4}$$

with d the wavelength-dependent decay length. The second term a on the right hand side corresponds to second order corrections to the exponential decay model.

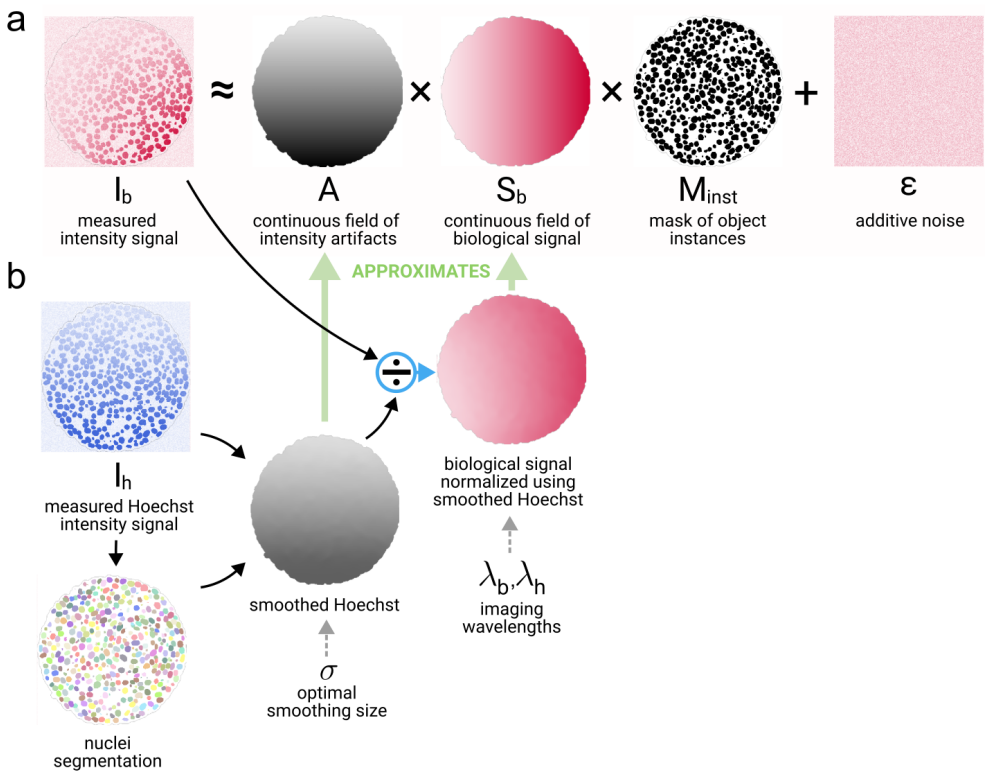

Figure 2.1: **Illustration of the intensity model and of the normalization procedure based on a ubiquitous signal. (a)** We model any arbitrary intensity signal I_b by decomposing it into three terms: (i) a continuous field A representing local intensity artifacts, (ii) a field S_b representing a continuous depiction of the biological signal, and (iii) an instance mask M_{inst}, which localizes the biological objects (e.g. nuclei or membranes) from where the biological signal originates. The first two fields are assumed to have variations at length scales equal or above the typical cell size. Finally, an additive term ε which varies at sub-cellular length-scales and encompasses all the remaining voxel-scale signal (e.g. measurement noise or intra-cellular gene expression heterogeneities) completes the model. **(b)** An ubiquitous signal (here Hoechst) can be used to approximate the field of intensity artifacts, which can be removed from the original intensity signal.

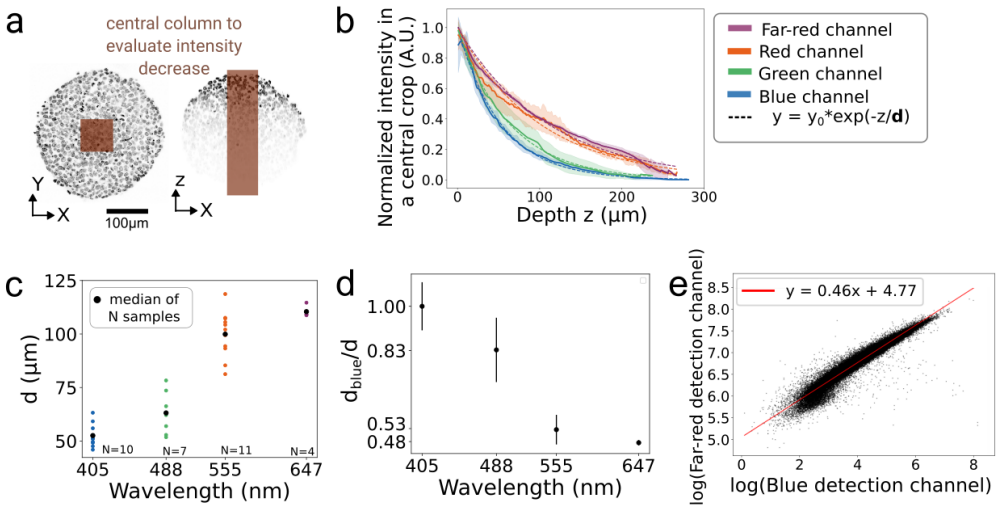

Figure 2.2: **The wavelength-dependent normalization method. (a)** Intensity decrease as a function of depth and wavelength is evaluated in samples immunostained with histone markers in four colors. Median intensity is measured in a central column to minimize border effects. **(b)** Normalized intensity in each channel, with shaded regions showing standard deviation. An exponential fit extracts the characteristic length d. **(c)** Length d as a function of wavelength, with mean and standard deviation. **(d)** Ratio $r = \frac{d_{blue}}{d}$ is plotted, with mean and standard deviation. **(e)** Far-red channel intensity vs. Blue channel intensity (log scale). Each dot represents a $10 \times 10\,\mu m^2$ square across N=5 samples.

In accordance with Beer-Lambert law, the value $d(\lambda)$ also likely depends on experimental parameters $\mathcal{E}$, like the nature of the mounting medium or the settings of the experiments, in a multiplicative fashion, i.e

$$d(\lambda, \mathcal{E}) = f(\mathcal{E})d'(\lambda), \tag{2.5}$$

so that in similar imaging scenarios, ratios of decay lengths at two wavelengths $d(\lambda_1, \mathcal{E})/d(\lambda_2, \mathcal{E})$ only depend on λ_1 and λ_2.

To estimate the decay lengths across the range of wavelengths in use, we measured the intensity decay in samples with ubiquitous nuclei stainings in four emission wavelengths: Hoechst for the blue channel ($\lambda = 405nm$), H2B-GFP amplified for the green channel ($\lambda = 488nm$), SPY-DNA555 or the endogenous H2B-Tomato for the red channel ($\lambda = 555nm$) and DRAQ5 for the far-red channel ($\lambda = 647nm$). Stainings for different wavelengths were done on separated samples to avoid signal crosstalk between the channels.

All samples were imaged with the same excitation and detection settings. For each sample, we extracted the intensity profile in a central column along the z-axis to minimize boundary artifacts, and fitted equation (2.4) to recover the decay lengths $d(\lambda)$ (Fig. 2.2(a-c)). We observe a monotonous relationship between median values of d across different samples and wavelengths. On Fig. 2.2(d), we plot ratios of decay lengths with the decay length of the Hoechst channel d_{blue}. In our model, these ratios are independent of imaging conditions, and the monotonicity and smoothness of the trend suggests that ratios $d_{blue}/d(\lambda)$ in the range $\lambda \in [405nm,\ 647nm]$ could be estimated by interpolating our results.

To confirm our measurements, we also computed decay lengths via a second method:

notice that the model of equation (2.4) leads to the following identity for two ubiquitous signals I_1 and I_2 at wavelengths λ_1 and λ_2 (neglecting the noise term ε in equation (2.3) and the second order term from equation (2.4))

$$\log(I_1) = \frac{d(\lambda_2)}{d(\lambda_1)} \log(I_2) + cst. \tag{2.6}$$

New samples were imaged with both Hoechst ($\lambda = 405nm$) and the Far-red nuclei stain ($\lambda = 647nm$). There was no significant spectral overlap between the two signals. By fitting equation (2.6), we obtained another estimate of the ratio $d(\lambda = 405nm)/d(\lambda = 647nm) \approx 0.46$, coherent with the value of 0.48 obtained with the previous method (Fig. 2.2e).

2.3.2.3 Intensity normalization based on a ubiquitous signal

In the experiments, it was not possible to image ubiquitous nuclei signals at the same wavelengths as relevant biological signals, while the Hoechst staining was often easily available. We thus developed a correction scheme based on the Hoechst staining to separate the contribution of artifactual gradients from true gradients in gene expression. The rationale of the scheme stems from the fact that equation (2.6) also allows us to relate two ubiquitous signals up to a multiplicative constant

$$I_1 \propto I_2^{d(\lambda_2)/d(\lambda_1)}. \tag{2.7}$$

Note that equation (2.7) is still valid for ubiquitous signals that result from our fusing procedure. In this case, the total intensity artifact field A is the sum of two exponential terms that decay in opposite directions, but equation (2.7) applies if one of the exponential term can always be neglected in a domain where the other is non-negligible, i.e the distance between the upper and lower ends of the samples is superior to the decay length of the exponential terms.

The Hoechst signal could therefore be used to predict the intensity profile from another "virtual" ubiquitous signal imaged at some wavelength $\lambda \neq \lambda_{blue}$ that coincides with the wavelength of the biological signal of interest.
First, we compute the Gaussian average of the reference Hoechst signal I_h at a scale σ using equation (2.2), in which we keep only voxel intensities that belong to the nuclei using a mask M_{nuc},

$$I_h^{\sigma, M_{nuc}}(\mathbf{x}) \approx S_{h,0} A(\mathbf{x}). \tag{2.8}$$

Here, σ is chosen large enough to average out the additive noise ε but small enough that the structure of the field of optical artifacts A is preserved. The practical choice of σ is discussed below.

We then normalize other biologically relevant intensity signals I_b by the Gaussian-

averaged nuclei field $\left(I_h^{\sigma,M_{nuc}}(\mathbf{x})\right)^{d(\lambda_{blue})/d(\lambda_b)}$ to get

$$\begin{aligned}\hat{I}_b(\mathbf{x}) &= \frac{I_b(\mathbf{x})}{\left(I_h^{\sigma,M_{nuc}}(\mathbf{x})\right)^{d(\lambda_{blue})/d(\lambda_b)}} \\ &= M_{inst}(\mathbf{x})\frac{S(\mathbf{x})}{S_{h,0}^{d(\lambda_{blue})/d(\lambda_b)}} + \varepsilon'(\mathbf{x}),\end{aligned} \tag{2.9}$$

with ε' the normalized version of the additive noise ε. This step effectively removes the field of optical artifacts from the signal. The choice of σ used for the Gaussian averaging of the ubiquitous signal depends on the length scale at which the field of artifacts A varies. Since this is not known in practice, we choose in our implementation the value σ that leads to the most homogeneous (i.e flat) normalized ubiquitous signal across all z-planes. Formally,

$$\begin{cases}\sigma = \mathrm{argmin}_\sigma \,\mathrm{MAD}\left(\mathcal{I}_h^\sigma\right) \\ \mathcal{I}_h^\sigma = \left\{\mathrm{median}\left(I_h^{\sigma,M_{nuc}}(z=Z)\right)\right\}_{Z\in[0,Z_{max}]},\end{cases} \tag{2.10}$$

with MAD the raw median absolute deviation and Z_{max} the depth of the last z-slice.

With the aim of comparing normalized signals across different samples acquired with the same imaging setup, we further rescale the normalized signal by a global multiplicative factor designed to mitigate the sample-dependent impact of the normalizing factor $S_{h,0}^{d(\lambda_{blue})/d(\lambda_b)}$ in equation (2.9). We multiply each normalized signal by a multiplicative factor such that the median intensity of the raw Hoechst signal computed in a region of brightest intensity is transmitted to the normalized Hoechst signal. The rationale is that the median intensity in the brightest regions reflects the expected intensity in the absence of artifacts, e.g. in outer cell layers close to the cover glass.

This normalization scheme accounts for heterogeneities in gastruloids of arbitrary shape as well as local cell density and thus variations in the optical paths of excitation and fluorescence light rays that cannot be captured by a simple geometrical model. We illustrate the typical normalization workflow on Fig. 2.13 on real data, in which the ubiquitous Hoechst stain is used to normalize the fluorescence intensities of E-Cadherin and Brachyury expression signals.

2.3.3 Quantifying cell deformation

To quantify object-scale deformation, we fit an ellipse to each segmented object. This can be done efficiently by computing the inertia tensor I. For a given object composed of N voxels at positions $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, we first compute the positions of the voxels relative to the object's centroid $\vec{c}$, $\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N\}$, with $\vec{r}_k = \mathbf{x}_k - \vec{c} = (r_{k,1}, r_{k,2}, r_{k,3})$. The inertia

tensor is then defined as

$$I_{i,j} = \sum_{k=1}^{N} (r_{k,i} r_{k,j} - \delta_{i,j} r_k^2), \tag{2.11}$$

with $r_k^2 = r_{k,1}^2 + r_{k,2}^2 + r_{k,3}^2$ and $\delta_{i,j}$ the Kronecker delta. The eigenvectors and eigenvalues of the inertia tensor correspond to the principal axes and moments of inertia of the object. In particular, the principal axes of I are the same as the principal axes of the ellipse that best fits the object. Note that this prevents the inertia tensor from accurately quantifying the deformation of shapes that deviate significantly from an elliptical model. However, in practice, this estimator exhibits sufficient robustness against minor segmentation errors, ensuring that small perturbations, such as those introduced by segmentation inaccuracy in StarDist3D, have little impact on the principal directions and lengths.

The lengths of the ellipse's semi-axes can be obtained from the moments of inertia: if (I_1, I_2, I_3) are the moments of inertia associated with the principal axes $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$, the lengths of the corresponding semi-axii (L_1, L_2, L_3) are given by

$$L_i^2 = \frac{5}{2N}(I_1 + I_2 + I_3 - 2I_i), \quad i = 1, 2, 3. \tag{2.12}$$

Computing the inertia tensor for each object leads to a sparse tensor field that can be studied at arbitrary scales using the sparse Gaussian convolution equation defined previously. For instance, it can be used to study local object alignments by focusing only on the largest semi-axis (the main axis of deformation of the object). In practice, quantifying local object alignment by averaging inertia tensors introduces a bias by over-representing large objects. To avoid this bias, we instead compute the true strain tensor from the inertia tensor. The true strain tensor T is defined as

$$T = V \; diag(T_1, T_2, T_3) \; V^T, \tag{2.13}$$

with V the matrix of eigenvectors of I and $T_i = \frac{1}{3} \log (L_i^3 / (L_1 L_2 L_3))$. The eigenvalues (T_1, T_2, T_3) of T represent the relative amount of deviation of each principal length from the radius of a sphere with the same volume as the object. Due to its relative nature, the true strain tensor is suited to compare the deformations of two objects with different sizes, and is thus appropriate for averaging in a local neighborhood.

Importantly, in the analysis of spatially averaged quantities derived from inertia or true strain tensors, e.g principal axis orientations, it is crucial to average the tensors themselves rather than their derived scalar or vector quantities. The desired quantities should then be computed from these averaged tensors at the appropriate scale. This approach preserves the tensorial nature of the data and ensures mathematically consistent results across different scales [115].

2.4 Results

2.4.1 Nuclei deformations bear spatial signatures of internal mechanical stresses

Tissue-scale patterns of cell/nuclei deformations in developing embryos or organoids are good indicators of the existence of mechanical stresses which can contribute to tissue morphogenesis [116] and affect in return cell differentiation [117]. To explore changes in tissue structure and organization during gastruloid development, we studied spatial heterogeneities in cell densities and cell deformations emerging during the first morphogenetic phase of gastruloids development, which consists in a global tissue elongation between 72 and 96h after their initial aggregation. Previous works [27, 36] have shown that such an elongation is associated with the polarization of the gastruloid, exhibiting a posterior pole rich in E-cadherin and constituted of a mixed population of both endoderm and nascent mesoderm tissues, and an anterior pole poor in E-cadherin constituted of cardiac and vascular mesoderm tissues which are mesenchymal. Here, to prevent any relaxation of cellular deformation caused by tissue fixation, we performed 3D acquisition of live gastruloids with an H2B-GFP marker.

Masked gaussian smoothing reveals patterning of properties related to cell packing from cellular to tissue scale The cell density field revealed a high level of heterogeneity at the nuclei scale, but analysis at different scales always showed the presence of two tissue-scale populations of low and high cell densities located respectively in the posterior and anterior part of the sample, separated by a sharp boundary characterized by the length scale of a few nuclei (see Fig. 2.5). These two populations and the boundary were also visible, although less clearly, from the nuclear volume and nuclear volume fraction fields, with the high-density population showing a higher nuclear volume fraction but lower nuclear volumes. The nuclear volume fraction indicates regions of high packing located in the high-density population, with a maximum of 61%, close to the theoretical value for random close packing of spheres (64%). As extra-cellular space volume fraction is low in such sample [27], this indicates high nuclei-to-cell volume ratios, which suggests that nuclei shapes are a good proxy for cell shapes, and thus nuclei could be relevant in cell-scale mechanotransduction processes.

Nuclei align along a stereotyped boundary between low- and high-density domains. To quantify cell deformation, we first computed the inertia tensor from each segmented nuclei instance and obtained the true strain tensor [118] for each cell (see Methods 2.3.3), which is a size-independent measure of object deformation. This justifies averaging these tensors among nuclei with large size variations by effectively removing the size bias. We plot the major axis (associated with the largest eigenvalue) of the true strain tensor for each nucleus on Fig. 2.3 at the nuclei scale. The major axis is oriented along the direction of the largest deformation of the nuclei. Close to the low-to-high density boundary, we observed that nuclei are elongated in the direction of the boundary, i.e perpendicular to the gradient

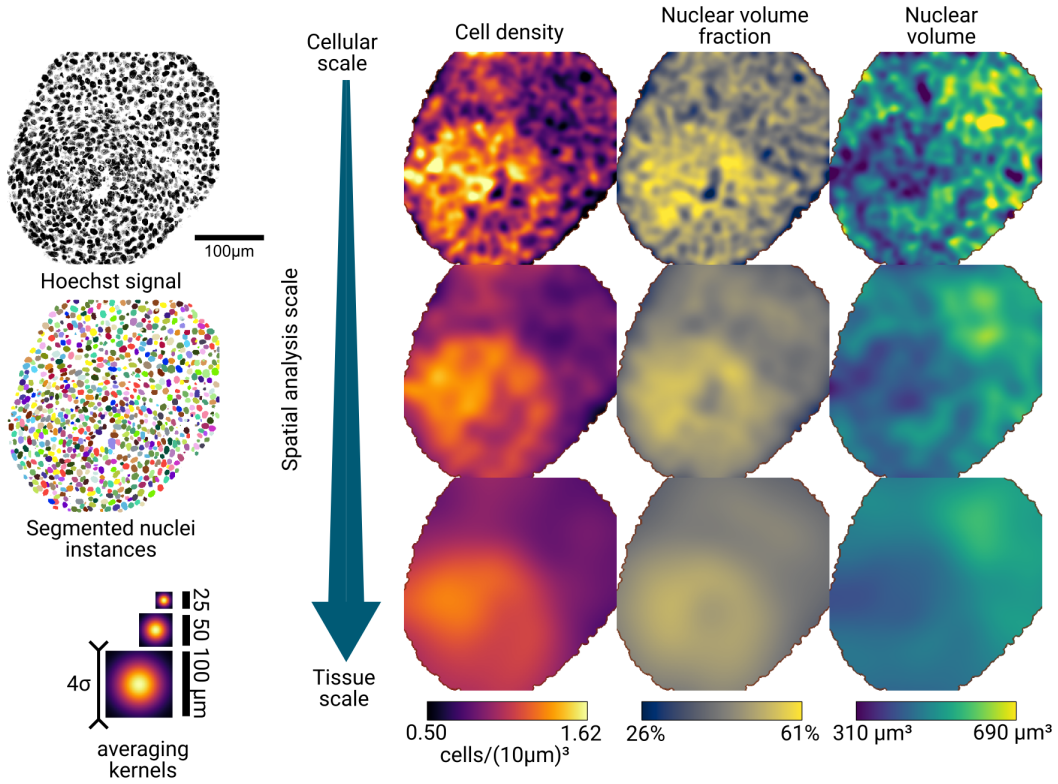

Figure 2.3: **Masked gaussian smoothing reveals patterning of properties related to cell packing from cellular to tissue scale.** 3D segmentation of nuclei allows quantification of cell density, nuclear volume fraction, and nuclear volume across scales, using Gaussian averaging kernels with $\sigma = 6,\ 12$, and $25\,\mu m$.

of cell density.

We further quantified this observation by computing the angle θ between the gradient of cell density and the major axis of the true strain tensors. In Fig. 2.4(a-b), we plot the continuous field representing the quantity $\cos^2(\theta)$ averaged at the tissue scale. For random orientations of the major axis with respect to the gradient of cell density, the average value of $\cos^2(\theta)$ is $1/3$, and the value approaches 0 or 1 for respectively perpendicular and parallel orientations. The field reveals a wide region with values below 0.1 around the boundary, confirming that nuclei are elongated along the tissue boundary. Other regions show comparatively smaller values of the local gradient amplitude or no tissue-scale alignment.

These descriptions appear reproducibly in several samples, see Fig. 2.5, and are consistent with an interpretation of the boundary as a region where mechanical stresses are generated between gastruloids anterior and posterior poles.

This analysis also shows how nuclei shapes at small scales are appropriate proxies for deformation inside and around tissue micro-structures like lumen and cavities, which allows for an accurate description of the local epithelial architecture. At the tissue scale, nuclei shapes can serve as a proxy for tissue mechanical stresses in the context of tissues modeled as visco-elastic liquids with residual elasticity in cell deformation [119].

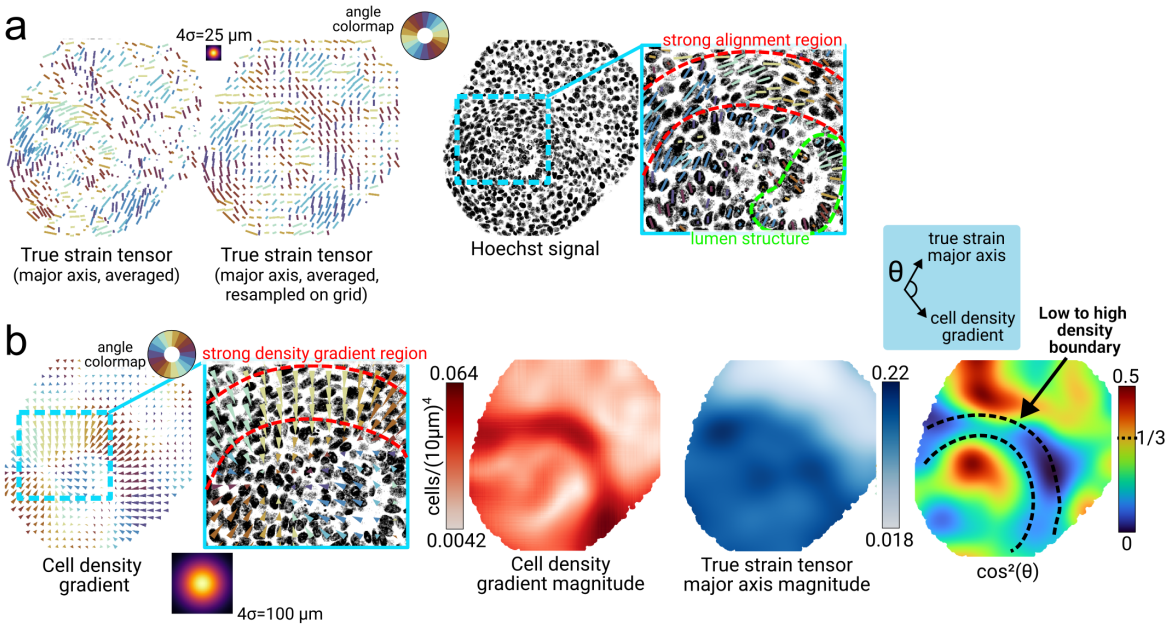

Figure 2.4: **Nuclei align along a stereotyped boundary between low- and high-density domains. (a)** Deformation fields at the cell scale, quantified by the major axis of nuclei true strain tensors, reveal spatially aligned regions overlapping with density boundaries and nuclei lining a lumen. A grid-sampled version is included for clarity. **(b)** Tissue-scale fields of squared cosine of the angle between the cell density gradient and strain tensor major axis highlight local anti-alignment. Vector magnitudes provide context in regions with weak gradients or deformation. All quantities are computed in 3D but shown as a single z-slice for convenience..

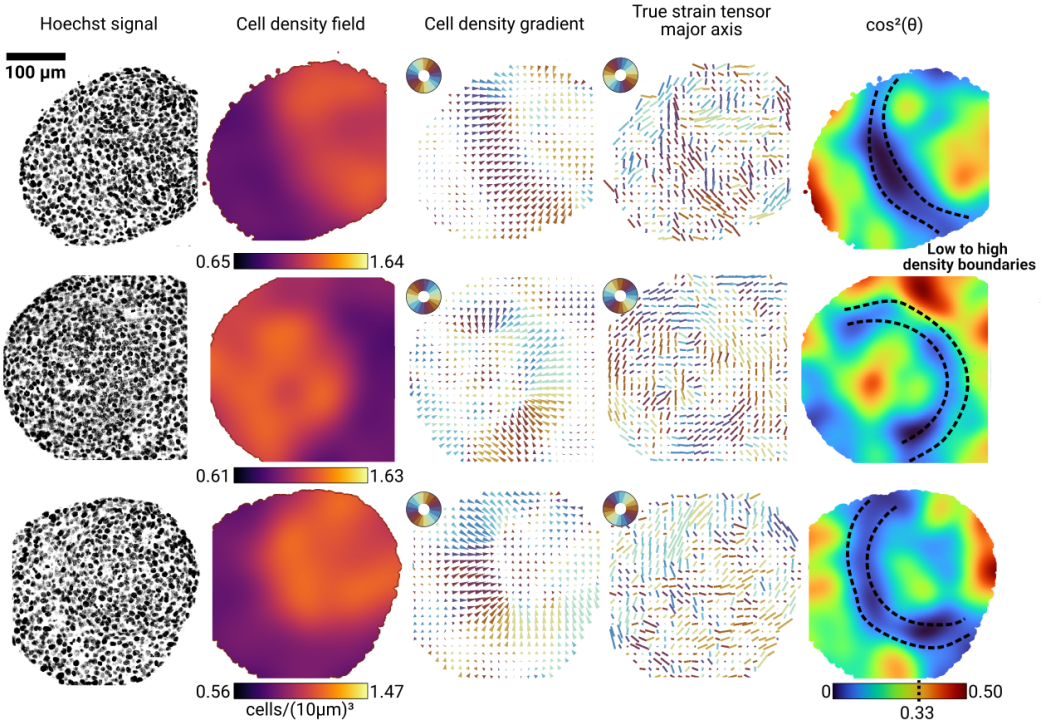

Figure 2.5: **Reproducibility of nuclei alignment along density boundary.** Consistent spatial colocalization of density boundaries and high alignment regions is illustrated across multiple gastruloids.

2.4.2 Gene expression can be quantified in deep gastruloids despite imaging artifacts by leveraging a model of fluorescence scattering

Intensity normalization retrieves homogeneous Hoechst spatial distribution despite intensity artifacts Quantifying gene expression in 3D tissue samples from imaging data is hindered by various sources of noise and optical artifacts that induce large scale intensity gradients (e.g. scattering, aberrations). We designed a method to remove these artifactual gradients. The method relies on one of the imaged channel being a ubiquitous signal, i.e a signal that is expected to be homogeneous within the sample provided that intensity artifacts were not present. In our case we mostly use the Hoechst signal, as it is often accessible in experiments. Assuming that intensity artifacts vary on a lengthscale superior to the lengthscale of variation due to genetic heterogeneities, the complete 3D spatial distribution of artifacts can be estimated by applying our masked convolution method, using a mask of the objects that express the ubiquitous signal. In the case of Hoechst, the nuclei segmentation can be binarized to a mask. The result of the convolution step is a coarse-grained map of nuclei intensity, that we use to locally normalize the original image.

Fig. 2.6a shows the average intensity of planes along the z-axis for the unnormalized and normalized Hoechst signal. THe normalization corrects for the exponential decay in intensity, and a flat profile compatible with the ubiquitous nature of Hoechst is retrieved. Fig. 2.6a demonstrates how the flatness of the profile along the z-axis depends on the length-

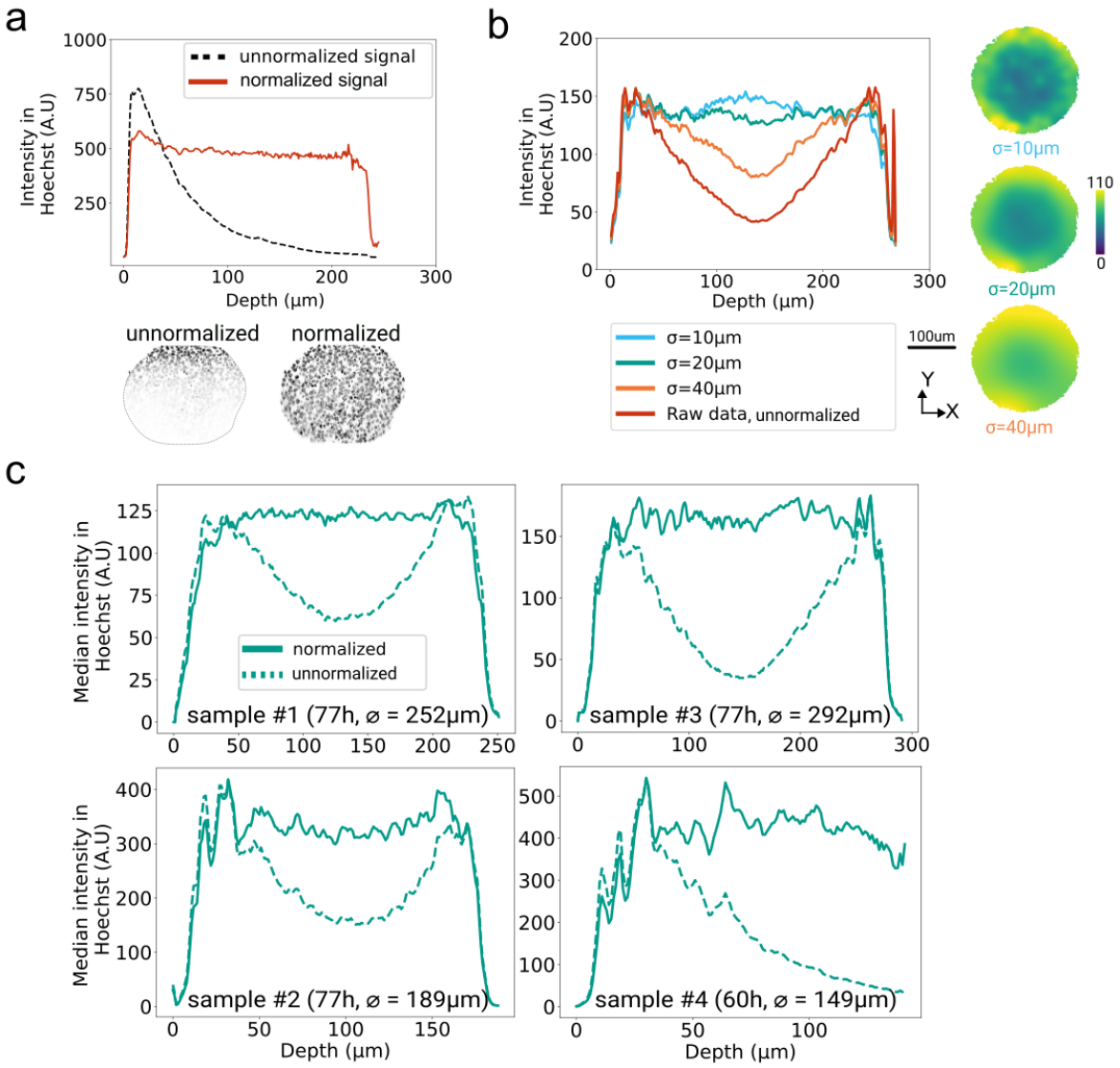

Figure 2.6: **Intensity normalization retrieves homogeneous Hoechst spatial distribution despite intensity artifacts (a)** Comparison between normalized and unnormalized Hechst signal. **(b)** Influence of the smoothing lengthscale σ on the intensity profile along the z-axis of the normalized Hoechst signal. **(c)** Robustness of the normalization method to retrieve homogeneous Hoechst intensity pattern in samples of varying timings and diameters.

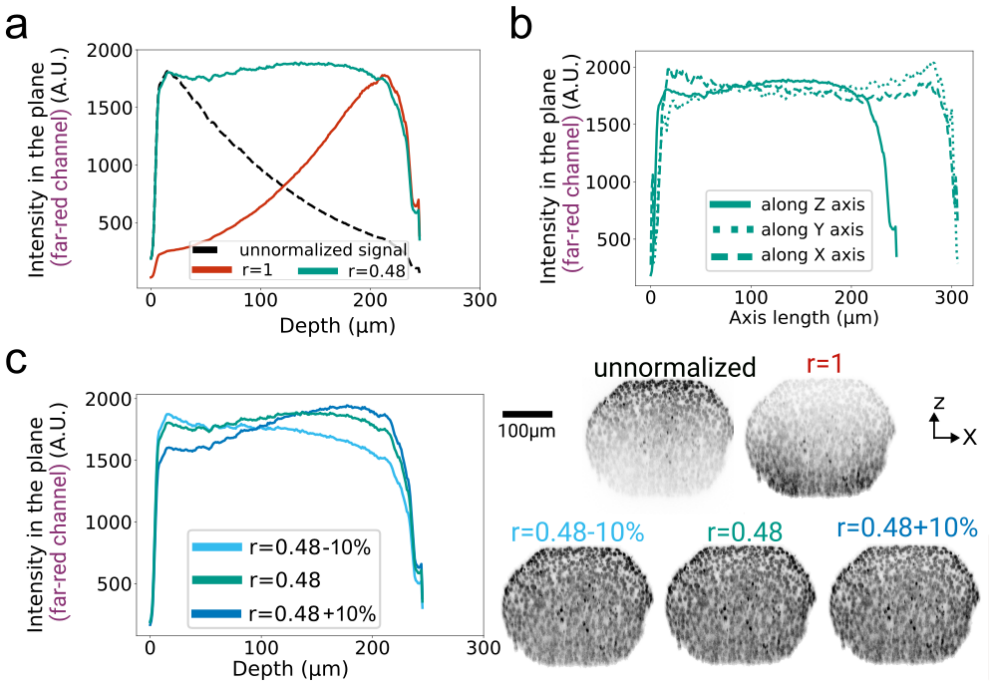

Figure 2.7: **Wavelength-dependence of intensity artifacts is mitigated using a physical model of scattering. (a)** Impact of the wavelength correction on the normalization of another ubiquitous signal. The experimentally calibrated decay length ratio between blue and far-red channels is used. The $r = 1$ profile ignores wavelength dependency. The corrected intensity profile can be appreciated on the XZ planes images. **(b)** Normalized profiles across Z, X, and Y axis show consistent flatness. **(c)** The method is robust to small perturbations ($\pm 10\%$) of r. Corresponding XZ planes are shown.

scale of smoothing used when computing the approximated field of intensity artifacts. In practice, we optimize for the profile that maximizes a robust heuristic flatness criterion (see equation (2.10)). We found empirically that the optimal σ, which gives insight on the typical lengthscale of variation of the intensity artifacts, is of the order of two nuclear diameters ($\sigma \approx 20\,\mu m$). Furthermore, the method effectively and robustly homogenized nuclei intensities across gastruloids from different acquisitions and imaging conditions (Fig. 2.6c).

Wavelength-dependence of intensity artifacts is mitigated using a physical model of scattering The intensity artifact profile depends on the imaging wavelength. This challenges our algorithm, as it is usually impossible to image both a ubiquitous and a biological signal of interest with the same wavelength. To solve this issue, we model the artifact profile as being exponentially decreasing. This predicts that two ubiquitous signals at different wavelengths can be related by defining a single positive exponent (up to a multiplicative factor), which is equal to the ratio of the decay lengths at the respective wavelengths. This ratio can be measured once using reference ubiquitous signal, and under mild assumptions it does not depend on experimental conditions.

We validate this approach by normalizing a ubiquitous in the "Far-red" acquisition channel using Hoechst ("Blue" acquisition channel). Neglecting the wavelength dependency pre-

vents our algorithm from retrieving the correct homogeneous intensity along the z-axis, but using our model with the experimentally determined ratio solves the issue (Fig. 2.7a). Interestingly, despite modeling only the exponential decay along the z-axis, the wavelength correction we propose also provides coherently flat profiles along the x- and y-axis (Fig. 2.7b).

Though we only experimentally determined the values of the decay length ratios at four different wavelengths, we believe our approach can be adapted to any pair of wavelengths by simply interpolating our data (see Fig. 2.2(c-d)). Our method is robust to small perturbations (e.g errors) in the choice of the decay lengths ratio. We show this on Fig. 2.7c, where the normalized profile remains valid up to a 15% error for perturbations of the r of 10% and shows little variation qualitatively.

Intensity normalization is required to measure spatial patterning of gene expression When we applied the correction method to the T-Bra signal (Fig. 2.8(a)) and evaluated normalized T-Bra intensity in a lateral section across the mid-plane of a 77 h gastruloid, we still observed a gradient with higher intensity in outer cell layers (Fig. 2.8b), while the nuclei signal recovered a flat profile. Besides reflecting a true T-Bra expression gradient, a possible explanation for this could be that the penetration of primary and secondary antibodies differs is considerably more difficult compared to Hoechst stain. To evaluate this, we compared immunostained T-Bra signal with the intensity of a T-Bra reporter (H2B-GFP expressed under a T-Bra promoter) in gastruloids grown from a transgenic cell line fixed at 72 h. We found a strong correlation between the two channels at all depths, excluding the hypothesis of penetration heterogeneities (see Appendix 6.5).

In addition to optical artifacts, our normalization method also corrects for intensity variations effectively induced by residual rotations of aggregates after flipping, which tilt the optical axis with respect to the axis of the first acquisition. This is illustrated in Fig. 2.8(b), providing an example where the nuclei signal exhibits an asymmetrical gradient due to the fusion of two opposite views that were subject to residual rotations. After normalization, the nuclei profile is homogeneous, whereas the profile in T-Bra drastically changed, showing a polarization in the opposite direction with respect to the initial gradient that was also visible both in the nuclei stain and T-Bra channel. This emphasizes the importance of our method to properly quantify gene expression gradients in 3D. The observed polarized T-Bra appearance is in agreement with previous studies in which it was reported that T-Bra expression already started to be polarized at 77 h in some gastruloids [36].

Measuring cell-scale spatial co-expression of genes brings biological insight on differentiation pathways Our ability to image multiple expression markers in the same acquisition without cross-talk and to reliably segment all cells allows studying spatial correlation in gene expression at the single cell level.
To demonstrate this, we applied the correction scheme to four-color 3D acquisitions of multiple gastruloids of different size (obtained by varying the initial cell number of the culture) imaged in the same slide, immunostained with T-Bra, the pluripotency marker Oct4, and the early endoderm marker FoxA2 (Fig.2.9(b)). We analyzed spatial correlations between T-Bra, Oct4, and FoxA2 after intensity normalization. Cell-scale spatial correlation between two genes is explored by plotting a 2D histogram in which each segmented nucleus is as-

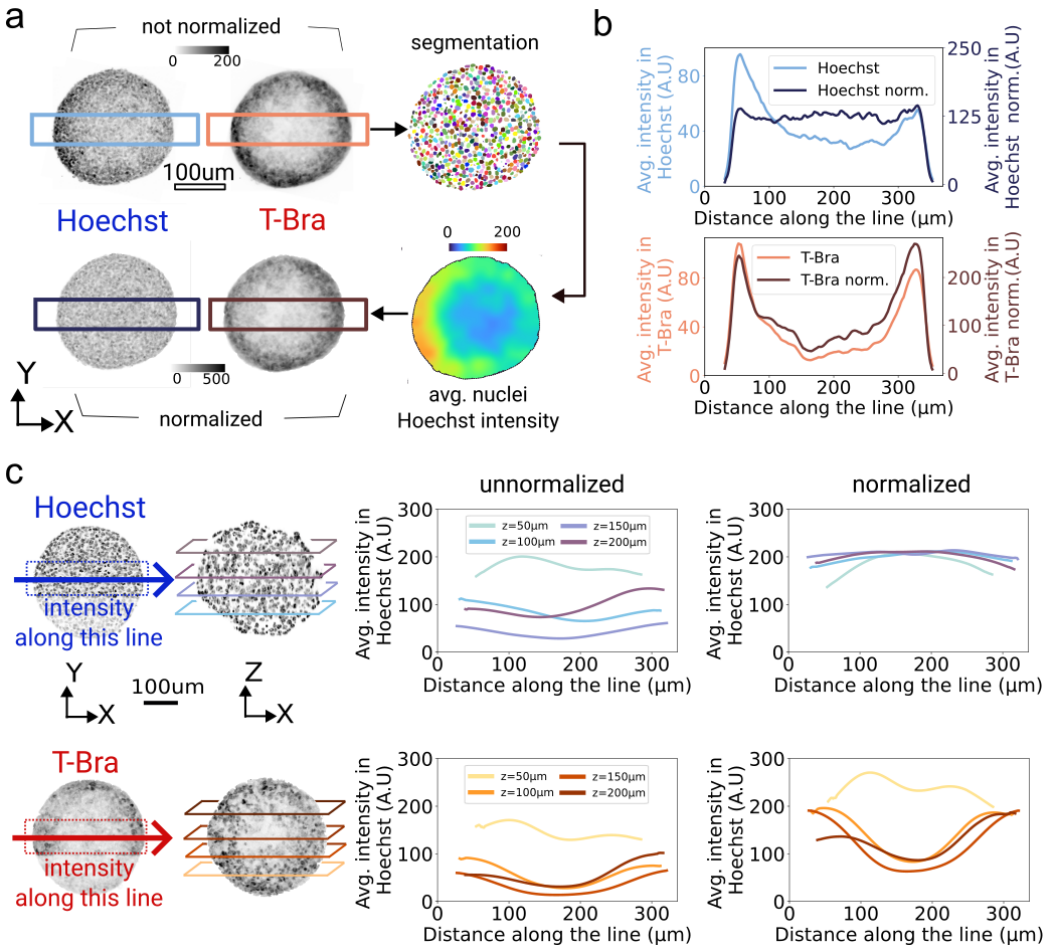

Figure 2.8: **Intensity normalization is required to measure spatial patterning of gene expression. (a)** The Hoechst signal is used to normalize the Brachyury (T-Bra) signal. **(b)** The method can also correct artifactual intensity gradients in the XY plane, e.g. due to additional rotations during the dual-view acquisition. **(c)** Comparison of lateral intensity profiles at different depths. We compute the intensity across rectangular sections of each image in 2D, as shown on the left. The graphs show the line profiles at different depths for the Hoechst channel (top) and the T-Bra channel (bottom), comparing profiles before (left plot) and after (right plot) normalization.

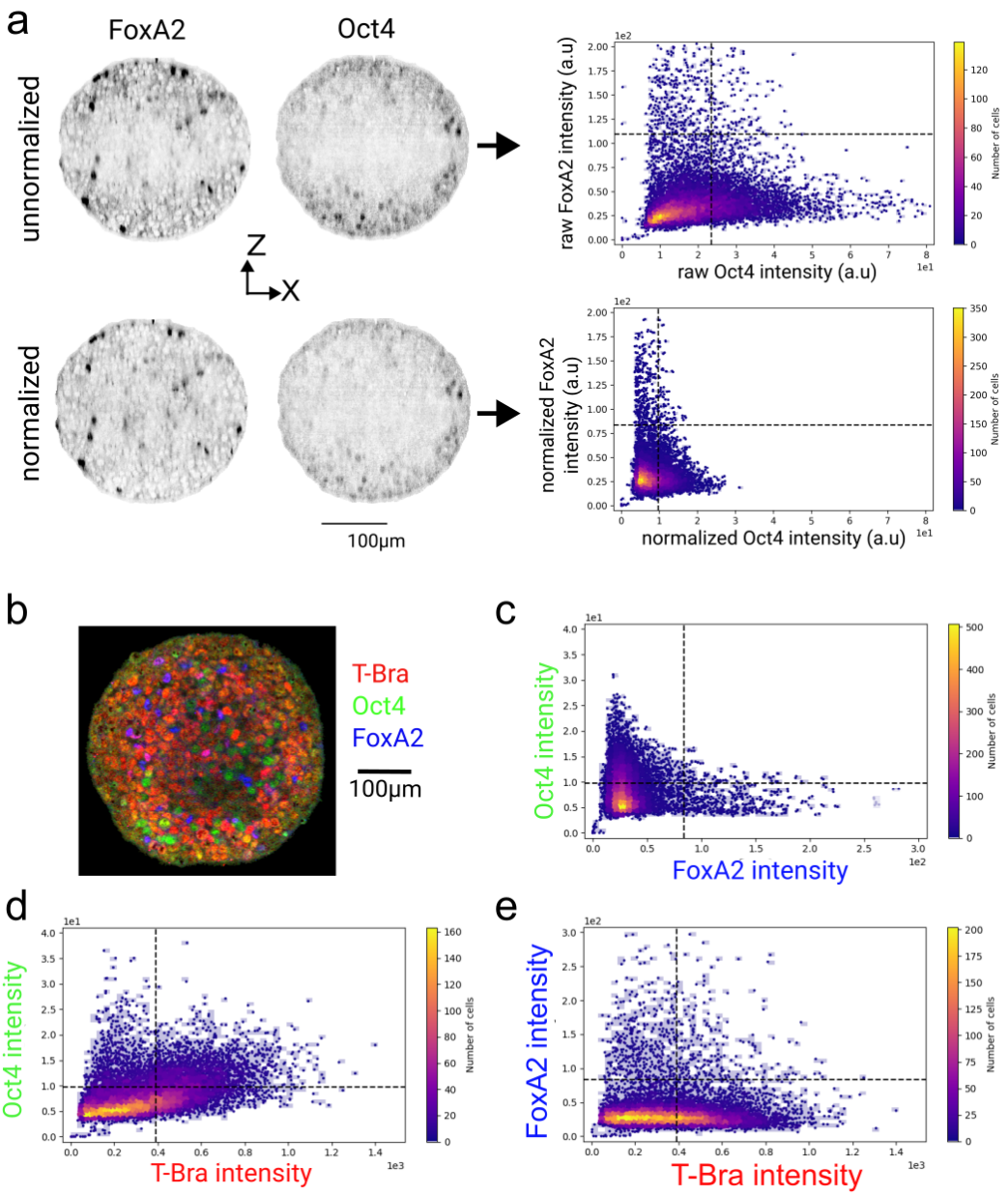

Figure 2.9: **Intensity normalization reveals spatial correlation features in key genes.** **(a)** Correlation maps between FoxA2 and T-Bra for non-normalized data (top) versus normalized data (bottom), and the corresponding images are shown on the left. **(b)** Three-color image of a 77 h gastruloid immunostained for T-Bra, Oct4, and FoxA2. The three plots show a cell-level correlation analysis of the three markers for the entire gastruloid. Each dot in the correlation plot represents an individual cell. Shown are all three marker combinations. Quadrants show regions of background signal (-) and actual expression (+) for each marker in the respective color, defined by Otsu thresholding.

sociated to a bin based on the average value of two genes within the voxels it encloses. Intensity normalization remained crucial to obtain unbiased results, as measuring faithful gene expression is only possible if the intensity is devoid of artifacts(Fig.2.9(a)). To discriminate marker-positive cells from cells only showing background levels, we applied Otsu thresholding on the cells' intensity distribution for each marker. This defines a threshold that separates the distribution between marker-negative and marker-positive populations, which was confirmed by visual inspection.

Based on previous studies [36], it is expected that FoxA2 and Oct4 are mutually exclusive since cells are either in the process of endoderm differentiation or still pluripotent. A similar behavior could be expected for the couples T-Bra/Oct4 and T-Bra/FoxA2.
We observed that Oct4 and FoxA2 appeared as two mutually exclusive populations, with almost no cells appearing in the top-right quadrant (Fig.2.9(c)). On the contrary, while the brightest Oct4 cells also showed low T-Bra signal and vice versa, a large pool of T-Bra and Oct4 positive cells appeared (Fig.2.9(d)), though we never observed a striking spatial pattern appearing in the data. This could suggest some co-expression since T-Bra marks very early mesendoderm differentiation *en route* to meso- or endoderm and Oct4 is a relatively long-lived pluripotency marker [120].
For FoxA2 and T-Bra, a strong lack of correlation was observed, with cells appearing significantly in all four quadrants (Fig.2.9(e)). Interestingly, we overall detected fewer "pure" FoxA2 positive cells than pure T-Bra positive cells. While we found many cells that were T-Bra positive and FoxA2 negative, most of the FoxA2 positive cells showed at least some T-Bra signal. This indicates that T-Bra expression might be required to activate FoxA2.

Size-dependent gene expression gradients reveal distinct expression patterns We sorted gastruloids based on their sizes and divided them into two groups: small, with an average diameter of $193 \pm 20\ \mu m$ (n=8), and big, with an average diameter of $291 \pm 15\ \mu m$ (n=5). We observed an approximately constant T-Bra expression for small aggregates, whereas big aggregates showed a clear decay of T-Bra expression towards the interior (Fig. 2.10(a)). Notably, we quantified T-Bra intensity as a function of distance to the gastruloid border using the *in toto* 3D image stack, thus reflecting expression gradients in 3D.

Because FoxA2 signal was sparse and noisy, we could not directly compute the map of expression as done previously for T-Bra. To properly analyze the spatial expression of FoxA2, we first thresholded the signal in segmented nuclei, using the interactive visualization tools developed in Napari (see Fig. 2.11) to select an appropriate intensity threshold and inspect the signal and distribution of FoxA2 positive cells. We analyzed the distribution quantitatively by generating a coarse-grained map of FoxA2 positive cell density fraction. This was done by computing the number of FoxA2 positive cells rescaled by the local nuclei density, using the same method as applied for proliferation analysis, except that we consider FoxA2 positive cells instead of ph3 positive cells. The coarse-grained map of FoxA2 positive cells was then analyzed radially for all gastruloids.

For FoxA2, a similar pattern was observed, but expression appeared in much fewer cells (Fig. 2.10(b)). We observed slightly higher FoxA2 levels in outermost cell layers, and in general more cells expressing FoxA2 in small gastruloids proportionally to the total number of cells.

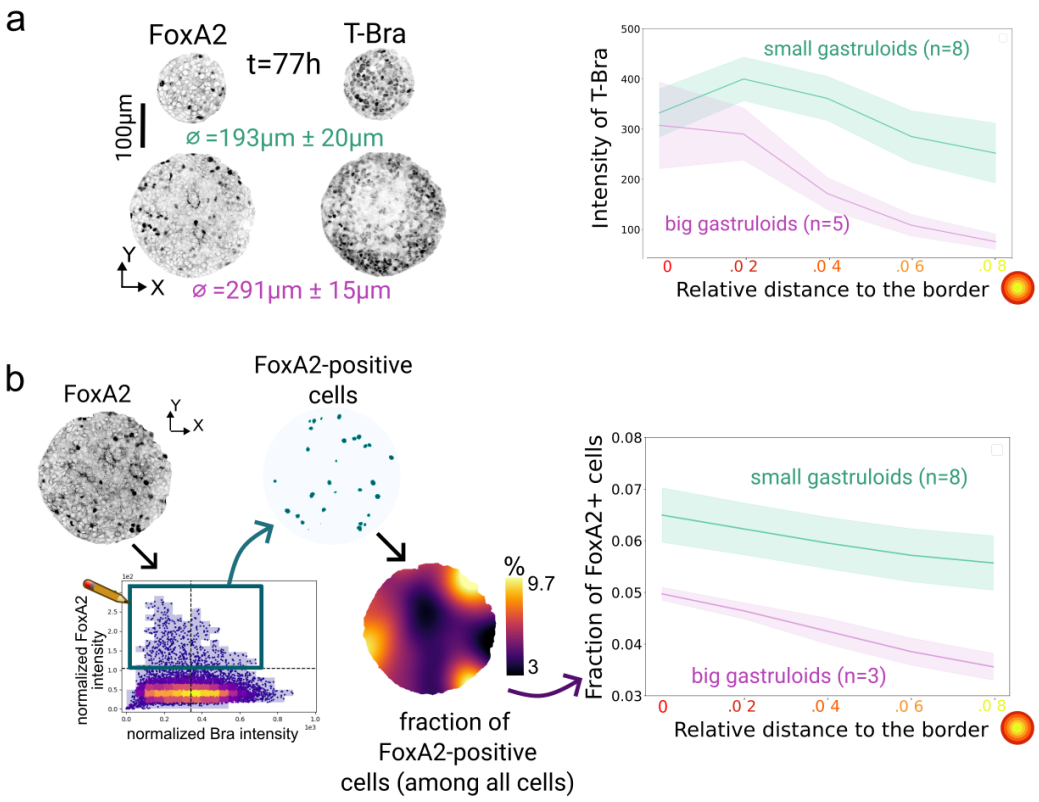

Figure 2.10: **Gastruloid exhibit radial expression patterns for T-Bra and FoxA2. (a)** Images of 77 h gastruloids (mid-plane) of different size, immunostained for T-Bra and FoxA2 gene expression markers. From the normalized 3D images, radial intensity plots were generated for big ($291 \pm 15\ \mu m$ average diameter, n=5) and small ($193 \pm 20\ \mu m$ average diameter, n=8) gastruloids as a function of distance to the border. Shaded regions show standard deviations. **(b)** For one of the previously described samples fixed at 77 h, FoxA2 positive cells are visualized from the normalized data using the interactive Napari tools and the image thresholded accordingly. Using the binary image of FoxA2 positive cells, a map of local fraction of FoxA2 positive cells, rescaled by the local cell density, was computed, showing a radial pattern. Analysis of the radial distribution, i.e. fraction of FoxA2 positive cells as a function of distance to the border, for big ($278 \pm 16\ \mu m$ average diameter, n=3) and small ($193 \pm 20\ \mu m$ average diameter, n=8) aggregates, computed from the coarse-grained maps as shown in (d). Shaded regions show standard deviations.

2.5 The *Tapenade* software suite brings interactive and user-friendly tools to empower researchers

2.5.1 The Python package

Driven by the recent success of software solutions for 2D analysis of organoid at the tissue scale [121] and for extensive 3D quantifications of cell properties in epithelia [122], we packaged our pre-processing and analysis pipeline into a *Python* library called *Tapenade* (for "Thorough Analysis PipEliNe for Advanced DEep imaging"). The code is freely available on *Github* at github.com/GuignardLab/tapenade/. The library comes with an extensive documentation and *Jupyter* notebooks accessible to non-specialist users. All datasets used to train our custom *StarDist* models to detect nuclei and mitotic events, along with the optimized weights for nuclei and mitotic events detection are currently available upon reasonable request. A non-specialist user can quickly segment a custom dataset by downloading the weights and loading them into one of the user-friendly implementations of *Stardist* like the *Stardist Fiji* plugin, the *Napari* plugin *stardist-napari*, or the *ZeroCostDL4Mic* notebook[91, 123].

2.5.2 The Napari plugins

While working with large and dense 3D and 3D+time gastruloid datasets, we found that being able to visualise and interact with the data dynamically greatly helped processing it. During the pre-processing stage, dynamical exploration and interaction led to faster tuning of the parameters by allowing direct visual feedback, and gave key biophysical insight during the analysis stage. We thus created four user-friendly Napari plugins designed around facilitating such interactions (Fig.2.11):

napari-file2folder To facilitate accessibility and usability of our code for non-expert users, including those working on standard laptops, we designed the data handling pipeline to accommodate large image datasets that may not fit entirely in memory. We employ the *bioio Python* library to read diverse bioimage file formats, such as `tif`, `ome.tiff`, `czi`, `lsm`, and `nd2`. To manage memory efficiently, we use *zarr* for lazy loading, allowing each frame, or more generally each element along any dimension, to be accessed and processed individually without requiring the entire dataset to be loaded into memory. This approach also allows us to save these frames independently to disk, further reducing memory overhead and simplifying downstream processing workflows with other plugins (Fig.2.11a). .

In bioimage analysis, the lack of a standardized convention for dimension ordering in Python-based workflows can lead to ambiguity. For consistency and compatibility with *Napari*, we adopt the dimensional order *CTZYX* (channel, time, z-axis, y-axis, x-axis). This convention streamlines interoperability between different parts of the processing pipeline and simplifies visualization and manipulation within the *Napari* software. The plugin is

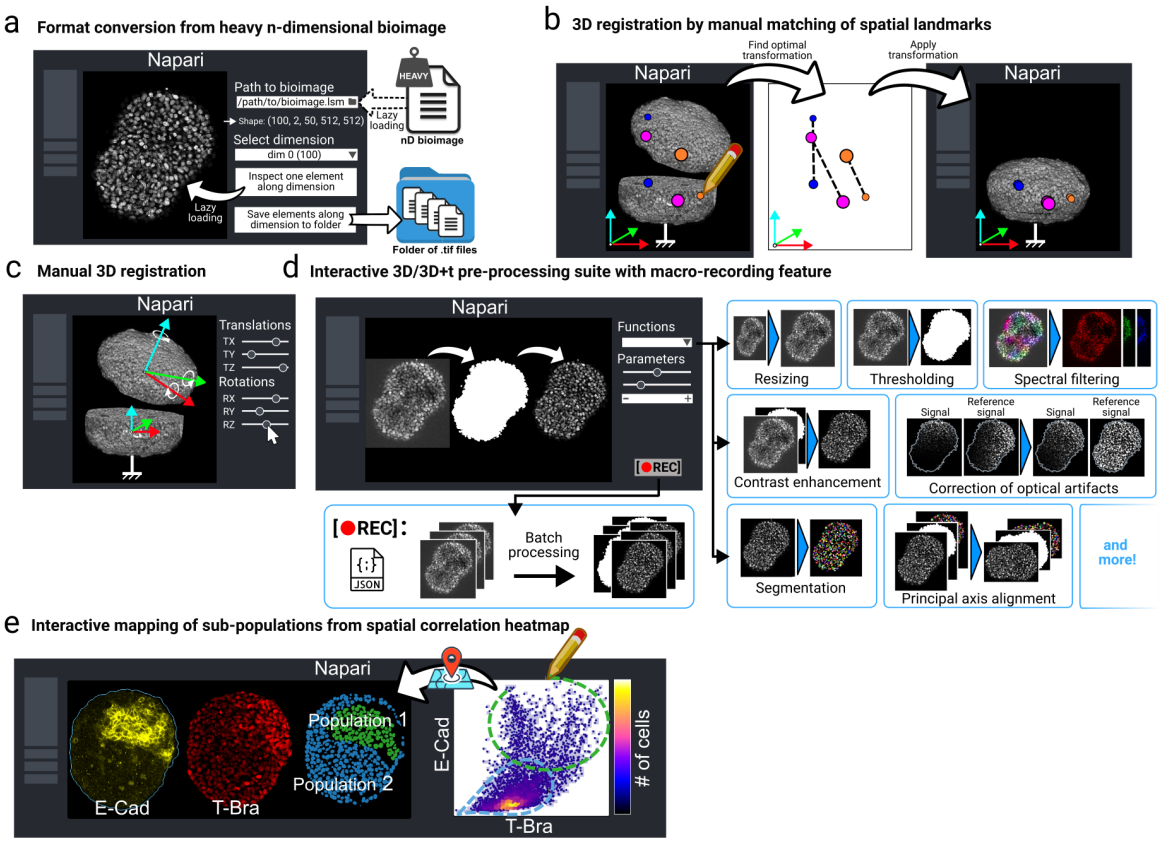

Figure 2.11: **Interactive Napari plugins for bioimage handling, manual registration, pre-processing, data exploration and analysis. (a)** A Napari plugin for exploring and processing large nD bioimages. Users can load datasets lazily, inspect slices along a chosen dimension, and export them as separate TIFF files. Compatible with common bioimage formats, it simplifies handling complex datasets. **(b,c)** To fuse data acquired with our dual-view setup, we provide a Napari plugin to assist users in manually defining a 3D rigid transformation either when it is trivial (e.g a single axial translation) or to initialize an automatic registration tool close to the optimum. The first mode (left) allows the user to select salient landmarks and match them between the two views. An optimal rigid transformation is found automatically. The plugin's second mode (right) lets the user define each element of the transformation (translations and rotations) until a satisfactory match is observed in 3D or on 2D planes. **(d)** A complete pre-processing suite is provided as a Napari plugin to transform raw 3D and 3D+time datasets into datasets ready for analysis. The Napari implementation allows for quick and visual exploration of parameters best suited to a given dataset. A recording feature can be toggled to save a complete user-defined pipeline into a *JSON* file, that can be used to process large datasets in batch or for sharing. **(e)** Interactive analysis of multiscale correlation heatmaps in Napari is provided by allowing users to change the analysis length scale and to draw regions of interest directly on the correlation plot to see which cells contributed to the region's statistics on a 3D view of the data.

freely available on the *Napari* hub or at github.com/GuignardLab/napari-file2folder.

napari-manual-registration When using our automatic registration tool to spatially register two views of the same organoid, we were sometimes faced with the issue that the tool would not converge to the true registration transformation. This happens when the initial position and orientation of the floating view are too far from their target values. We thus designed a Napari plugin to quickly (typically less than a minute for a confirmed user) find a transformation that can be used to initialize our registration tool close to the optimal transformation. From two images loaded in Napari representing two views of the same organoid, the plugin allows the user to either (i) annotate matching salient landmarks (e.g bright dead cells or lumen-like structures) in both the reference and floating views, from which an optimal rigid transformation can be found automatically using singular value decomposition [124] (Fig.2.11b), or (ii) manually define a rigid transformation by continually varying 3D rotations and translations while observing the results until a satisfying fit is found (Fig.2.11c). The plugin is freely available on the *Napari* hub or at github.com/GuignardLab/napari-manual-registration.

napari-tapenade-processing From a given set of raw images, segmented object instances, and object mask, the plugin allows the user to quickly run all pre-processing functions from our main pipeline with custom parameters while being able to see and interact with the result of each step (Fig.2.11d). We provide typical computation times on Table **??**. For large datasets that are cumbersome to manipulate or cannot be loaded in Napari, the plugin provides a macro recording feature: the users can experiment and design their own pipeline on a smaller subset of the dataset, then run it on the full dataset without having to load it in Napari. The plugin is freely available on the *Napari* hub or at github.com/GuignardLab/napari-tapenade-processing.

napari-spatial-correlation-plotter This plugins allows the user to analyse the spatial correlations of two 3D fields loaded in Napari (e.g. two fluorescent markers). The user can dynamically vary the analysis length scale, which corresponds to the standard deviation of the Gaussian kernel used for smoothing the 3D fields. If a layer of segmented nuclei instances is additionally specified, the histogram is constructed by binning values at the nuclei level (each point corresponds to an individual nucleus). Otherwise, individual voxel values are used. We took inspiration from the plugin *napari-clusters-plotter* [125] in letting the user dynamically interact with the correlation heatmap by manually selecting a region in the plot. The corresponding cells (or voxels) that contributed to the region's statistics will be displayed in 3D on an independant Napari layer for the user to interact with and gain biological insight (Fig.2.11e). The plugin is freely available on the *Napari* hub or at github.com/GuignardLab/napari-spatial-correlation-plotter.

Preprocessing function	Laptop	Workstation	Most limiting steps
Isotropization	10 s	7 s	Pixel-value interpolation
Masking (in/out)	30 s	20 s	Gaussian smoothing
Local contrast enhancement	9 s	7 s	Pixel-value interpolation
Intensity normalization	40 s	30 s	Gaussian smoothing
Segmentation	190 s	25 s	• Inference (if no GPU) • Reconstruction procedure
Axis alignment	7 s	5 s	Pixel-value interpolation
Segmentation postprocessing	4 s	4 s	N/A
Cropping	1 s	1 s	N/A
Total	∼290 s	∼100 s	

Table 2.1: **Preprocessing steps with corresponding typical computation times** for a standard 3D volume of shape 300×512×512 with ∼6000 nuclei, reported for a standard scientific laptop with an 11th Gen Intel Core i7-1185G7 CPU, and a a computer with an Intel Xeon Gold 6242R CPU and a Nvidia RTX A6000 GPU.

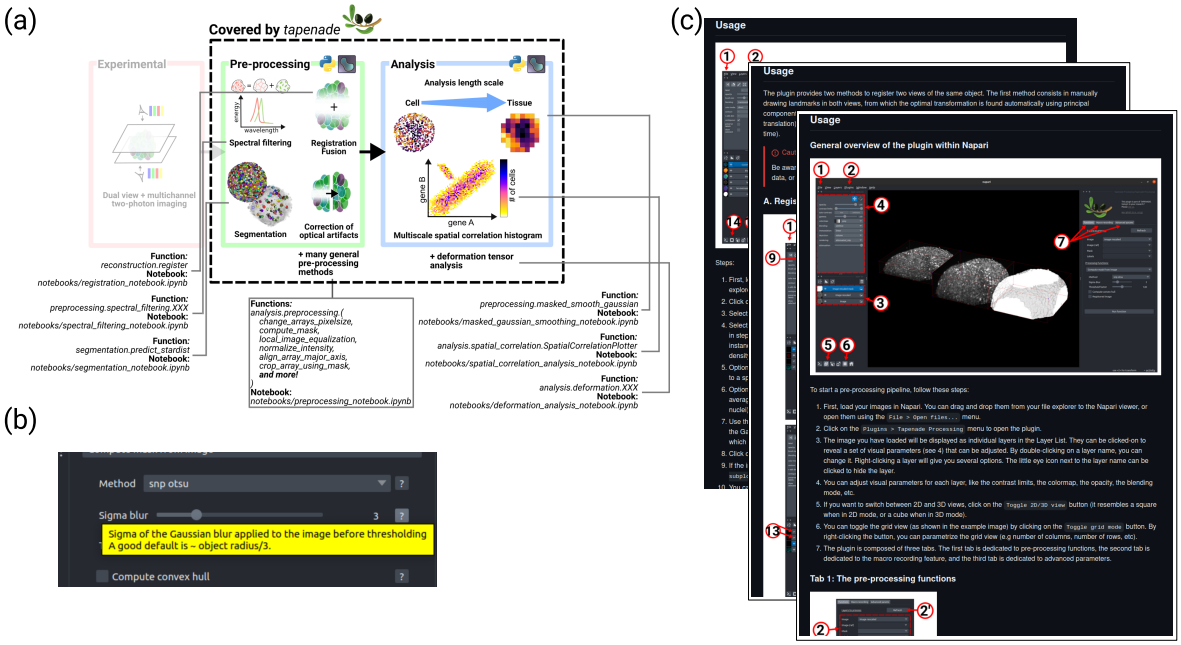

Figure 2.12: **The Tapenade software suite was designed with user-friendliness in mind.** **(a)** Tapenade packages *Python* tools that are designed for accessibility to non-expert users (through detailed Jupyter notebooks and documentation), and modularity for computer-savy users. **(b)** In our Napari plugins, most parameters come with a "Help" button which provides information about how to use it best. **(c)** All plugins come with extensive documentation, videos, and installation instructions.

2.5.3 User-friendliness of the tools

We believe our pipeline offers a distinctive and effective integration of 3D data exploration and processing, providing a valuable tool for the organoid and tumoroids research community. All tools are designed with a focus on modularity for expert users, accessibility to non expert users (Fig. 2.12(a)), and general user-friendliness through instinctive graphical interfaces (Fig. 2.12(b)) and detailed documentations (Fig. 2.12(c)).

2.6 Discussion

2.6.1 Multiscale imaging and analysis pipeline

We have developed an integrated pipeline for *in toto* imaging of whole-mount gastruloids and automated image processing, illustrated on Fig. 2.13. A key innovation of our approach is its specific optimization for dense 3D tissues, such as gastruloids, which present similar imaging challenges and artifacts as tumors and spheroids, unlike many organoids that are shallow and less light-diffusive. Unlocking deep whole-mount imaging is crucial for researchers working with such dense tissues, as cryosectioning strategies cannot fully substitute for comprehensive 3D *in toto* analysis [39, 126]. Reconstructing 3D tissue mi-

crostructure from cryosectioned data is complex and challenging, involving slicing tissue into thin sections and imaging them individually. To achieve a 3D reconstruction, these 2D images must be meticulously aligned and stitched together, a process complicated by frequent issues such as tissue deformation and section loss. In contrast, whole-mount imaging preserves spatial relationships and tissue integrity, making it the method of choice for studying cellular interactions and tissue architecture in their native context. Additionally, whole-mount imaging reduces sampling bias and enhances the detection of rare cell populations or phenotypic variations that might be missed in thin tissue sections. This limitation also apply to state-of-the art spatial transcriptomic techniques, due to the fact that only few RNAs are typically captured per cell for each individual gene and many cells do not show any counts for a gene of interest expressed at low level [127].

Here, we developed a wavelength-dependent normalization scheme based on a simplified columnar model. Future work could focus on creating a complete model that accounts for photon paths in a complex y, for example in elongated gastruloids, and integrate the spectral filtering approach to the wavelength-dependent normalization scheme. We assume that the nuclear signal is the most reliable for normalizing other signals due to its extensive coverage of the tissue (over 60% in gastruloids). Instead of normalizing signals at the level of individual nuclei, we developed a correction scheme that identifies the optimal scale of characteristic gradients in nuclear intensities to generate a continuous field for renormalization. This approach has the added advantage of being applicable to markers with different subcellular localizations (e.g., nuclei, cell membrane, intracellular organelles).

2.6.2 Exploring gastruloid development in 3D

Our methodology utilizes segmentation of the Hoechst-stained nuclei channel to (1) identify individual cells for cell-scale analysis and (2) normalize other channels using the nuclei signal. By developing a diverse training set for nuclei segmentation, we achieved precise determination of nuclei shapes. This precision enables systematic studies of the relationship between nuclear morphology and gene expression patterns, and can be extended to cell morphology using membrane stains such as *ZO-1* and cell shape segmentation methods like CellPose [95] or more specialized approaches for challenging membranes [128]. Nuclei deformation patterns show large-scale alignment at the gastruloid level (see Fig. 2.4 and Fig. 2.5) and localized alignment near lumens and cavities, providing potential insights into the structuring of epithelial and mesenchymal tissues during gastruloid development. In older gastruloids with cavity-containing structures, such as somitic structures [35], automated cavity detection will be crucial for accurately quantifying coarse-grained fields of cell or division densities.

In this work we have not used refined masks of the gastruloids that would account for and exclude hollow structures like lumen. In Fig. 2.3 and 2.4(a), we noted that this lead to noticeable patterns in the packing-related coarse-grained fields and in the local alignment of the true strain tensor major axis. Furthermore, obtaining precise masks is essential to extend our analysis to inherently hollow datasets, like intestinal gastruloids. To improve our current implementation, our masking approach could be augmented by a second method used to specifically detect these structures. Pixel classification based on Random-Forests

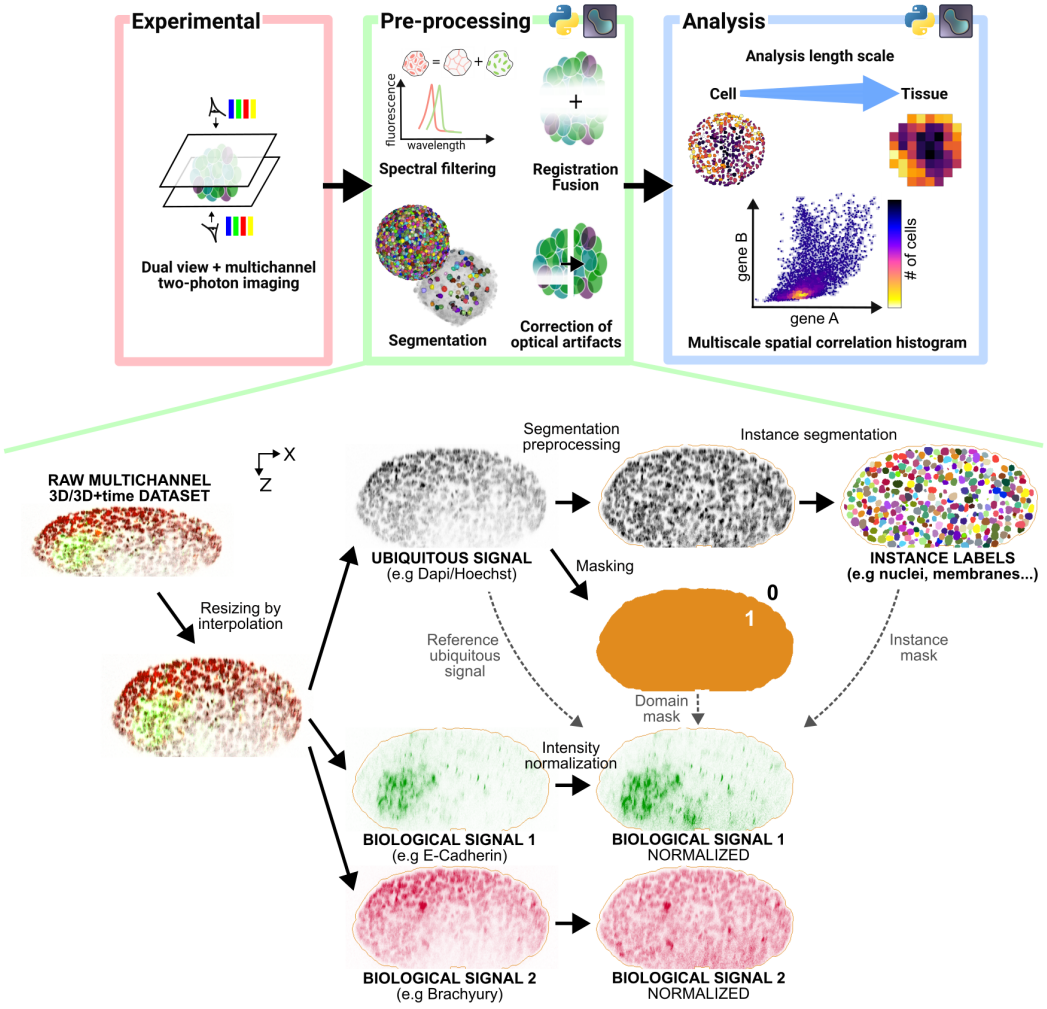

Figure 2.13: **Overview of the experimental and computational pipeline for *in toto* imaging of organoids at cellular resolution and multiscale analyses.** A typical workflow from the preprocessing block is illustrated on the bottom half using a slice of a multichannel gastruloid dataset in the YZ plane to show how our tools handle challenges related to depth.

[129] could be used to get accurate segmentation using only a few manually annotated lines. These approaches are now widely accessible to non-specialist users in software like *Fiji* (with *Labkit* [130]), *Ilastik* [131], or *Napari* (with *napari-APOC* [132]).

In the context of gastruloid development, it is crucial to understand how spatial heterogeneities in gene expression arise at both the cell microenvironment scale and the larger gastruloid scale. During early gastruloid symmetry breaking and elongation, radial and later antero-posterior gradients of the gene T-Bra, a nascent mesoderm marker, emerge. These radial gradients can be captured with our coarse-grained analysis (see Fig. 2.8 and Fig. 2.10). However, while both the endoderm marker FoxA2 and T-Bra exhibit radial gradients at the coarse-grained scale, their anti-correlated expression is only evident at the single-cell level (see Fig. 2.9). This example highlights the importance of a multiscale analysis pipeline that allows simultaneous visualization of tissue-scale properties and local cellular events. Overall, the spatial-correlation studies we performed might suggest a sequence of cellular transitions from Oct4 progressively to T-Bra, from which a fraction of cells, presumably the ones expressing low T-Bra levels, will commit to FoxA2. To further explore such transitions and localize specific cell populations, tools that allow interactive mapping of cell populations selected in the correlation plot to the actual gastruloid are needed. We have, therefore, integrated this functionality into our Napari plugin (see Fig2.11(e)). Overall, the 2D correlation histograms are practical to quickly and interactively explore striking et and salient co-expression (or anti-expression) patterns, but they lack proper statistical background and physical gene expression modeling to give definitive answers for most situations in which ambiguous correlations arise.

In the future, we will extend the analysis to dynamic features by integrating motion estimation at both the cell scale and a coarser level (e.g with single-cell tracking and optical flow) in our pipeline. This will allow us to explore the connection between gene expression and variations in cell motility. For instance, we could study spatial patterning of strain rate or effective viscosity coarse-grained at different scales in both space and time [133]. Another prospect is automated mining strategies to identify coordinated patterns in cell populations dynamics [134].
Recent studies also suggest directly learning spatial fields of mechanical quantities from raw fluorescence signals down to the cellular scale using deep-learning [135]. These methods bypass the challenges of object (cells or nuclei) segmentation entirely and could be trained using results from our pipeline. Ultimately, they could lead to improved accuracy and robustness in physical quantification, and could even benefit from physical model that can be directly injected in the deep learning framework [136].

2.6.3 A Python integrated pipeline for non-specialists, with Napari-enabled 3D exploration

We built our pipeline in *Python* to leverage its rich ecosystem of scientific computing libraries [137] while keeping the code as accessible and user-friendly as possible. We provide detailed installation instructions and *Jupyter* notebooks with step-by-step instructions designed to be used by non-specialists, with a focus on ease of use and reproducibility. To tackle the challenges that arise from manipulating deep and dense 3D data, we provide *Na-*

pari plugins for manual registration, preprocessing, and analysis. *Napari* provides a user-friendly graphical interface which enables real-time visualization and exploration of 3D data during each step of the preprocessing and analysis. This was essential for the development of the pipeline, as the direct visual feedback allowed us to quickly iterate and optimize the processing steps, while gaining insights into the 3D organization of gastruloids. We believe our pipeline offers a distinctive and effective integration of 3D data exploration and processing, providing a valuable tool for the organoid and tumoroids research community.

The bioimage field is saturated with software, toolboxes, and plugins, making it challenging to find tailored solutions. Special guides [138] and community forums like *ImageSc* help users navigate available tools. While our pipeline fills a clear gap in deep 3D tissue analysis, its long-term diffusion and maintenance remain uncertain, as dedicated software engineering positions for niche tools are often inaccessible to research teams. Community efforts address this by consolidating methodologies under unified platforms with dedicated engineering support, such as TrackMate [139] for tracking and ZeroCostDL4Mic [123] or Biapy [140] for easily accessible deep learning solutions for e.g. segmentation, denoising, or super-resolution. In the future, our methodology could integrate such platforms, as CartoCell [122] has within Biapy's ecosystem.

Part II

Live datasets

Chapter 3

Cell tracking strategies in developmental biology: challenges associated to stochastic dynamics and uncertain measurements

3.1 Chapter overview

This chapter addresses the complexities of cell tracking in developmental biology, emphasizing the difficulties posed by stochastic cell dynamics, uncertain measurements, and inherent technical limitations. We systematically review existing tracking strategies, critically assessing their applicability to deep gastruloid tissues. Furthermore, we explore the crucial role visualization tools play in parameter tuning and method validation, setting the stage for the Bayesian tracking solutions introduced in the following chapters.

My contributions:

1. reviewed cell tracking algorithms and visualization software for large 3D+time datasets, with an emphasis on how they relate to the challenges of developmental biology
2. designed and implemented a visualization method for Napari, which allows to display highly customizable and interactive orthorendering of 3D data

3.2 Challenges in cell tracking strategies for developmental biology

Visual inspection alone is insufficient to study 3D biological movies, as humans are incapable of following even a few cells for extended periods of times using 2D slices of the data. Automated cell tracking is a powerful technique that aims to follow cells across multiple generations, providing insights into temporal dynamics, lineage trees, spatial organization, and cellular movements.

3.2.1 Aleatoric and epistemic uncertainties in gastruloid cell observation

Two aspects complicate cell tracking in gastruloids. First, cellular stochasticity is a fundamental aspect of biological systems, playing a crucial role in development, differentiation, and tissue homeostasis [22, 29, 141]. This inherent variability manifests in various cellular processes, including gene expression, cell movement, and morphology, and hinder predictive modeling of their dynamics.
Second, experimental constraints prevent freely exploring the space of experimental conditions. For instance, photon exposure has to be kept low enough during long acquisition as to prevent phototoxicity and fluorescence bleaching, which means that high framerate (i.e continuous motion and shape dynamics), and high illumination/resolution (i.e better constrast and object delineation) are inaccessible.

This fundamentally results in

- Aleatoric uncertainty: as cells' natural stochasticity prevents deterministic predictions of their motion, morphological, or genetic dynamics.
- Epistemic uncertainty: as (i) low SNR and spatio-temporal resolution creates regions in the movies where even human experts cannot visually resolve the different objects and their dynamics, and (ii) our custom StarDist3D model trained in Chapter 1 produces detection errors in the form of split/merged nuclei or false negatives which cannot reasonably be manually corrected. This introduces a form of uncertainty in the detection and shape reconstruction processes.

3.2.2 Features required by an effective tracking strategy for cell lineage reconstruction in gastruloids

In this context, an effective 3D nuclei tracking strategy must address several key challenges:

1. Handling faulty detection results due to highly deformable (i.e not ellipsoidal) objects, which vary significantly between frames, low SNR and extreme nuclei packing.
2. Modeling stochastic motion and morphological dynamics, accommodating mid-track transitions between motion patterns, multiple cell populations with distinct movement patterns, and sudden shape change induced by compressed nuclei or mitotic events.
3. Leveraging robust object association techniques to manage lineage constraints such as the possibilities of cell division (while preventing cell merge), cell death, and cells entering or leaving the field of view, particularly in *non-toto* imaging.
4. Implementing data-efficient algorithms, as scalability is essential since gastruloid datasets can involve tracking tens of thousands of cells across hundreds of frames.

3.3 Multiple particle tracking without epistemic uncertainty: case-study on canonical Brownian dynamics

In this chapter, we aim to simplify the problem of cell trajectory reconstruction in organoids by systematically reducing its complexity to a solvable form before gradually reintroducing biological and experimental challenges. We begin by considering the fundamental case of tracking N point-like particles undergoing idealized Brownian motion, following the seminal approach of Reid et al. [142], adapted e.g for tracking brownian colloids in a study by Crocker and Grier [143]. Even in this simplified setting, we encounter significant difficulties that highlight the inherent challenges of particle tracking. By the end of this illustrative case, we extend our analysis to include organoid-specific complexities, demonstrating how they compound the difficulties already present in the textbook example. This motivates the need for a methodology which directly models aspects of cellular stochasticity, which we develop further in subsequent sections.

3.3.1 Maximum-likelihood solution for multiple Brownian particles tracking without measurement uncertainty

To reconstitute the motion of N particles undergoing Brownian dynamics from a set of images, we alleviate the challenges that come with detecting objects in images and assume that the first detection step has been done already and yielded perfect results. We thus leverage the time-tested approach of "detect-then-track" [73, 75, 139, 144–146], which consists in treating the detection and association tasks independently. This paradigm is traditionally used when temporal context is not necessary to obtain perfect detection results.
This approach is consistent with the pioneering work of Bao et al. [144] who performed automated lineage tracking in *Caenorhabditis elegans* embryos imaged with a confocal microscope. Due to the small size (relative to the size of the cells) and the sparsity of nuclei in their images, they obtained a quasi-perfect detection result using a simple local-maximum algorithm. However, access to perfect detections/segmentations is usually not a marginal assumption in developmental biology studies, as detecting cells or nuclei in embryos is a challenging task. Most methods adopt a different paradigm that will be further addressed later.

The simplest toy model for stochastic dynamics is the Langevin equation, which was originally introduced to describe Brownian motion, which refers to the seemingly erratic motion of a particle immersed in a fluid, caused by collisions with the surrounding fluid molecules. The equation was also used in several biophysical studies to describe biological phenomena at diverse scales. At the molecular level, it has been applied to describe the motion of Brownian particles, providing insights into the behavior of biomolecules in solution [147, 148]. At larger scales, the Langevin formalism has been adapted to model the dynamics of entire biochemical systems, enabling researchers to describe to characterize cell migration phenomena for cancer cells and human granulocytes [149].

The Langevin equation describes a particle subject to viscous drag from its surrounding environment and to a stochastic force term which accounts for random fluctuations. The Newton's second law in the overdamped limit reads

$$\mathbf{0} = \mathbf{F}_{\text{drag}} + \mathbf{F}_{\text{brown.}} = -\gamma \mathbf{v} + \boldsymbol{\eta}(t) \,. \tag{3.1}$$

In this equation, $\mathbf{v}$ represents the particle's velocity, γ is the damping coefficient, and m denotes the mass of the particle. The stochastic force $\boldsymbol{\eta}(t)$ is modeled by a Gaussian probability distribution, with instantaneous expected value $\mathbb{E}[\boldsymbol{\eta}(t)]$ represented by a drift term $\mu(t)$, and component-wise auto-covariance function given by

$$\mathbb{E}\left[(\eta_k(t) - \mu_k(t))(\eta_j(t') - \mu_j(t'))\right] = \sigma^2 \delta_{k,j} \delta(t - t') \,, \tag{3.2}$$

where η_k and μ_k represent respectively the k-th component of $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$, σ is the standard deviation of a component's amplitude, and where $\delta_{k,j}$ and δ are respectively the Kronecker δ and the Dirac δ-function. The time-correlation expressed as a δ-function indicates that the random force at any given moment t has no correlation with the force at any other time, and that two different components of the force are not correlated.

Such a first order differential equation is usually solved by specifying an initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$. However, due to its stochastic nature, it is not possible to directly solve for the position of the particle. Instead, we can model the time evolution of the probability distribution function (PDF) conditioned on the initial condition $p(\mathbf{x}, t|\mathbf{x}_0, t_0)$, which can be interpreted as the particle's likelihood to belong in a certain spatial neighborhood at a given time.

To simplify this case study to the maximum, we make four assumptions to proceed with the demonstration

- Though cells in dense tissues are by definition constantly interacting to maintain the integrity of the tissue, we make the assumption that the motion of the particles between two consecutive frames is not dominated by the inter-particle interactions, and that particles are thus independent time scales comparable to the framerate.
- The objects are virtually point-like (feature-less), and thus perfectly indistinguishable, so that only proximity of a particle from one frame to the particles in the next frame will be used to compute the most likely assignment.
- Particles have no drift velocity.
- Particles have the same diffusion coefficient. Its value does not need to be known at this stage.

Now if we postulate a given one-to-one assignment (each particle in a frame is linked to a single particle in the next frame) which results in observing N individual displacements $\{\boldsymbol{\delta}^l\}_{l\in[1;N]}$ for the N noninteracting particles, the most likely set of trajectories could be constructed by iterating over all consecutive pairs of frames and finding the optimal assignment resulting in optimal displacements $\{\boldsymbol{\delta}^l_*\}$ maximizing $p(\{\boldsymbol{\delta}^l\}|\tau)$ among all possible

$\{\boldsymbol{\delta}^l\}$ of the pair.
An exact derivation can be done (see Appendix), which yields

$$\{\boldsymbol{\delta}_*^l\} = \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmax}}\, p(\{\boldsymbol{\delta}^l\}|\tau) = \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmin}} \sum_{l=1}^{N} \frac{(\delta^l)^2}{4D\tau} + \frac{Nd}{2}\log\left(4\pi D\tau\right). \tag{3.3}$$

Due to the assumption that all particles have equal diffusion coefficient D, this finally reduces to a problem of minimizing the sum of squared distances

$$\{\boldsymbol{\delta}_*^l\} = \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmin}} \sum_{l=1}^{N} (\delta^l)^2 \tag{3.4}$$

This case-study provides an examples of how probabilistic frameworks can naturally incorporate physical models and biological insights on object stochasticity to aid motion reconstruction, as this case even leads to a closed-form solution.

In the following, we will gradually introduce fundamental and technical challenges that are common in tracking biological objects. We will start by introducing a fundamental limitation of tracking Brownian movement beyond a limit defined by a criterion relating dynamics diffusiveness, framerate, and object density. Then, we will focus on the technical problem of solving the assignment problem in practice, especially when introducing objects appearances and disappearances (i.e when the number of objects differ between two consecutive frames). Finally, we will how tracking can be achieved when there is heterogeneity in the individual properties of the particles (e.g different values of the diffusion coefficient D) and their motion patterns.

3.4 The intrinsic limitations of diffusive dynamics, framerate, and density

A fundamental limitation of the aforementioned approach arises when attempting direct interpretation of the rationale behind the optimization step. Minimizing the sum of squared displacements intuitively implies that particles do not move "too much" between two consecutive frames so that they remain close to their previous positions. This instinctive argument can be formalized by comparing the order of magnitudes of two typical lengths linked to the system: the typical displacement due to diffusion between two consecutive frames separated by a time τ

$$\ell_{dyn} \sim \sqrt{2dD\tau} \tag{3.5}$$

and the typical half-distance between nearest neighbors (assuming a regular arrangement of particle positions)

$$\ell_{nn} \sim \frac{1}{2}\left(\frac{V}{N}\right)^{1/d} = \frac{\rho^{-1/d}}{2} \tag{3.6}$$

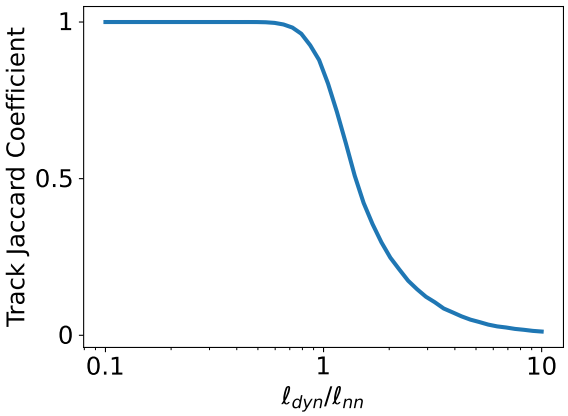

Figure 3.1: **Illustration of the tracking complexity as a function of the adimensional number ℓ_{dyn}/ℓ_{nn}.** A simulation of 1000 particles obeying the 2D Langevin dynamics is used. In the regimes $\ell_{dyn}/\ell_{nn} \ll 1$ and $\ell_{dyn}/\ell_{nn} \gg 1$, the Track Jaccard Coefficient (see Appendix 6.5), which quantifies the quality of inferred tracks, is respectively 1 and 0.

for N particles in a volume V, or equivalently for particle density ρ.

Though coarse, these estimates indicate that the algorithms is bound to perform poorly when $\ell_{dyn} \gg \ell_{nn}$, i.e when particles' displacements are so large that two particles are likely to "exchange" positions between two frames. This criterion clearly links the diffusion coefficient, the framerate $1/\tau$, and the density of particles in a single parameter ℓ_{dyn}/ℓ_{nn} which fully determines the tracking complexity of the system in terms of dynamical parameters. This criterion is illustrated on Fig. 3.1 with a simple simulation of 2D Langevin dynamics. This problem is extensively studied in [150].

Despite the simplicity of the idea behind this criterion, it reflects a fundamental challenge in all tracking problems regardless of the complexity of the system of interest, with direct impacts on experimental imaging protocols: by involving the framerate, whose value is constrained e.g by phototoxicity or hardware limitations, it becomes clear that there is a fundamental trade-off between the image acquisition speed and the complexity of the tracking problem.

3.5 The challenge of solving the assignment problem with lineage constraints

3.5.1 Early strategies based on frame-to-frame assignment

The naive strategy proposed above consists in iterating over consecutive pairs of frames and going through all possible assignment hypotheses for each pair to find the one associated to the highest likelihood. In practice, finding the optimal assignment which minimizes the sum of squared displacements by running through all the $N!$ possible assignment hypothesis and

computing the N pair-distances becomes intractable for large N. For instance, 100 particles lead to $100! \approx 10^{158}$ assignments to compute. Though many extended studies in the fields of radar detection and aircraft tracking had proposed well-grounded methodology for the problem of tracking multiple particle with an assignment problem [142, 151, 152], these tools were not immediately used in biology.

Both Crocker et al. [143] and Bao et al. [144] solve the assignment problem by drastically pruning the space of possible assignments, considering only candidate matches in a radius R around a particle. This process is called gating. If R is chosen smaller than the typical distance to a particle's nearest-neighbor, as is done by Crocker et al. [143], the problem of finding the best global assignment reduces to several problems of finding best local assignment for each particle independently, which the author solve by running through all possible local assignments in a brute-force manner. This reduces the use of their algorithm to extremely dilute systems (or equivalently systems imaged with a high framerate as seen in 3.4) without the biological lineage constraints inherent to developmental biology. In Bao et al. [144], the authors face the issue of finding the optimal assignment in the presence of cell division. They design a heuristic procedure mixing nearest-neighbor search and handcrafted heuristic scores based on daugther-mother distance and visual features to rank association hypotheses. Though sufficient for lineage reconstruction in *C. elegans*, for which framerate is high enough that apparent cell movements are small, the scoring rules defined for division detection and the lack of advanced motion modeling prevent directly applying this method to more complex datasets.

3.5.2 Integer linear programming for frame-to-frame assignments

3.5.2.1 Formulating frame-to-frame assignment as a Linear Assignment Problem

In Jaqaman et al. [73] (which acknowledges the work of [142]), the authors provide a framework that formalizes the reconstruction problem as a bipartite-graph matching problem, which can be solved efficiently. In practice, this requires computing once a matrix $\mathbf{C} \in \mathcal{M}_N(\mathbb{R})$ of all possible assignment log-likelihoods (or equivalently the squared pair-distances between particles if the simplifying assumptions we have used hold). The authors set the matrix element c_{ij} to the squared distance $(\delta^{i\to j})^2$, where the superscript $i \to j$ represents the link from particle i at frame t and particle j at frame $t+1$. In this framework, a candidate global assignment is represented by a binary matrix $\mathbf{A} \in \mathcal{M}_N(\{0,1\})$, with $a_{ij} = 1$ signifying $i \to j$, $a_{ij} = 0$ otherwise. As such, $\mathbf{A}$ is subject to constitutive constraints: due to the one-to-one nature of the assignment, the values in each line and column sum to 1, which respectively reflect that each particle at frame t can be linked to exactly one particle in the next frame, and that each particle at frame $t+1$ can be linked to exactly one particle in the previous frame. The sum of squared distances for a given assignment hypothesis is then elegantly expressed as

$$\sum_{\substack{i,j\in[1,N]^2,\\ s.t\ i\to j}} (\delta^{i\to j})^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}c_{ij}. \tag{3.7}$$

The assignment problem is finally formalized as the optimization problem of finding the best assignment matrix $\mathbf{A}_*$ satisfying

$$\begin{aligned} \mathbf{A}_* &= \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} a_{ij}, \\ \text{s.t} \sum_{i=1}^{N} a_{ij} &= 1, \sum_{j=1}^{N} a_{ij} = 1. \end{aligned} \tag{3.8}$$

Equation (3.8) is called a linear assignment problem and can be solved efficiently with the Hungarian algorithm in $\mathcal{O}(N^3)$ complexity [153]. Note that the solution obtained with a matrix $\mathbf{C}$ is the same as with the matrix $\tilde{\mathbf{C}} = \alpha\mathbf{C} + \beta$, $\alpha \in \mathbb{R}_*^+$, $\beta \in \mathbb{R}$. To further reduce computation time in case of large number of particles, a gating radius R can be provided so that distances greater than R are not considered when creating the pairwise squared distance matrix $\mathbf{C}$. This renders $\mathbf{C}$ highly sparse, making it efficiently computable using data structures like *Kd-Trees* [154], now widely available in standard programming languages [114].

Note that the algorithm we have sketched at this point optimally solves the problem of assigning N perfectly detected indistinguishable point-like particles undergoing Brownian motion, given perfect object detections.

3.5.2.2 Unbalanced assignment problems

In Jaqaman et al. [73], the authors apply the tracking algorithms based on the linear assignment problem to the reconstruction of the trajectories imaged with fluorescence microscopy. They face the issue that the objects appear and disappear from the field of view due to (i) parts of the trajectories leaving and reentering the field of view, (ii) objects being misdetected (creating false negative detections), or (iii) noise creating false positive detections. These types of issues are common in cellular datasets, and at least systematically appear in organoid datasets. In practice, this means that the number of detected particles may change between two consecutive frames, and that trajectories are expected to start or end when the object appears or disappears from the field of view.

Again, this problem can be solved in the LAP framework by adding virtual "dummy" objects to account for the possibility of disappearances and appearances [73, 155]. Let N and M be the number of detected objects in two consecutive frames t and $t+1$. A new square cost matrix $\mathbf{C} \in \mathcal{M}_{N+M,M+N}(\mathbb{R})$ is constructed with M added rows and N added columns, which represent respectively dummy appearance and disappearance nodes in the graph matching problem. The matrix elements are defined as

$$c_{ij} = \begin{cases} c_{i,j}, & i \leq N,\ j \leq M, \text{ (top-left sub-matrix)} \\ c_b^j, & i > N,\ j \leq M, \text{ (bottom-left sub-matrix)} \\ c_d^i, & i \leq N,\ j > M, \text{ (top-right sub-matrix)} \\ c_{\text{aux}}, & \text{else, (bottom-right sub-matrix).} \end{cases} \tag{3.9}$$

where $c_{i,j}$, c_b^j and c_d^i are respectively costs for linking object i to object j appearances and disappearances (births and deaths), and c_{aux} is a small constant auxiliary cost injected in the sub-matrix in the bottom-right part of the full cost matrix, which only serves to satisfy the constitutive constraints of the assignment problem.
In the earliest implementation of Jaqaman et al. [73], the authors chose to set $c_{i,j} = (\boldsymbol{\delta}^{i \to j})^2$ instead of properly defining the linking likelihood $-\log\left(p(\boldsymbol{\delta}^{i \to j})\right)$ as it is a well-grounded approximation (see equation (3.4)) and a robust choice for most simple tracking tasks.

The optimal assignment matrix $\mathbf{A}_* \in \mathcal{M}_{N+M,M+N}(\{0,1\})$ is then obtained by solving the assignment problem of equation (3.8) using the full cost matrix.

3.5.2.3 Enforcing lineage constraints

As stated, this assignment framework produces optimal decisions based on spatially-global contexts, but remained greedy in time, as the assignment is done on pairs of consecutive frames. Thus, it can produce gaps in reconstructed trajectories, and it lacks temporal context to enforce lineage constraints. Jaqaman et al. [73] propose solving a second linear assignment problem called gap-closing, dedicated to solving these issues. In this framework, the start and end points of all tracks obtained after the initial assignment phase are considered for potential fusion between themselves or with intermediary points of neighbor trajectories. This accounts (i) for gaps by allowing starts and ends separated by a few frames (generally less than 3 frames) to fuse, (ii) for splits, e.g cell divisions, by allowing a track start (e.g belonging to a daughter cell) to fuse to a point of an existing track (this implies that when a division happens, the mother object always ends up linked to one of the daughter object), (iii) for merges, by allowing a track end to fuse to a point of an existing track.
This again requires to set merge/split/gap-closing costs (and equivalently costs to reject the fusion hypotheses if the tracks starts and ends obtained in the first step are legitimate).

Btrack by Ulicna et al. [74] implemented a similar algorithm for the gap-closing step following the initial frame-to-frame assignment. Due to advanced probabilistic modeling of cellular dynamics based on visual features, the authors are able to assign to each cell a precise probability of it belonging to a set of typical states of the cell-cycle (e.g interphase during standard migration, metaphase before a division, apoptosis before a cell death, etc). By injecting this information directly in the cost matrix of LAP in the frame-to-frame assignment, they can restrict biologically infeasible links directly at the first stage of the tracking This prevents for instance a cell from going from anaphase directly to interphase, as is possible in *u-track* [73]. To solve the gap-closing step, the authors used a method called multiple hypothesis testing [156], which considers the same hypotheses as the gap-closing algorithm of *u-track*, but better fit their probabilistic framework.

3.5.3 Generalizing integer linear programming for global enforcement of biological lineage constraints and multiple-candidate assignment

3.5.3.1 Improving segmentation based on temporal context

To deal with epistemic uncertainty in tracking coming from systematic errors in detection and shape reconstruction procedures, many state-of-the-art trackers in developmental biology implement an assignment strategy that can take candidate segmentations as input and use temporal context during the tracking procedure to refine the segments (e.g by choosing to split a segment if the local trajectories in a local spatiotemporal window suggest an undersegmentation error).

Pioneering work by Amat et al. [46] introduced TGMM, an efficient tracking framework designed for large-scale fluorescence microscopy datasets, notably applied to *Drosophila melanogaster* embryogenesis. As nuclei are mainly concentrated along the surface of the drosophila's oocyte, the authors used lightsheet microscopy, which allowed high framerate, thereby reducing the complexity related to cell dynamics. Besides, this type of datasets bears computational challenges that resemble those present in gastruloid datasets, mainly (i) packed nuclei and rapid deterioration of the SNR in some parts of the sample, hindering nuclei segmentation, and (ii) the computational challenge of scalability, as the datasets analyzed reached terabyte-scales.

By modeling nuclei shapes via ellipses and fitting Gaussian distributions directly on the image intensity data, TGMM provided a novel method to segment and link nuclei jointly. In a first step, this method iteratively fits a Gaussian mixture model on the image voxel intensity data, initialized by propagating the result of the fit from the previous frame. This fitting procedure requires that the number of Gaussian distributions stays constant between consecutive frames, which lead to cells (e.g daughter cells in a division event) being fitted under a common Gaussian distribution. In a second step, division events are detected by finding distributions that can be split to improve the Gaussian mixture fit. Finally, sliding windows are ran over the frame sequence and heuristic laws are applied to detect and correct unlikely track starts/ends, fuse tracks, etc.

Guignard et al. [157] expanded the application of TGMM [46] beyond lineage reconstruction to cell communication in ascidian embryos. The authors studied contact-mediated neighbor interactions between cells, and thus designed a protocol which allowed segmenting cell membranes accurately, and leveraging the sequential tracking and segmentation introduced by TGMM to propagate membrane segments between frames and refine them based on the local intensity pattern.

McDole et al. [47] tackled nuclei tracking in post-implantation mouse whole-embryos, monitoring around 10,000 cells over 48 hours. Again, the authors chose to analyze cells that were located close to the surface of the embryo, and thus could maintain high framerates throughout. Their adaptive tracking methodology based on TGMM [46], combined with efficient storage and visualization pipelines, enabled large-scale tracking while maintaining computational feasibility. However, due to the highly dynamic nature of the tissue,

lineage reconstruction remained incomplete over long durations. Instead, the authors analyzed coarse-grained collective cell flow obtained from the incomplete tracks. While this large-scale flow analysis is useful for global pattern detection, it falls short in organoids where comprehensive lineage information is needed to detect rare differentiation events and short-range cell-cell interactions. The authors also generated mosaic data, in which a fraction of cells are fluorescently labeled, to obtain sparse tracks that were easier to reconstruct, but could not give insight on local cell-cell interactions.
This algorithm and analysis protocol was also used with minor changes and visual calibration to study EMT during mouse embryo heart morphogenesis [41].

3.5.3.2 Widening temporal context

The linear programming framework defined previously for frame-to-frame assignments can be extended to a multiframe version to improve temporal context [158]. Formulated in a particle tracking context in [159], the generalized multi-frame assignment problem bears similarity with equation (3.8), in the sense that it defines the assignment via a binary tensor $\mathcal{A} \in \mathcal{M}_{N_1,N_2,\ldots,N_{w-1},N_w}(\{0,1\})$ and associated assignment cost tensor $\mathcal{C} \in \mathcal{M}_{N_1,N_2,\ldots,N_{w-1},N_w}(\mathbb{R})$, with $\{N_1, N_2, \ldots, N_{w-1}, N_w\} \in \mathbb{N}^w$ the number of objects in w consecutive frames, with w the size of a sliding window, which could theoretically be chosen as the total length of the movie. The minimization problem reads

$$
\begin{aligned}
&\mathcal{A}_* = \underset{\mathcal{A}}{\operatorname{argmin}} \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \cdots \sum_{m=1}^{N_{w-1}} \sum_{n=1}^{N_w} c_{k,l,\ldots,m,n}\, a_{k,l,\ldots,m,n}, \\
&\text{s.t} \\
&\sum_{l=1}^{N_2} \cdots \sum_{n=1}^{N_w} a_{k,l,\ldots,m,n} = 1, \quad k = 1, \ldots, N_1 \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
&\sum_{k=1}^{N_1} \cdots \sum_{m=1}^{N_{w-1}} a_{k,l,\ldots,m,n} = 1, \quad n = 1, \ldots, N_w
\end{aligned}
\tag{3.10}
$$

This multiframe assignment framework can serve as a canvas to design advanced laws for both direct assignment and gap-closing by leveraging features present in several frames. This framework was exploited in [159, 160] in the context of Bayesian particle tracking, in which particles tend to appear and disappear frequently and detections are plagued with clutter (many false positive detections).

In the field of developmental biology, Malin-Mayor et al. [161] implemented *Linajea* which applies the same network flow integer programming algorithm to track nuclei first in *C. elegans*, then zebrafish and drosophila embryos imaged with lightsheet microscopy. After segmenting nuclei using a U-Net-based CNN [92], the authors solve the assignment problem by deploying an integer linear programming (ILP) framework called network flow integer programming that allows formalizing and solving both the associa-

tion and biological-lineage-constraints-enforcement tasks in a single optimization problem. Though no method capable of finding the globally optimal solution exist, many optimization software implement efficient approximate methods that still lead to improved global coherence of the reconstructed lineages thanks to the "one-shot" formulation of the mixed problems.

The authors record drastically better performance than TGMM in lineage reconstruction, and show that the network flow ILP formulation can appropriately incorporate sparse ground-truth annotations as fixed constraints to help the optimization algorithm resolve ambiguous case. They also point out that network flow ILP is a scalable solution as it conveniently lends itself to parallel processing by solving the optimization problems in spatio-temporal windows, with added constraints to stitch the results.

The same workflow was also used in the algorithm *OrganoidTracker* by Kok et al. [145] to reconstruct lineages in images of intestinal organoids imaged with confocal microscopy. In their datasets, nuclei are essentially motionless and can be imaged with good resolution and high SNR. Due to the nature of their studies, they choose to not perform shape reconstruction, and only detect nuclei centers using a U-Net-based CNN [92]. They obtain excellent detection results, and so the assignment reduces to appropriately finding cell divisions and cell deaths. Local errors in detections do lead to lineage reconstruction errors, but the numbers of nuclei and errors are low enough that the authors report visually inspecting the tracking results to manually correct erroneous associations.s

3.5.3.3 Joint multiple-candidate assignment and global enforcement of biological lineage constraints

New tracking algorithms have recently appeared, which jointly implement the two methodologies we presented: (i) generalizing linear programming for wider (potentially global) temporal context windows, and (ii) leveraging temporal context to decide on the optimal segmentation during the tracking process.

Like TGMM [46], *ECLIP* by Turetken et al. [162] refines imperfect nuclei segmentation by modeling nuclei as ellipses. ECLIP first uses a segmentation algorithm that is designed to favor undersegmentation over oversegmentation of nuclei. From a candidate segment (which can actually be composed of several nuclei), ECLIP builds a hierarchy of all potential ellipses (from large to small) that could reasonably fit the candidate segment. The algorithm then formulate the choice of choosing the best non-conflicting hierarchy elements directly into a network flow ILP.

The newest algorithm following this design is *Ultrack* by Bragantini et al. [163], designed to perform tracking in terabyte-sized datasets of zebrafish embryos. The algorithm improves upon the approach of Turetken et al. [162] by using segmentation hierarchies built from watershed segmentations, which allows much more general shape modeling than ellipsoidal geometries, and thus provides a universal algorithm for tracking objects with arbitrary shapes.

3.5.4 Designing robust linking weights

The trackers discussed in this section rely on an assignment procedure formulated as an optimization problem that minimizes a cost function aggregating individual linking costs, e.g object linking, division, or track initiation/termination. While we have primarily examined the assignment frameworks employed by each algorithm, we have not yet addressed the definition of robust costs. As demonstrated in our case-study on multiple Brownian particle tracking, optimal costs arise from accurately modeling transition likelihoods. This, in turn, requires precise characterization of the dynamical transitions in observable object features, such as motion, morphology, or gene expression.

3.6 The challenge of modeling cell dynamics

Though we derived a closed form solution in equation 6.10 for the linking likelihoods in our case-study, we made two unrealistic assumptions about the particles, namely that they all have (i) no predictable drift velocity, which discards cell migration and collective flow and (ii) the same diffusion coefficient, which prevents modeling cell-to-cell variability or population-dependent dynamics. We also assumed that particles were point-like and thus feature-less so as to not introduce the added complexity of modeling dynamical behaviors in cell morphology or gene expression.

In this section, we will focus on how different tracking methods design linking costs.

3.6.1 Heuristic cost definitions

When first introduced, *u-track* [73] based its linking scores on the squared-distance function, a popular choice taking roots in the types of derivations of Section 3.3.1. This also promotes defining search radii around objects to limit the number of candidate matches, which reduces computational complexity drastically.

Properly defining the birth and death costs in terms of likelihoods is often an arduous task that requires many assumptions on the probabilistic dynamical behavior of the system (e.g modeling how cells behave close to the boundaries of the field of view, or the frequency of detection errors, like false positives/negatives). To help users define track deaths and births costs during frame-to-frame linking and gap-closing, and merge, split, and fusion costs during gap-closing, the authors defined handcrafted heuristics to score different hypotheses (e.g by quantifying if the local arrangement of objects is compatible with a cell division, or injecting biological priors to restrict the minimal time between division events). They set both costs to a high (e.g 80th) percentile of the reduced cost matrix value distribution to insure a conservative tracking (i.e to favor tracks ends and starts rather than erroneous links), or to an upper bound of the distribution's maximum (i.e $c_b = c_d = \alpha \max_{i,j} c_{ij}$, $\alpha > 1$) when gating is already used to prune the cost matrix entries and remove the most unlikely hypotheses.

The authors praise the modular formulation and generalizability of their framework, suggesting advanced strategies like probabilistic modeling of cell dynamics to improve the robustness of the costs. Nonetheless, such heuristics have proven robust and permitted the wide-scale dissemination of this algorithm beyond the field of bioimage analysis and to non-expert users thanks to collective efforts in software engineering and user-interface design in projects like *Trackmate* [139] or *u-track* [73, 75]. More fine-tuned values for these parameters can be obtained by using a search strategy (e.g grid-search) over a hyperspace of candidate parameters.

Elephant by Sugawara et al. [101] is a tracking software designed around accessibility, user-friendliness, and interactive manual annotation and correction of segmentations and tracks. Elephant promotes a human-in-the-loop approach, in which the user quickly draws a few manual ellipsoidal objects over objects (mostly nuclei) and link them across few frames. The software then trains two custom U-Net-based CNN [92] to respectively segment and predict dense velocity flows on the sparse annotations. The predicted objects are then advected along the predicted flows and a simple nearest-neighbor algorithm is used to match objects between consecutive frames, with a heuristic search technique based on proximity and shapes of neighbor ellipses to detect divisions.

In OrganoidTracker [145], the authors define heuristic hypotheses rules based on distance between candidate matches, and absolute change in volume and intensity measured in ellipsoidal regions around nuclei centers. These scores are all aggregated after being multiplied by a factor that depends on relative importance chosen between the different hypotheses. The authors recommend changing these factors rather than the heuristic rules when applying their methods to new data. They do not specify if they found the parameters they use by grid search, or via visual validation. After these results appear, the user can interact with the predicted ellipses and links to reassign or remove them. A new training phase ensues, and the process can be repeated until satisfactory results are reached. This approach efficiently tracks nuclei in *C. elegans* embryos.

In Ultrack, the linking costs between segments contained in hierarchies is simply IoU^{γ}, with IoU the intersection-over-union score, and $\gamma \in \mathbb{R}^{+},\ \gamma > 1$ a constant used to mitigate oversegmentation (without γ, if the hierarchy of an object does not vary over the course of several frames, e.g if the object is static, the ILP might decide to systematically choose the set of smallest members of the hierarchy). This formulation is simple but particularly robust in datasets imaged with a high framerate, as is often the case with lightsheet microscopy, since objects have small movements between frames and so their shape and position is well approximated by simply propagating the information from the previous frame. Costs for divisions and tracks starts and deaths are determined using grid-search thanks to large amounts of in-house annotated tracks.
When applying this method in later stages of zebrafish development, the authors find rapid collective flows of cells due to key morphogenesis steps (e.g gastrulation) which reduce the accuracy of lineage reconstruction. To solve these issues, they compute a dense approximation of the coarse-grained velocity field using optical-flow [52], which uses the spatio-temporal correlation of a pixel's intensity to estimate smooth velocity fields in local boxes. They use it to advect the segments from one frame to the next in the hope that the two segments will overlap. This approach is again very effective for late zebrafish or

drosophila embryo datasets, but assumes coordinated drift at large spatial scales. Applied to gastruloid movies, these methods were shown to provide sensible velocity fields at spatial resolutions down to a few cell diameters [27, 36], which is not sufficient to resolve the highly heterogeneous dynamics (both in time and space) of individual cells.

3.6.2 Learning from manually annotated tracks

Machine-learning methods can be used to infer linking costs directly from manually annotated tracks.

Turetken et al. [162] define a well-grounded probabilistic framework to formulate how to use linking likelihoods to find the best tracking result in their use case. To compute linking and division likelihoods, they use classical (non-deep) classifiers to learn scores from a small set of hand-picked features (e.g distances between objects, difference in intensities, or features of the ellipsoidal hierarchies they use for segmentation).

Interestingly, in Linajea [161, 164], the authors choose a cost function similar in design to that of OrganoidTracker [145], but show that their formulation allows using sparse ground-truth annotations of tracks (i.e only a fraction of tracks need to be annotated, and they do not need to be chosen close to each other spatio-temporally) to directly learn optimal parameters (equivalent to the multiplicative factors in OrganoidTracker) by solving a structured support vector machine problem [164]. Like Ultrack [163], they also use CNN-predicted optical flow to refine their distance-based cost by computing the distance between detections advected by the local flow and candidate linking detections.

More recently, Gallusser et al. [165] train transformers to predict likelihoods directly from matched regions of interest taken from raw pixel images of the data and corresponding segmentations. The network is trained by teaching the network to classify whether a link exists between two objects based on raw images and corresponding segmentations cropped in bounding-boxes around the two objects of interests.

Given enough training data, this potentially allows extracting as much information as possible from the data by letting the deep neural network find the best set of features directly from pixel values, alleviating human-picked features that might be biased or inefficient. The authors incorporate division constraints directly into the transformers architecture, which allows for a single network to determine all potential hypotheses likelihoods. They then use network flow ILP to solve the lineage reconstruction task.

3.6.3 The case of deep methods

Advances in deep-learning have also created entirely new paradigms for cell tracking.

Delta by [166] uses a 3D U-Net-based CNN [92] that takes as input a pair of consecutive images, a binary mask delineating a cell of interest, and a binary mask delineating all cells (all four inputs are stacked on a virtual third dimension), and outputs a mask of the matched cell in the next frame, a mask of a potential daughter cell, and a mask of background pixels.

Though straight-forward, the method was successfully applied to track 2D movies of highly dividing *Escherichia coli* cells trapped in a "mother machine" microfluidic device.

In Ben-Haim et al. [146], authors use a graph-neural-network (GNN, which generalizes operations from CNN to graph-structures) to represent the entire tracking graph composed of all linking hypotheses, and train the network to classify each link as real or not from features learnt automatically from pairs of images and annotated tracks.

Though not available in their current implementations due to their requirement for large amounts of annotated tracks, these options could be useful in the future if a version of the deep sparse training (as is implemented in Elephant [101]) can be adapted to the algorithms.

3.6.4 Direct probabilistic modeling via state-space models

Circling back to the start of this chapter, in which we have shown how appropriate physical modeling combined with probabilistic frameworks can be used to infer optimal linking likelihoods, we present here probabilistic methods which have been successfully used to solve diverse problems of bioimage tracking.

Chenouard et al. [160] used precise probabilistic modeling of clutter (false positive), movement, and object lifetime to track point-like molecules imaged with fluorescence microscopy. The Bayesian probabilistic framework allows computing approximated likelihoods based on well-grounded equations. To insure robustness to rapidly accelerating particles, the authors implemented a multiframe association framework and showed that they could learn dynamical parameters (allowing individual motion prediction tailored to each object) for each object individually, and that solving the multiframe association problem using their formulae to compute likelihoods led to nearly optimal tracking.

The u-track [73] algorithm was gradually updated to incorporate Bayesian motion modeling of individual objects via Kalman filters (KF), which, combined with their existing robust gap-closing framework, allowed reliable cell tracking in scenarios challenging both due to the important motion and to the lineage constraints [167]. These methods will be presented in greater technical details in the following chapter.
Later, an advanced probabilistic algorithm based on locally stationary motion models allowed modeling datasets in which objects brutally switch between motion models (e.g in capture scenarios where molecules brutally switch from confined diffusive dynamics to highly directed motion) [168].

Btrack [74] is one of the most advanced Bayesian tracker used for cell tracking currently. In their original publication, the authors train machine-learning classifiers to predict precise probabilities of a cell to belong to a each elements of a set of cell-cycle states based on images, which they incorporate in their frame-to-frame association and gap-closing frameworks along with Bayesian motion modeling with KF to fully represent the transition probability that an object will switch from a given triplet of (position, morphology, gene expression) to another.

3.6.5 Summary

Accurate cell tracking in developmental biology requires addressing multiple interconnected challenges, from the fundamental uncertainties in stochastic motion to the computational complexity of linking cell trajectories over time. While classical methods based on frame-to-frame assignments provide a foundation, more sophisticated approaches leveraging integer linear programming (ILP) and deep learning have significantly improved robustness and accuracy. Modeling cell dynamics allows representing motion and cell-state transitions more accurately, which is mandatory for accurate lineage reconstruction. While heuristic cost functions are still routinely implemented in state-of-the-art ILP-based tracking strategies, these algorithms cannot handle large motion and heterogeneous dynamics. In this matter, Bayesian frameworks and deep learning have been successfully applied to track complex cell dynamics by modeling dynamical heterogeneities at the scale of individual objects. Though deep-learning methods remain inaccessible for gastruloid datasets due to the amount of manually annotated tracks they need to be trained on, we believe that Bayesian tracking methods can provide robust solutions adapted to cell tracking [74, 167], and that they have an important role to play in helping researchers address current challenges in developmental biology.

3.7 The challenge of direct visualization in cell tracking strategies

3.7.1 All tracking methods require parameter tuning based on some form of visualization

All methods presented in this chapter rely on parameters that need to be adjusted to the user's specific use-case. Whether for visual assessment and corrections of tracking results, or the use of ground-truth annotated tracks (which are produced by drawing on images of the data) for parameters calibration, visualization tools play an important role in current strategies used to reconstruct lineages reliably in cell tracking and in developmental biology.

3.7.2 State of the art

In bioimage analysis, validating tracking results is crucial for precise parameter tuning during image acquisition and analysis, as well as for ensuring accurate biological measurements.

Biophysicists still rely heavily on direct visualization before processing images to (i) gain modeling insights and (ii) design new experiments. After processing, such as denoising, segmentation, or tracking, visualization is also routinely used to qualitatively assess the performance of these steps. Incidentally, platforms that promote joint visualization

and processing are prominent in biological community [69, 169–171]. Those approaches provide an outside view of the 3D data and a range of tools to investigate the volume manually (e.g., in focused ROIs or slices). They also reveal many challenges associated with visual occlusions, 3D perception on a flat screen, interactions, and time-consuming volume exploration.

BigDataViewer [169] has tackled the occlusion and perception challenges indirectly by focusing on the caching and rendering of 2D slices at arbitrary orientations. The primary effect is to accelerate rendering at the expense of 3D context, but it also avoids distortion of pixel-level information. As such, the Mamut [172] software has demonstrated that combining BigDataViewer and 3D rendering of vectorial information enables annotation in large and complex 3D trajectories as long as the density remains low. The immersive quality of virtual reality (VR) headsets has also been exploited to better perceive and interact with complex 3D datasets. [173, 174]

Advances in GPU rendering methods also bring powerful open-source solutions for real-time interaction with terabyte-sized datasets. For example, InTracktive [175] provides a browser-based graphical interface to investigate single tracks and whole ROIs. Though technically impressive, this method is showcased on cell trajectories in whole-mount zebrafish embryos, which primarily lie on a pseudo-2D subspace. On the other hand, Napari [69] provides excellent built-in interactivity option and virtually-limitless customization through plugins.

Finally, a complementary approach to full volume visualization is the presentation of software-defined ROIs. The embryoMiner package [171] implements this idea by building static ROIs from groups of trajectories that are then used to systematically visualize the data in a series of 3D tiles. Approaches implemented in Matlab in the Bayesian tracking suite u-track3D [75] also belong to this category. However, the ROIs adapt over time to the changing geometry of collective processes, which allows the inspection of specific particle dynamics in a crowded and dynamic surrounding by looking at 2D slices in perpendicular directions.

As exposed above, Bayesian tracking frameworks are highly appropriate to incorporate biological and physical priors, which leads to particularly interpretable parameters that can often be intuitively visualized (e.g covariance matrices to represent spatial uncertainty). This helps in effectively reducing the number of parameters that are to be fitted by exhaustive parameter-space searches by promoting visual validation instead. utrack-3D leverages this principle within their dynamic ROIs by allowing users to superimpose key variables from the probabilistic modeling of individual objects on top of raw intensity images or detections. utrack-3D also implements a tracking ambiguity metric defined at the scale of a single link, called trackability. The authors show how this unsupervised metric obtained through their Bayesian filtering tracking framework allow automated detection of ambiguous tracks and rapid dynamical events to facilitate inspection.

3.7.3 Our custom visualization tool allows interactive and modular exploration in 3D and 2D views

When exploring dense organoid datasets, direct visualization enables researchers to choose appropriate priors and motion models for Bayesian tracking, which are crucial for capturing the nuanced dynamics of cellular behavior. This insight ensures that tracking algorithms align more closely with the physical realities of the biological processes being studied. Moreover, visualization allows iterative tuning of Bayesian tracker parameters. These parameters, which often govern aspects such as search radii and covariance matrices, are not always intuitive and require careful adjustment. By employing a user-in-the-loop approach, researchers can iteratively observe tracking outcomes, refine parameter settings, and repeat the process. This feedback loop, driven by visual assessment, enhances the accuracy and reliability of tracking results.

To address the challenges inherent in exploring 3D+time biological datasets, we present a visualization tool built upon Napari, a Python-based software platform known for its capacity to provide several modalities to handle, explore, and annotate n-dimensional datasets, making it particularly well-suited for bioimage analysis. Our tool extends Napari's capabilities to allow researchers to display volumetric datasets in three dimensions and define regions of interest (ROIs) through intuitive manual selection, such as drawing 3D bounding boxes (Fig. 3.2(a)). These ROIs can then be examined in detail using an orthoviewer, which offers simultaneous visualization of three orthogonal 2D planes (Fig. 3.2(b)). This approach greatly enhances the ability to navigate and interpret complex, dense 3D environments. The tool is available at github.com/jules-vanaret/napari-orthorendering.

The interactivity of the tool further empowers researchers by enabling dynamic customization of visualization parameters. Users can modify display settings, such as toggling the visibility of images, tracks, and segmented objects, adjusting contrast levels, altering track lengths, and applying colormaps to tracks and segmented objects based on attributes like volume, velocity, or trackability. This flexibility promotes a deeper understanding of the data and allows for tailored exploration when addressing specific research questions.

This visualization tool is particularly useful to explore segmentation and tracking in organoid studies, where traditional ground truths are often unattainable. The extreme density of these systems, poor signal-to-noise ratios and heterogeneous dynamics, make automated validation unreliable. By offering an intuitive and interactive visualization environment, our tool enables the user to critically assess and refine their segmentation and tracking results. Ultimately, this contributes to a more robust understanding of the underlying biological phenomena, facilitating discoveries that might otherwise remain elusive.

3.7.4 The role of manual editing of tracks

Manual editing of tracks plays a critical role in cell tracking strategies, particularly in the context of complex developmental biology datasets, where automated algorithms often struggle to achieve perfect accuracy. Despite significant advances in computational approaches, including model-based tracking and deep-learning-based segmentation, errors

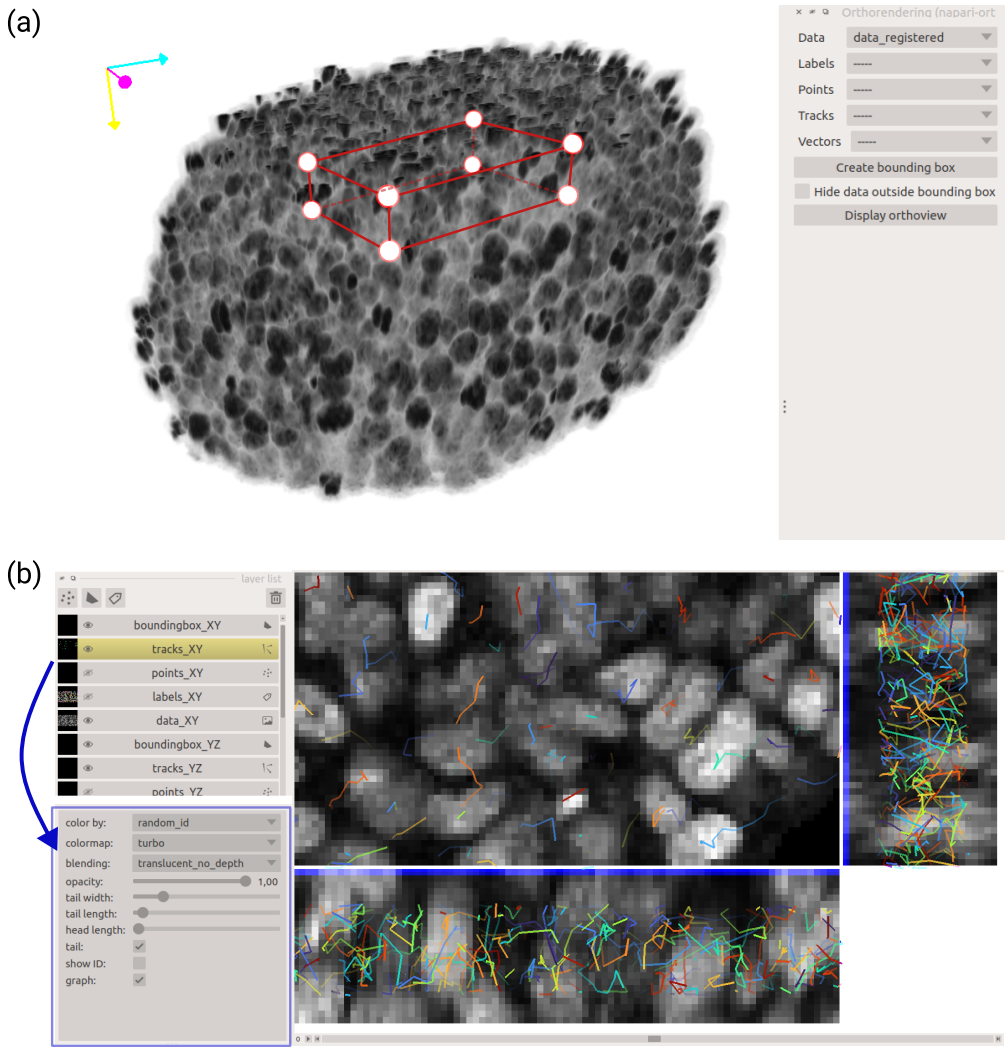

Figure 3.2: **Napari visualization tool overview. (a)** Napari provides a graphical interface that allows manually locating the vertices of a 3D region of interest (ROI), either on a 3D view, or from 2D planes. **(b)** The ROI can then be explored in an orthoviewer, which displays 2D views along the three main directions. This graphical interface benefits from all Napari customization options to tailor the exploration to the data.

remain inevitable due to factors such as cell divisions, close cell interactions, and imaging artifacts. As a result, user intervention remains systematically necessary to refine automatically generated tracks, correct misassignments, and resolve ambiguities in cell lineage reconstructions.

One example of crucial integration of manual correction into cell tracking workflows is the development of the CATMAID annotation toolkit by the Keller team [176], which facilitated the manual curation of automated tracking results in the context of their pioneering work on large-scale lineage reconstruction [46]. Given the complexity of embryonic development and the inherent limitations of their automated method, manual editing was necessary to correct errors in cell lineage assignments and ensure the accuracy of reconstructed developmental trajectories. The authors report that thanks to their custom software, which incorporate efficient user-experience designs like shortcuts for editing and jumping to high-error-probability regions, a novice user can currate a full drosophila dataset of tracked 10^5 objects (number aggregated from the individual counts of each frames) in less than 100 hours. A very conservative estimate performed in [161] gives more than 7000 hours for a user to manually annotate the full lineages from scratch.

More recently, the MaMuT/Mastodon software tools [172, 177], incorporated into the BigDataViewer visualization software of the Fiji platform, has provided an interactive environment for curating cell tracks in large-scale microscopy datasets. Designed to integrate with automated tracking methods, MaMuT/Mastodon enable users to visually inspect, correct, and validate tracking results and lineage trees, ensuring biologically meaningful lineage reconstructions.

These examples highlight the indispensable role of manual editing in cell tracking pipelines. While automation has dramatically improved the scalability and efficiency of cell tracking in developmental biology, expert intervention remains essential for ensuring high-fidelity lineage reconstructions, ultimately enabling more accurate quantitative analyses of cell behavior during embryogenesis. This motivates further integration of such methods in our custom Napari-based visualization tools.

3.7.5 Discussion

Recent breakthroughs in cell tracking methods applied to developmental biology have been brought by new object association frameworks based on integer linear programming which led to algorithms formulating the association task while enforcing biological lineage constraints over the full duration of the movie in a single optimization step. Methods have used this framework to also perform joint optimization of object detection/segmentation via contour evolution methods or candidate segmentation hierarchies, which improves accuracy by leveraging wide temporal contexts. However, these methods cannot handle complex heterogeneous motion, as they assume collective motion patterns that can be estimated at a coarse-grained state in a preprocessing step. On the other hand, Bayesian tracking methods excel at incorporating physical priors to model individual dynamics.

No tracking method is capable of perfectly reconstructing lineage in challenging developmental biology datasets, and the integration of visualization tools remains essential for

refining tracking results and ensuring biological validity. Future advancements will likely depend on the synergy between improved motion models, optimization techniques, and interactive visualization systems that facilitate both automated and expert-driven refinement. Originating from the Bayesian tracker u-track [73], u-track3D provides a full suite of complementary solutions to form an extensive tracking strategy, with (i) a Bayesian tracking algorithm robust to the challenges of cell tracking [167], (ii) GUI-based and script-based visualization via dynamical 3D ROIs, with ortho-rendering options, which allows exploration, visual calibration of tracking parameters, and validation of tracking results, and (iii) automated salient event detection via an unsupervised tracking ambiguity metric. Notably, the recent advances in quantifying assignment-uncertainty via the unsupervised tracking ambiguity metric trackability developed by Roudot et al. [75] suggest several promising applications to the challenges of cell tracking in developmental biology, from helping visualization and guiding inspection, to allowing automatic selection of reliable tracks which reduces the need for globally perfect lineage reconstruction.

The following chapters address these challenges by introducing and validating a Bayesian tracking framework tailored for complex biological contexts, along with practical applications leveraging uncertainty quantification to enhance tracking reliability and algorithmic calibration.

Chapter 4

Bayesian filtering provides a probabilistic framework for unsupervised uncertainty quantification

4.1 Chapter overview

This chapter introduces a basic Bayesian tracking framework designed specifically to manage heterogeneous and rapidly changing cellular movements observed in gastruloid systems. Central to this chapter is the introduction of the trackability score, a Bayesian metric designed in recent studies, which quantifies uncertainty in tracking assignments. Trackability captures variations in tracking quality resulting from particle density, motion complexity, and imaging parameters, making it particularly useful for dense 3D bioimages. Its versatility allows robust visual validation, detection of abrupt dynamical events, and reliable prediction of tracking quality at various scales. This versatility is foundational, as subsequent chapters leverage trackability for new applications, including unsupervised calibration and segmentation algorithm ranking.

My contributions:

1. implemented a modular Bayesian tracker in Python
2. implemented the trackability metric
3. implemented a modular version of the PMMS method in Python

4.2 Bayesian tracking to solve complex and heterogeneous dynamics

4.2.1 Motivation

Bayesian filtering serves as a fundamental framework for estimating and updating the state of a target over time. The Bayesian filtering framework operates through two primary steps: prediction (time update) and correction (measurement update), enabling both forward pre-

diction of the target's state and refinement based on new observations. This approach is particularly valuable for its ability to (i) handle uncertainty, explicitly accounting for both measurement process noise and imperfections in the target's motion model, and (ii) dynamically learn the motion patterns of each particle independently. This capability is crucial when operating in noisy environments with heterogeneous objects, as is often the case in biological imaging [74, 75, 168, 178].

Bayesian filtering also excels in information fusion, providing a natural mechanism for integrating data from multiple sensors or diverse measurement types, which is essential in multi-sensor tracking systems. In the context of bioimage analysis, this property has been leveraged to combine different imaging modalities or to integrate prior biological knowledge into the tracking process [74, 168, 179].

Advanced Bayesian filtering techniques, such as particle filters, are capable of processing nonlinear state transitions and non-Gaussian noise models, making them suitable for complex tracking scenarios where linear approximations fall short. These methods have found particular success in tracking subcellular structures and molecules in live-cell imaging [180]. However, their numerical complexity prevent these methods from scaling to datasets composed of thousands of objects in several hundreds of frames [179, 180].

Kalman filters [181] have provide a scalable but powerful solution to Bayesian tracking challenging bioimage datasets. In bioimage analysis, researchers have adapted Bayesian filters to track diverse biological entities, from individual proteins to whole organisms [75, 167].

4.2.2 Bayesian filtering with hidden Markov models

Bayesian filtering is widely used to recursively estimate the parameters of dynamical systems based on noisy and limited sensor measurements. Here, we model the tracking task a Hidden Markov Model (HMM) in which each object has a hidden state $\mathbf{x}_t$ at time t, which must be estimated using only partial and noisy measurements $\mathbf{z}_t$. Bayesian filtering in this framework consists in recursively using the new measurements to update a state estimate over time.

The system state at time $t \in \mathbb{N}$ is denoted by $\mathbf{x}_t \in \mathbb{R}^N$, with N the number of state variables. For example,

$$\mathbf{x}_t = (x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t)) \tag{4.1}$$

can be used to model the state of a point-like particle in 3D moving in a directed fashion.

The state evolves according to a hidden Markov process. Observations $\mathbf{z}_t$ are conditionally independent given the state. Assuming a first-order Markov process, as illustrated on the hierarchical schematics of Fig. 4.1,

$$p(\mathbf{x}_{t+1} \mid \mathbf{x}_{1:t}) = p(\mathbf{x}_{t+1} \mid \mathbf{x}_t), \tag{4.2}$$

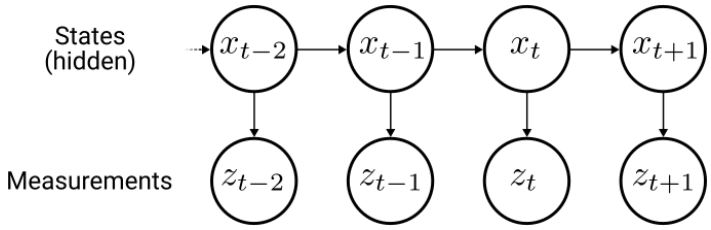

Figure 4.1: **Hierarchical schematics of the hidden Markov model used in Kalman filtering.** In classical Kalman filtering, the hidden state evolves based only on the value of the state at the previous frame. The measurements are generated from the states, and are conditionally independent from each other given the state.

where $\mathbf{x}_{1:t} = \{\mathbf{x}_1, \dots, \mathbf{x}_t\}$. The goal of the filtering framework is to compute the posterior distribution $p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t+1})$ using a recursive approach.

Bayesian filtering consists of two steps: prediction and update. The prediction step computes the prior distribution $p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t})$, as

$$p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t}) = \int p(\mathbf{x}_{t+1} \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \, d\mathbf{x}_t. \tag{4.3}$$

When a new measurement $\mathbf{z}_{t+1}$ becomes available, the posterior distribution is updated using Bayes' rule

$$p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t+1}) = \frac{p(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t})}{p(\mathbf{z}_{t+1} \mid \mathbf{z}_{1:t})}, \tag{4.4}$$

where the denominator is a normalization constant ensuring the probability distribution sums to one. The combination of both steps in a unified equation yields the Bayesian filtering equation

$$p(\mathbf{x}_{t+1} \mid \mathbf{z}_{1:t+1}) = \frac{1}{p(\mathbf{z}_{t+1} \mid \mathbf{z}_{1:t})} \, p(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1} \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \, d\mathbf{x}_t. \tag{4.5}$$

Under the Markovian chain hypothesis that $\mathbf{x}_{t+1}$ can be determined exclusively from $\mathbf{x}_t$, one can propose a dynamical model $f : \mathbb{R}^N \times \mathbb{N} \to \mathbb{R}^N$ for $\mathbf{x}_t$ known up to some precision represented by $w_t \in \mathbb{R}^N$ defined as a realization of a random variable that represent the model (or process) noise

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, t) + \mathbf{w}_t. \tag{4.6}$$

$\mathbf{w}_t$ could represent how an over-simplistic model deviates from reality, or real stochastic terms in the equation of the dynamics, like in the Langevin equation (3.1). The complete state variable $\mathbf{x}_t$ is hidden and information about the object can only be obtained via the measurement of variable $\mathbf{z}_t \in \mathbb{R}^M$, with $M \leq N$ (e.g position but not velocity in the example above). The process of measurement is modeled by a function $h : \mathbb{R}^N \times \mathbb{N} \to \mathbb{R}^M$ corrupted by $v_t \in \mathbb{R}^M$ defined as a realization of a random variable modeling a measurement noise such that

$$\mathbf{z}_t = h(\mathbf{x}_t, t) + \mathbf{v}_t. \tag{4.7}$$

In biological applications, $\mathbf{v}_t$ could represent measurement errors due to optical limitations (e.g particles below the diffraction limit appearing as point-spread functions) or systematic errors in detection algorithms.

4.2.3 Kalman filtering applied to a single particle

Simplifying assumptions are often made to make the posterior PDF tractable [73]. Notably the models f and h for the dynamics and measurement are assumed to be linear, i.e $f(x_t, t) = \mathbf{F}x_t$ and $h(x_t, t) = \mathbf{H}x_t$, with transition and observation matrices $\mathbf{F} \in \mathcal{M}_N(\mathbb{R})$ and $\mathbf{H} \in \mathcal{M}_{M,N}(\mathbb{R})$, and the noises variables $\mathbf{w}_t$ and $\mathbf{v}_t$ are assumed to be sampled from Gaussian laws with covariance matrices Q and R. In this case, $p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t})$ and $p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t+1})$ are normally distributed too.

The definitions of $\mathbf{x}_t$, $\mathbf{F}$, and $\mathbf{H}$ depend on the motion observed. Classical choices include $\mathbf{x}_t = (x(t), y(t), z(t))$, $\mathbf{F} = \mathbf{H} = I$ for Brownian (caged) motion, and

$$\mathbf{x}_t = (x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t))\,,$$

$$F = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

for constant-velocity (directed) motion. Note how the diagonal of 1 in the upper-right part of $\mathbf{F}$ can be interpreted as the hypothesis that objects have (supposedly) constant velocities, and that the absence of 1 on the last elements of the diagonal of $\mathbf{H}$ indicate that the velocity cannot be directly measured by a single snapshot of a movie.

Under these assumptions, Kalman filters give an optimal [181] estimation of the parameters of the dynamics by providing a recurrent formula to compute the prior and the posterior PDF at each frame [168]. Kalman filters can be implemented efficiently for thousands of objects in parallel, as routinely appear in biological imaging, e.g endocytic events, microtubule polymerization at the molecular level and developing embryos at the cellular level.

We note $\hat{\mathbf{x}}_t$ and $\hat{P}_t$ the mean and covariance matrix of $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, and $\bar{\mathbf{x}}_t$ and $\bar{P}_t$ those of $p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t})$. The Kalman filter provides recurrent relations to compute the prediction and update steps of equations (4.3) and (4.4) analytically. The subscript is dropped for $\mathbf{H}.$ and $\mathbf{F}.$ to reduce text overcrowding.

For the prediction step

$$\begin{cases} \bar{\mathbf{x}}_{t+1} = \mathbf{F}\hat{\mathbf{x}}_t \\ \bar{P}_{t+1} = \mathbf{F}\hat{P}_t\mathbf{F}^T + Q \end{cases}, \tag{4.8}$$

and for the update, during which $\mathbf{z}_{t+1}$ is provided

$$\begin{cases} \mathbf{K}_{t+1} = \left(\mathbf{H}\bar{P}_{t+1}\mathbf{H}^T + R\right)^{-1} \bar{P}_{t+1}\mathbf{H}^T \\ \hat{\mathbf{x}}_{t+1} = \bar{\mathbf{x}}_{t+1} + \mathbf{K}_{t+1}(\mathbf{z}_{t+1} - \mathbf{H}\bar{\mathbf{x}}_{t+1}) \\ \hat{P}_{t+1} = (I - \mathbf{K}_{t+1}\mathbf{H})\bar{P}_{t+1} \end{cases}, \tag{4.9}$$

where $\mathbf{K}_{t+1}$ is the Kalman gain matrix, which weighs the relative contribution of the predicted state $\bar{\mathbf{x}}_{t+1}$ and of the new measurement $\mathbf{z}_{t+1}$ to form the parameters of the updated state $\hat{\mathbf{x}}_{t+1}$. I is the identity matrix. We note $\mathbf{y}_{t+1} = \mathbf{z}_{t+1} - \mathbf{H}\bar{\mathbf{x}}_{t+1}$ the innovation of the process, representing the difference between the real and predicted measurements. As such, $\mathbf{K}_i y_i$ can be interpreted as the prediction error.
Owing to their linearity, KFs can be implemented efficiently for thousands of objects, and are thus well suited to tracking applied to biology.

4.2.4 Relevance to the simulation model of 5.4.3.1

In 5.4.3.1, we presented how linear dynamics of the type $\mathbf{v} = \mathbf{F}_{OU}$, with $\mathbf{F}_{OU}$ following an Ornstein-Uhlenbeck (OU) process, can simulate both caged motion and directed motion through the choice of the persistence time τ_p. Though it is possible to bring the OU process in the framework of linear Kalman filtering, this would require to know in advance the values of τ_p for each nucleus beforehand, or to estimate them during the tracking process. When filtering a single object, this is usually solved by switching to non-linear methods [182, 183], which do not scale to thousands of objects. Another possibility could be to develop an hybrid HMM in which the dynamics of the state is modeled by a KF in parallel to a discrete model for the persistence time. To our knowledge, this has not been done before, and could be an interesting lead for future developments.
Similarly, the Kalman filter cannot handle the rest of the dynamics of equation (5.7) due to the nonlinear interactions between the particles. Classical methods for non-linear filtering, like extended Kalman filters [184], are typically used when the dynamical model f (equation (4.6)) is non-linear and known. In this case, the predicted state is obtained by applying equation (4.6) directly, and the predicted covariance is computed using equation (4.8) with a linear model motion matrix $\mathbf{F}_{t+1}$ given by the Jacobian of f

$$\mathbf{F}_{t+1} = \frac{\partial \mathbf{f}}{\partial \hat{\mathbf{x}}_t}. \tag{4.10}$$

Such an approach cannot be used in our case to directly filter using the non-linear simulation model of 5.4.3.1, as the parameters of the model are not known *a priori* and vary from one nucleus to another and with time. Methods coupling dynamical model estimation [56, 57] and filtering in an iterative scheme have not been investigated to our knowledge, but could be developed to address this issue.

However, by assuming that the time Δt between two frames is smaller than the typical timescale of variation of the forces in (5.7) (e.g $\Delta t \ll \tau_p$ for highly directed motion), the forces can be considered approximately constant during this period, and a Kalman filter

with a constant velocity model is expected to give satisfying results, as long as the velocity is appropriately updated.

4.2.5 Bayesian filtering in the multiple-hypothesis context

4.2.5.1 Mixing assignment and filtering

The Bayesian filtering scheme introduced in equations (4.3) and (4.4) is only valid when filtering the dynamics of a single object or when the association between trajectories and detections is known in advance (e.g if the detections are labeled), with state $\mathbf{x}_t$ which is updated once a single measurement $\mathbf{z}_t$ is acquired. In the context of multiple-particle tracking, a set of new measurements $\mathbf{z}_{t+1}$ is proposed and the tracking framework has to decide between each candidate match for the most sensible one. The Bayesian filtering equation for one object i with state $\mathbf{x}_t^i$ is now marginalized over the full set of new measurements $\mathbf{Z}_{t+1}$ instead of a single measure $\mathbf{z}_{t+1}$

$$p(\mathbf{x}_{t+1}^i \mid \mathbf{Z}_{1:t+1}) = \frac{1}{p(\mathbf{Z}_{t+1} \mid \mathbf{Z}_{1:t})}\, p(\mathbf{Z}_{t+1} \mid \mathbf{x}_{t+1}^i) \int p(\mathbf{x}_{t+1}^i \mid \mathbf{x}_t^i) p(\mathbf{x}_t^i \mid \mathbf{Z}_{1:t})\, d\mathbf{x}_t^i. \tag{4.11}$$

Evaluating the new likelihood term $p(\mathbf{Z}_{t+1} \mid \mathbf{x}_{t+1}^i)$ requires considering all possible linking hypothesis with the new measurements, i.e all possible associations resulting from linking the track i to the current measurements. Conditioning the likelihood over the the different assignment hypotheses $\mathbb{H}_t^i = \{\eta^{i\to j} :$ object i at t linked to measurement j at $t+1\}$

$$\begin{aligned} p\left(\mathbf{Z}_{t+1} \middle| \mathbf{x}_{t+1}^i\right) &= \sum_{\eta^{i\to j} \in \mathbb{H}_t^i} p\left(\mathbf{Z}_{t+1}^j, \eta^{i\to j} \middle| \mathbf{x}_{t+1}^i\right) \\ &= \sum_{\eta^{i\to j} \in \mathbb{H}_t^i} p\left(\mathbf{Z}_{t+1}^j \middle| \eta^{i\to j}, \mathbf{x}_{t+1}^i\right) p\left(\eta^{i\to j} \middle| \mathbf{x}_{t+1}^i\right), \end{aligned} \tag{4.12}$$

where the first term in the sum corresponds to the likelihood associated with the hypothesis $\eta^{i\to j}$, while the second term is the predicted probability of the hypothesis $\eta^{i\to j}$.

In practice, the evaluation of this sum is untractable, and so a common approximation consists in only evaluating the single-measurement likelihood $p\left(\mathbf{z}_{t+1}^{j^*} \middle| \mathbf{x}_{t+1}^i\right)$ using the optimal measurement j^*

$$p\left(\mathbf{Z}_{t+1} \middle| \mathbf{x}_{t+1}^i\right) \approx p\left(\mathbf{z}_{t+1}^{j^*} \middle| \eta^{i\to j^*}, \mathbf{x}_{t+1}^i\right). \tag{4.13}$$

The notion of optimality of the measurement is again decided by solving a bipartite assignment problem between current states $\mathbf{X}_t$ and new measurements $\mathbf{Z}_{t+1}$. Combining the formalisms of equation (3.8) and of the Bayesian filtering framework, the optimal assignment (represented by the binary assignment matrix A_*) maximizes the sum of individual

linking likelihoods of assignment over both all states and all new measurements

$$\begin{aligned} A_* &= \underset{A}{\operatorname{argmin}} \sum_{i=1} \sum_{j=1} -\log p\left(\mathbf{z}_{t+1}^j \middle| \mathbf{x}_t^i\right) a_{ij}, \\ \text{s.t} &\sum_{i=1} a_{ij} = 1, \sum_{j=1} a_{ij} = 1, \end{aligned} \tag{4.14}$$

which allows expressing j^* for a given state i as $j^* = \operatorname{argmax}_j a_{*,ij}$.

Thus, Bayesian filtering generalizes equation (3.8) in the context of arbitrary motion models.

4.2.5.2 Adaptive gating to prune assignment hypotheses

Managing large amount of objects requires somehow pruning the set of linking hypotheses to consider only the most likely ones. For this end, a global gating radius $\text{SR}^{\text{glob.}}$, over which associations are deemed impossible, can be defined to reduce false positive links. A major improvement in the global gating approach comes from the possibility to learn the per-track general error covariance matrix S_t^i on-the-fly by exploiting the innovation defined above. u-track3D estimates S_t^i by computing the variance of the set of past prediction errors

$$S_t^i = Var\left(\left\{K_s^i y_s^i\right\}_{s\in[1:t]}\right). \tag{4.15}$$

The standard deviation matrix associated to the projection of S_t^i in the measurement space gives the typical scales at which the prediction deviates from the true state and so it can naturally be used to define an adaptive search radius SR_t^i that effectively changes according to the past history of an object's dynamics. u-track3D sets

$$\text{SR}_t^i = \lambda\sqrt{H^T S_t^i H}, \tag{4.16}$$

with λ a proportionality constant based on studies from [185].

4.3 Trackability: probing local tracking ambiguity to measure global performance

Automatic validation methods fall into two main categories: ground-truth-based approaches and error-inference-based approaches.

In the first category, a widely accessible yet time-consuming method involves using manual annotations to create a ground truth for comparison with the measured dataset. [139, 172] Even in 2D, these techniques are prone to selection bias, as annotating bright, well-separated objects is easier. In 3D, the complexity of visualization and particle selection further exacerbates these challenges. Recent advances in joint visualization and annotation

methods, e.g *Mamut*, integrated in *BigDataViewer* on *Fiji* or *syGlass*, a software for visualization of large datasets in virtual reality, help address this challenge, but still cannot cope with complex dynamics in dense environments. An alternative ground-truth-based approach involves simulating image sequences that replicate the acquisition process.[28, 178, 186] Unlike manual annotation, simulation enables automated predictions of tracking performance in simplified scenarios where image formation and particle dynamics are well-known. Advanced, well-grounded methods have been developed in biophysical studies of 3D aggregates to study large scale polarization [187], handle multiple tissue types and their interactions [188]. They provide good starting points to generate organization and dynamics that are reminiscent of gastruloid conditions [28].

The second category of methods seeks to infer tracking error likelihoods directly from the data. Early approaches used heuristics that combined motion and density measurements to identify error-prone regions. [150, 189] A more refined strategy involves directly analyzing the optimality of reconstructed trajectories. Cardinale et al. [190] proposed this concept and applied it to estimating the scale, intensity, and position of the spindle pole body, along with associated confidence intervals. While these confidence intervals capture uncertainty in position and velocity, they do not assess the correctness of trajectory-to-measurement associations. More recently, the software u-track3D [75] implemented a per-link score of association quality without the need for ground-truth tracks. By leveraging a Bayesian tracking framework, the score, called trackability, was applied to single molecule tracking and allowed detection and inspection of tracking error, automated selection of reliable tracks for analysis, and was also shown to provide a faitfhful readout of the Linking F1 score (see Appendix 6.5) when averaged over the entire tracking graph. Another study on intestinal organoids demonstrated that a similar concept could be introduced outside of the Bayesian tracking framework by training a deep neural network to predict association likelihoods from image pairs. As intestinal organoids are quasi-2D and shallow structures, the authors were able to gather sufficient annotated tracks to train the neural network and to learn a calibration mapping [191] from the network output to real, interpretable likelihoods. However, applying a similar strategy to deep gastruloids seems unreliable due to the lack of tracking ground-truth.

4.3.1 Trackability as an estimate for multiple hypothesis likelihood

The quality of the frame-to-frame assignment depends on the ability of the physical priors (through the measurement and dynamics models) to accurately represent the data. As a perfect model is often unavailable, a certain level of ambiguity persists in the data, which can lead to errors in the assignment. To quantify ambiguity within the optimal assignment, we built upon previous works that define an estimator of the probability of each individual assignment hypothesis (the quantity $p\left(\eta^{i,j}|\mathbf{x_t}\right)$ in equation (4.12)) called trackability. The estimator is computed by evaluating the prevalence of the optimal hypotheses in new assignment trials under Monte Carlo resampling of the particles' states $\mathbf{x_t} \sim p\left(\mathbf{x_t}|\mathbf{Z_{1:t-1}}\right)$. The approach is illustrated on Fig. 4.2.

Formally, we denote $T^{i,j}$ the trackability score for the given hypotheses $\eta^{i,j}$, and express

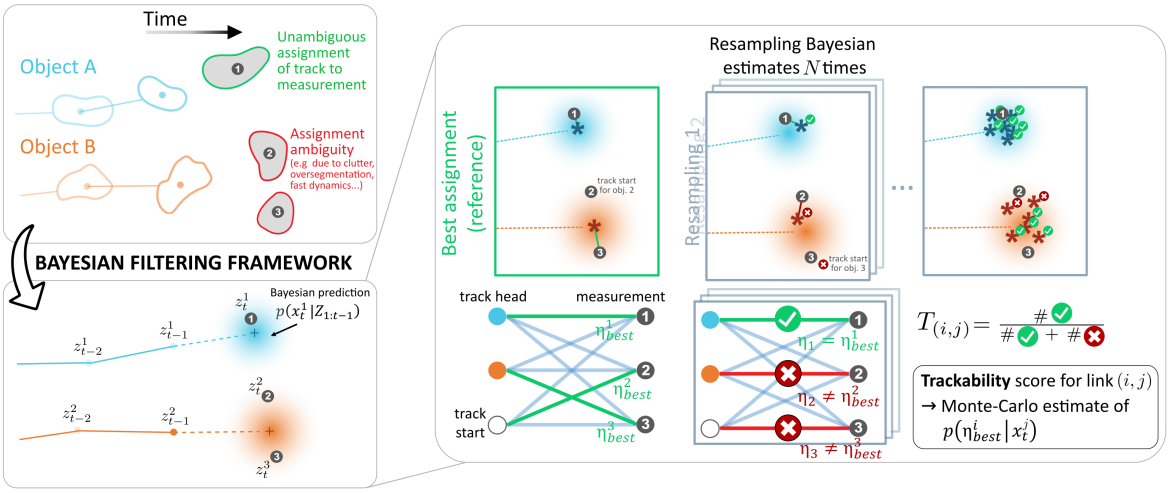

Figure 4.2: **Trackability as a Monte Carlo estimate of assignment stability in a Bayesian filtering framework.**

$T^{i,j}$ as

$$T^{i,j} = \frac{1}{N_{\mathrm{MC}}} \sum_{n=1}^{N_{\mathrm{MC}}} \left[\eta^{i,j} = \eta_n^{i,J}\right] \underset{N_{\mathrm{MC}} \to \infty}{\sim} p\left(\eta^{i,j} | \mathbf{x_t}\right), \tag{4.17}$$

where we perform N_{MC} Monte Carlo resamplings of the particles' states, $\eta_n^{i,J}$ is the newly computed assignment $i \to J$ during the n-th resampling, and $[\cdot]$ are the Iverson brackets, which equal 1 when the proposition evaluated is true, and 0 otherwise. As such, the local trackability score is defined as the ratio between the number of times a link has remained unchanged after resampling and the total number of resampling steps. A lower value of the score $T^{i,j}$ reflects a larger instability in the optimal assignment.

In our implementation, the assignment step is done with linear bipartite matching, and thus the hypothesis $\eta^{i,j}$ is uniquely associated to the value of its associated coefficient a_{ij}. This means that the term in the Iverson brackets in (4.17) is evaluated as $[a_{ij} = a_{iJ}^n]$ in practice.

The order of magnitude of the time spent solving linear assignment problems (without computing the trackability score) is around ∼50% for a typical tracking task in which the search radii can be kept reasonably small (which reduces the number of possible links allowed in the assignment problem). As computing the trackability score involves repeating the assignment step N_{MC} times (we choose $N_{\mathrm{MC}} = 20$ except specified otherwise), the time spent solving the assignment step represents ∼95% of the total compute time (the ratio of compute times with and without computing trackability effectively scales as $\mathcal{O}\left(N_{\mathrm{MC}}\right)$).

4.3.2 Trackability is a versatile measure

The trackability score is particularly valuable in 3D dense bioimages, where the complexity of dynamic and crowded environments hinders accurate validation. By quantifying ambiguities in particle tracking, it captures variations in tracking quality due to spatial density,

particle motion, and imaging parameters. This makes it a versatile tool for assessing performance and optimizing both imaging acquisition and analysis.

A key strength of the trackability score lies in its multiscale nature, allowing aggregation across different substructures of the tracking graph to convey distinct information. At the scale of a single link (where track starts and ends are also treated as links with dummy variables in the extended LAP framework), it estimates the probability of the corresponding hypothesis, measuring linking ambiguity and the stability of the optimal assignment (Fig. 4.3(a-d)). This quantity extends naturally to larger tracking substructures through averaging.

Averaging the score over all links in a trajectory facilitates visual validation by distinguishing high-quality (high trackability) from low-quality (low trackability) trajectories (Fig. 4.3(e-g)). When averaged per frame, trackability acts as a detector of abrupt dynamical events, which manifest as dips in the frame-level score due to increased ambiguity (Fig. 4.3(h-i)). At the movie scale, it serves as a robust predictor of tracking quality, shown to perfectly correlate with the Jaccard index over links (see Appendix 6.5) in simulations with varying particle densities, velocities, and motion heterogeneity (Fig. 4.3(j-k)).

4.4 Limitations of this framework for direct cell lineage reconstruction in gastruloids

The framework we have built in this chapter quantifies objects' individual aleatoric and epistemic uncertainties by learning their dynamical parameters during a trajectory reconstruction task. However, this basic framework is bound to suffer from a lack of robustness to the main challenges of cell lineage reconstruction in gastruloids.

4.4.1 Robustness to dynamical heterogeneity

The algorithm cannot handle heterogeneous cell motion, large movement, and sudden changes in object morphology. To address this problem, we implemented the PMMS algorithm [168], which allows modeling sudden changes in dynamical behavior by running as many KFs in parallel for each individual object as there are expected behavior patterns, and choosing the most appropriate one for the current data while leveraging past and future measurements. We refer the reader to the Appendix 6.5 for further technical details about the method.

4.4.2 Robustness to segmentation uncertainty and cell lineage constraints enforcement

After the initial linking step, traditional frame-to-frame Bayesian tracking algorithms [73, 74, 167] recover mostly tracklets, portions of actual tracjectories, even with advanced mo-

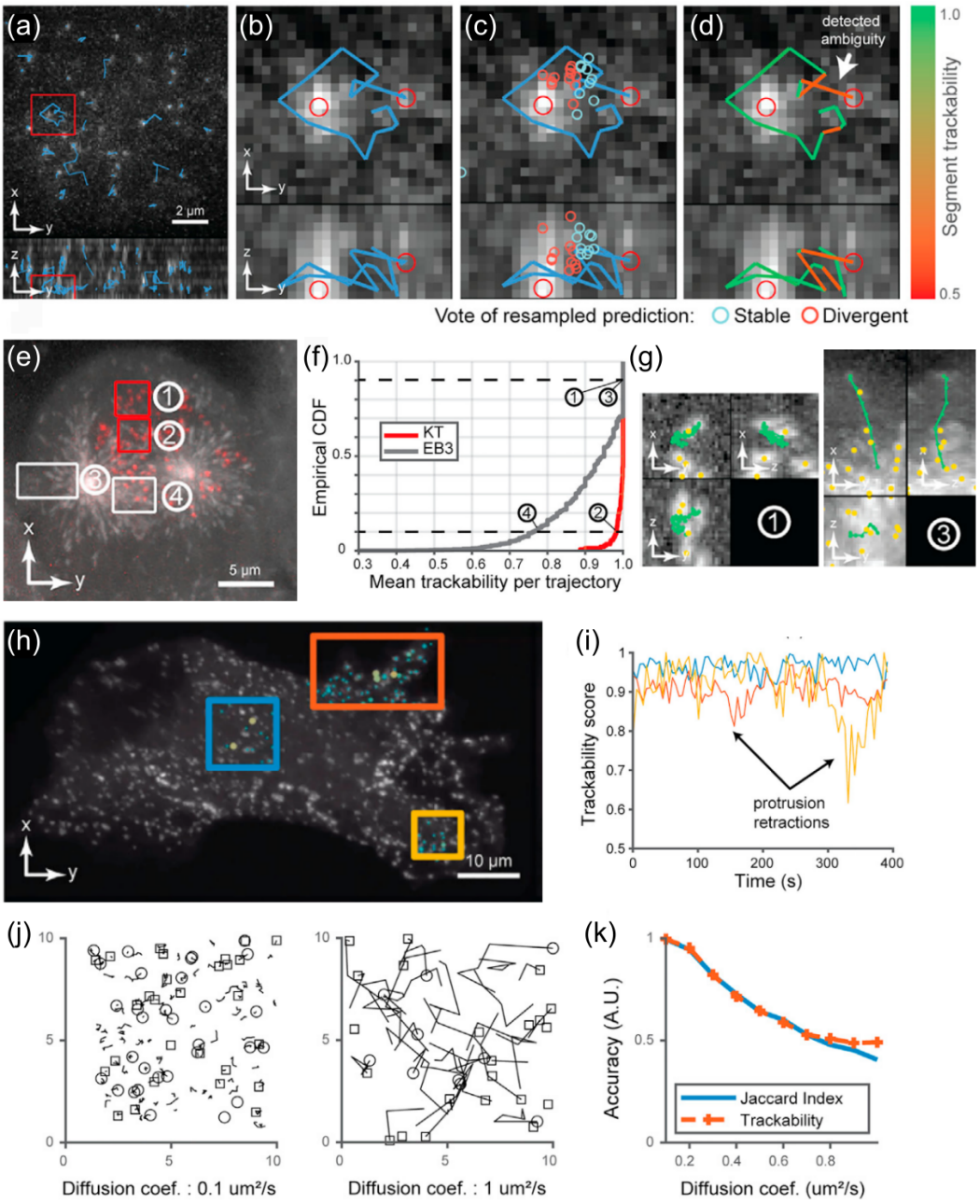

Figure 4.3: **Roudot et al. [75] showed that trackability is a multiscale mesure of tracking ambiguity and quality. (a-d)** Trackability defines link-scale ambiguity in track-to-measurement association. **(e-g)** Averaged over the links of a trajectory, trackability filters robust tracks from erroneous ones. Here, high and low thresholds on the trackability score cumulated distribution function were set (see (f)) to detect high- and low-confidence tracks **(h-i)** Averaged at the scale of a frame, trackability detects local sudden motion. In the render (h), ROIs are placed in the bulk of a breast cancer cell and close to protrusive parts of the cell membrane. In (i), rapid protrusion events lead to dip in the trackability score. **(j-k)** In simulations (j), aggregating trackability over a full movie perfectly estimates a supervised metric for tracking accuracy (k). Figures modified from [75].

tion models [168]. This can be due to aleatoric uncertainty, e.g when the objects stochasticity prevents perfect modeling of their dynamics (for instance blinking objects as often appear in molecule tracking scenarios), or epistemic uncertainty, e.g if the object detections/segmentations contain errors which may cause gaps or early track deaths or if lineage constraints are not enforced during the linking step (for instance when divisions are not modeled, which may cause a track end or an ambiguous single mother/daughter link).

To solve these issues, a separate post-processing procedure called gap-closing is generally used, in which an optimization problem involving knowledge about lineage constraints is solved globally both in space and time. Some methods [73, 75, 168] use hand-crafted heuristics to inject biological knowledge into their gap-closing procedure, while trackers designed for situations where the objects dynamics is precisely modeled via a probabilistic approach can directly use well-grounded hypothesis likelihoods [74]. While designing our Bayesian filter, we have not focused on implementing gap-closing, as our goal was neither to build a Bayesian cell tracker, nor to be tied to the choice of heuristics or advanced modeling priors for lineage constraints.

4.4.3 Calibrating the required priors

Second, we lack a strategy to efficiently tune the parameters introduced by the priors. While certain parameters, such as search radii derived from covariance matrices, can be visualized and adjusted heuristically, the overall behavior of the tracker can remain unpredictable. Moreover, adaptive strategies are paramount to handle heterogeneous datasets with varying cell motilities and densities. This sensitivity complicates the generalization of Bayesian tracking models across different experimental conditions and necessitates extensive empirical optimization (often visual and labor-intensive) or automated parameter-space searches based on ground-truth tracks, which are unavailable in our case.

4.5 Conclusion

Bayesian methods have been presented as a powerful framework to inject physical and biological priors in cell tracking algorithms. We have presented our implementation of a simple, modular Bayesian tracking algorithm, which incorporates the computation of the trackability score, which gives a link-scale estimation of assignment ambiguity and was previously established as a powerful, unsupervised, and multiscale estimator of trajectory ambiguity and quality. Its usefulness to validate tracking and automatically explore complex datasets raises interest in applying it to alleviate characteristic challenges of gastruloid analysis.

Our Bayesian tracking algorithm still suffers from a lack of robustness to (i) detection/segmentation uncertainty, (ii) global enforcement of lineage constraints, and (iii) the need for automatic parameter calibration methods. In this chapter, we established trackability as a versatile Bayesian metric capable of quantifying uncertainty across various scales, greatly improving the interpretability and reliability of tracking outcomes in challenging

cellular environments. Building on this foundation, the subsequent chapters explore practical applications of trackability, demonstrating how it enables unsupervised optimization and algorithmic benchmarking, thus directly addressing the methodological gaps identified earlier.

Chapter 5

Application: Trackability allows unsupervised ranking of segmentation algorithm with minimal priors

5.1 Chapter overview

This chapter extends the application of the trackability metric to evaluate and rank different segmentation algorithms in an unbiased, automated manner. We validate the robustness of this ranking approach through extensive simulations of cellular motion and varied detection errors, demonstrating strong coherence with standard performance metrics such as the F1 score. Additionally, we apply this method to experimental datasets, showcasing its sensitivity and adaptability to diverse cellular environments.

This chapter is part of the publication [97]:
VANARET, Jules, DUPUIS, Victoria, LENNE, Pierre-François, et al. A detector-independent quality score for cell segmentation without ground truth in 3D live fluorescence microscopy. IEEE Journal of Selected Topics in Quantum Electronics, 2023, vol. 29, no 4: Biophotonics, p. 1-12.

My contributions:

1. created an unsupervised metric to rank detection/segmentation algorithms based on the effect of segmentation error on track temporal coherence
2. designed and implemented a modular particle simulation framework, which can mimic key elements of nuclear dynamics in dense organoids
3. validated the metric using simulations generated with the simulator, incorporating added detection errors
4. validated the method on benchmark and sparse gastruloid datasets

5.2 Motivations

Building upon the success of trackability in precisely quantifying trajectory reconstruction accuracy, we will show that it can also be used to assess detection/segmentation quality. In what follows, we propose a novel score estimator to evaluate the accuracy of cell segmentation tools in live cell fluorescence microscopy imaging. In a nutshell, our score exploits the variation in temporal consistency to predict detection errors while remaining fully independent from the type of detector used (Figure 5.1c). The idea behind the approach comes from the observation that segmentation errors provide a time-varying response that is different from the usual dynamics. For example a cell cluster can be mis-detected as a single cell, a cell mis-detected as two cells, and a false positive detected in the background. To measure those inconsistencies, we first infer the parameters of cellular dynamics for each trajectory hypothesis, then, we evaluate for each detection the stability of the optimal set of trajectory by combining discrete optimization and a statistical resampling of the trajectory-to-detection likelihood.

A few wide-ranging studies [77, 192] have used manually-annotated data to analyze the performances of multiple cell segmentation algorithms. The metrics used for comparison typically distinguishes between the ability to segment a cell, using a metric of the overlap between estimated and annotated cell mask, and the ability to detect a cell, using a measure of object instance matching accuracy. On the one hand, a study focusing on cell morphogenesis will need a precise matching of the measured contour with the ground truth. On the other hand, a study analyzing cell count, lineage and cycle will require high detection accuracy against missed detection, false positives, split and merge artifacts. Since the seminal challenge on cell detection and tracking algorithms [192], deep learning approaches have enabled a breakthrough in both cell segmentation and detection. In their detailed and unbiased comparison of the capacity of conventional and deep-learning-based approaches to measure biologically relevant metric, Caicedo and colleagues [77] have shown that deep learning approaches are indeed performing better overall. However, their results also show that benchmarking techniques remain more important than ever. For example, while the count of false negatives obtained with U-net [92] is almost twice lower than with conventional adaptive thresholding approaches, the amount of cells artificially split by the neural network is higher than a threshold-based approach and a random-forest-based classifier. They also show that the performances of deep neural networks can vary significantly from one network structure to another or depending on the amount of annotations available. Importantly, while the relative performance of a conventional approach with respect to a specific task can be intuitively interpreted from the parameters and features that are used, the best purpose of a neural network can only be done with comprehensive benchmarking of its breakpoint. As such, comparing detectors remains a critical task in the routine use of cell segmenter.

Whether it is for benchmarking or training, annotation in 2D images is a time consuming process that becomes even longer, challenging and bias-prone in datasets of larger dimension. First, each Z plane and time point must be annotated to match the degree of accuracy in the entire volume. For example, in [77], authors measure 50 hours to annotate 11500 objects in 2D static images and efficiently train the network. Reproducing this

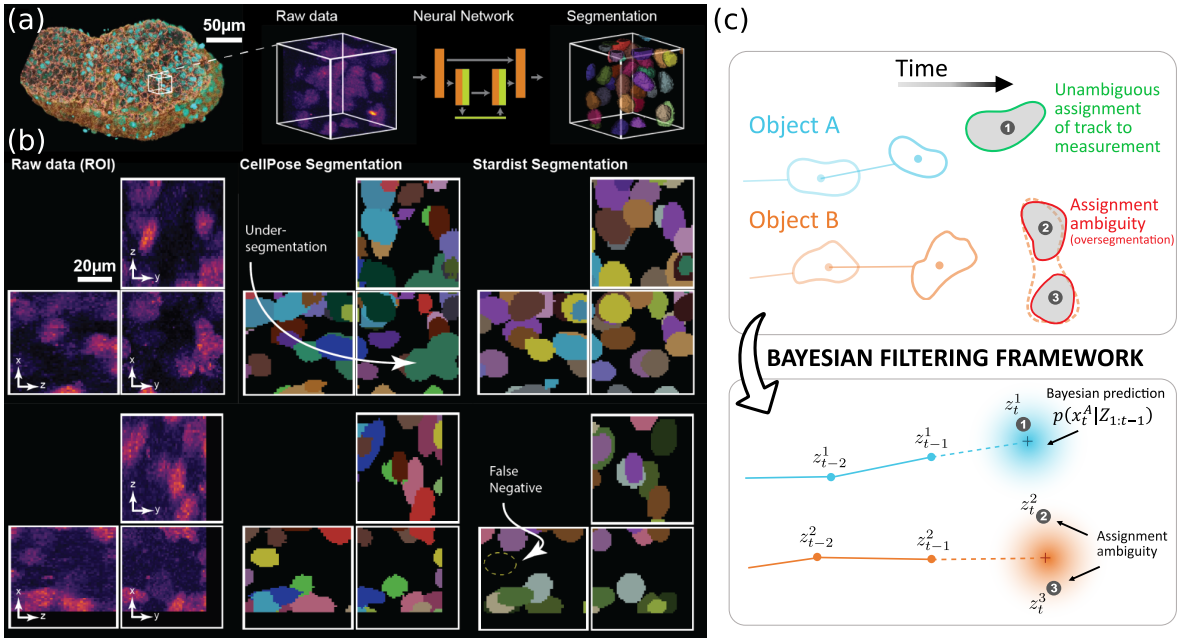

Figure 5.1: The density and heterogeneity of cellular objects challenge the comparison and validation of leading-edge segmentation approaches based on deep neural networks. **(a)** 3D volumetric rendering of a gastruloid imaged with two-photon microscopy along with a maximum intensity projection (MIP) rendering of a region of interest (ROI) and its associated segmentation. **(b)** Two ROIs rendered as orthogonal MIPs showing different artifacts ranging from the detection of nuclei clusters as a single nucleus (under-segmentation) to false negatives. **(c)** Our accuracy score uses temporal consistency to evaluate the likelihood of errors. Without segmentation errors (top trajectory), the detector provides a single measurement compatible with the cell dynamics. A segmentation error, here a split due to an oversegmentation (bottom trajectory), will result in two measurements with ambiguous associations to the trajectory.

process for a 3D dataset (with typically 5-10 2D Z-planes per object) would take up to 500 hours, even when ignoring the need to annotate all time points. Second, the 3D nature is difficult to interpret for the user. Correspondences must be made between Z-plane and time point, a non-trivial task when both images are not seen at the same time. 3D rendering is also prone to bias, as two different views may not result in the same visual interpretations [193]. To overcome those challenges, several papers have been focusing on improving annotation techniques for multidimensional datasets. In [194], authors propose to solve an inverse problem to recreate a 3D annotated volume from a collection of 2D annotations on maximum intensity projections. However, this approach is designed for sparse objects (e.g. vessels images) and poorly suited for the segmentation of the cluttered scene of cellular aggregates. In [195], authors experiment with the eye-tracking capacity of a virtual reality headset to annotate cell location and displacement from mere observation. While the tracking is seamless and robust to wavering attention by the annotator, this remains impractical for a large number of cells and ill-adapted to very dense scenarios. As such, there is to our knowledge no efficient solution to the annotation of 3D+time dataset, a challenge when comparing various approaches.

5.3 State of the art

Considering the ever-growing amount of data produced by next generation microscopy and its multidimensional nature, a few approaches have been developed for annotation-free quality control through the use of prior information on the object shape. First, goodness-of-fit analysis is routinely used to evaluate detection quality when the objects follow a simple template. For example, in a study on high-throughput super-resolution imaging, Beghin et al [196] use chi-square maps to control point-source detection across all the imaged wells. Since the cell intensity does not follow such a strict diffraction-limited model, those approaches cannot be used in our scenario. Fehri et al [197] proposed to alleviate this limitation with a graph-based criteria that characterize contour detection quality through a criteria on object intensity smoothness and object-to-background transition. More recently, Audelan et al [198] proposed a probabilistic approach score for the validation of generic segmentation tools based on the same intensity-base assumptions. However, those quality scores are only designed to evaluate the precision of the segmented contour as opposed to the accuracy of cell detection, a fundamental aspect of quantitative cell biology. More importantly, all those approaches make assumptions on object appearances that may not hold in experimental acquisition and especially in the challenging case of interest: low signal-to-noise ratio can corrupt intensity and blur the line between background and object, high scattering can modify this intensity across the object itself and low optical section typically makes the evaluation of contour quality more difficult in the axial orientation. To tackle those limitations, our contributions in the paper focus on a detector-independent quality score that makes no assumption on the object appearance, but rather on its dynamics. In particular, we focus on the capacity of this score to discriminate between false positive, negative, merging and splitting events.

5.4 Methods

5.4.1 Bayesian filtering based on hidden Markov models

Our approach aims at detecting segmentation errors through the dynamic footprint of objects' dynamics using the Bayesian filtering framework defined in the previous chapter. In this section, we first present the framework of Bayesian filtering to learn motion parameters, then we explain how we exploit ambiguities in local trajectories-to-measurement associations to detect potential errors.

One of the key challenges we face in inferring ambiguities is that a measurement with a lower trajectory-to-measurement likelihood does not necessarily mean that the segmentation is false (if the cell changes shape for example). In order to assess the presence of false positive, merging or splitting errors (leaving isolated false negative aside for now), we estimate the likelihood of association between a measurements and neighboring trajectories and test if the optimal solution is unique, i.e. if other local combinations have a significantly lower likelihood. In theory, this would require the filtering of every possible

sequence of measurements, which would grow exponentially even with efficient pruning of hypotheses. As such, similarly to MHT approaches, the comparison is carried out on a per-frame basis.

5.4.2 Search radius definition

Compared to the tracking framework proposed before, the definition of the gating radius must be modified. While tracking approaches use the motion prediction to accept or reject measurements into the optimization problem, we need to explore the measurement candidates in an area corresponding to possible measurement errors, including merging and splitting which may be larger than the area defined by motion only. Thus, we use a search radius SR that is both bigger than the prediction error d_{pred}, provided by the covariance of the innovation [168] and segmentation error d_{err}. Similarly, SR should be smaller than the average distance between particles d_0 to ensure that our score does not measure ambiguities from linking a particle to a track it does not belong to. As a result we chose to set SR to an intermediate scale given by the geometric mean of the upper bound d_0 and of whichever scale is largest between d_{pred} and d_{err}

$$SR \sim \sqrt{\max\left(d_{err}, d_{pred}\right) \cdot d_0}. \tag{5.1}$$

5.4.3 Simulation of cellular motion dynamics

To allow validation of our methods using ground-truth based metrics, we implemented a numerical model of interacting particle dynamics that we will tune to approach nuclear dynamics as observed in real organoid data. In particular, to convincingly mimic real nuclear dynamics and the challenges that come with it, our simulator should (i) prevent interpenetration between nuclei, (ii) incorporate heterogeneous motion, with a mix of caged and directed motion. Overall, this should promote a high probability of exchanging neighbors for each nucleus at each frame, as observed in real data.

We adopt and customize a model similar to previously published works, e.g [199] in which the model was used for the simulation of monolayered cell colonies which display liquid-like properties that can be precisely characterized by the model's parameters, or [28] which used it to study motility-driven unjamming during gastrulation in zebrafish.

5.4.3.1 Modeling individual cells and their interactions

In early embryonic development, particularly during gastrulation, cell nuclei have been observed to occupy a significant portion of the cell volume, often up to 60%. This high nuclear-to-cytoplasmic ratio makes the nucleus a reliable proxy for the overall cell dynamics. Moreover, the nuclei of these cells typically exhibit convex, approximately ellipsoidal shapes. Given these characteristics, modeling cells as point-like particles centered

at their nuclear centroids provides a reasonable approximation of their spatial distribution and movement while significantly reducing computational complexity.

The point-like representation is particularly suitable for capturing cell-cell interactions, overall tissue dynamics, and emergent patterns in developing embryos, where the precise details of cell shape may be less critical than their relative positions and movements. This framework has been successfully used to recreate many complex biological phenomena, like cell-sorting, convergent extension, or active cell jamming [28, 199].

The equation of motion, expressing the acceleration $\ddot{\mathbf{x}}^i$ for a particle i of mass m^i at position $\mathbf{x}^i$, reads

$$m^i \ddot{\mathbf{x}}^i = \mathbf{F}^i_{prop.} + \sum_{j \text{ neigh.}} \mathbf{F}^{ij}_{inter.} + \mathbf{F}^i_{drag}, \tag{5.2}$$

where the terms on the right-hand side are the forces exerted on particle i, namely propulsive forces, interaction with nearest-neighbor particles, and drag.

In cell aggregates, the dynamics is over-damped (friction dominates) which motivates us to neglect the inertia term in the equation of motion. Forces exerted on a cell then effectively balance at all times:

$$\mathbf{0} = \mathbf{F}^i_{drag} + \sum_{j \text{ neigh.}} \mathbf{F}^{ij}_{inter.} + \mathbf{F}^i_{prop.}, \tag{5.3}$$

Propulsion Cell propulsion emerges from the intricate coupling of mechanical and biological phenomena at multiple scales. At the molecular level, active stresses generated by cytoskeletal components like actin and myosin interact with cell-substrate adhesions to produce forces that drive motion. These internal forces are transmitted to the extracellular environment through mechanosensitive complexes such as integrins and cadherins, which can trigger biochemical signaling cascades in response to mechanical cues. The interplay between protrusive forces at the cell front, often driven by actin polymerization, and contractile forces at the rear, typically mediated by actomyosin contraction, creates an asymmetric stress distribution that propels the cell forward. As precisely incorporating all these phenomena would require implementing a joint simulation of genetic signals propagation and their interaction with cells, we opt instead to represent propulsive forces using stochastic variables.

The Langevin-type force introduced in equation (3.1) is the canonical choice, as it can appropriately yield diffusive trajectories, but cannot model directed trajectories characterized by high temporal correlation of the force direction for extended periods of time. For this matter, we employ an Ornstein-Uhlenbeck random force F_{OU}. Let $F^i_{k,OU}$ be the k-th component of the propulsion force term $\mathbf{F}^i_{OU}$ (e.g $k \in \{x, y, z\}$ in 3D). It satisfies

$$\mathrm{d}F^i_{k,OU}(t) = -\frac{F^i_{k,OU}(t)}{\tau^i_p}\mathrm{d}t + f_0 \mathrm{d}W^i(t), \tag{5.4}$$

with persistence time τ^i_p and $f_0 \mathrm{d}W^i(t)$ the differential of a Wiener process of variance $f_0^2 = 2D^i$, with D^i the effective diffusion coefficient of particle i. This leads to the magnitude

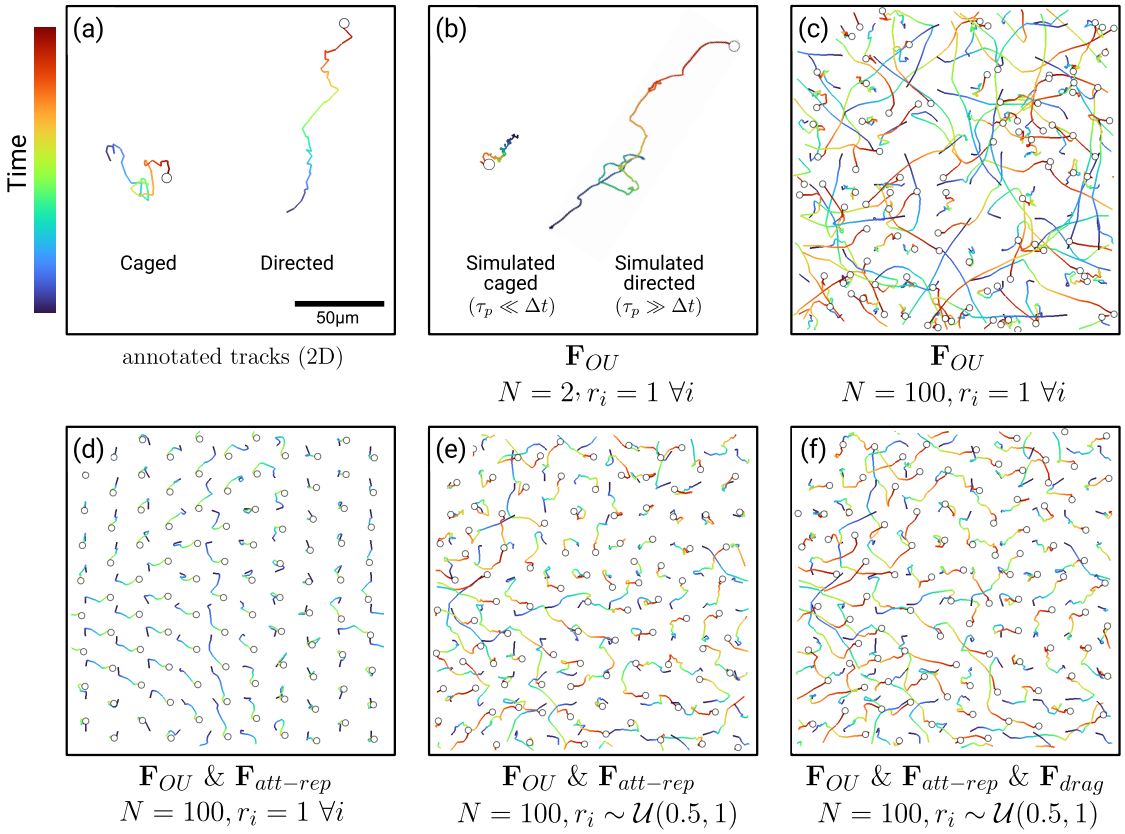

Figure 5.2: **Breakdown of each component of the simulation, and their effect on the simulated dynamics. (a-b)** Adjusting the persistence time τ_p relative to the simulation timestep Δt enables particles under an Ornstein-Uhlenbeck force to exhibit both caged and directed trajectories, reflecting behaviors observed in real data. **(c-d)** When increasing particle density, adding interaction forces prevents interpenetration and mimics tissue cohesion, though the stability of the hexagonal lattice configuration reduces dynamical heterogeneity. **(e)** Introducing slight variations in particle radii relaxes the lattice structure, enabling neighbor exchanges. **(f)** Adding a drag force locally couples particle dynamics, improving the representation of local correlations in real aggregates.

of the stochastic force having a variance of $\sigma^2 = 2dD^i\tau_p^i$, with d the number of spatial dimensions.

This can be interpreted as a classical Langevin force whose direction can vary randomly in all directions, but with added temporal persistence, allowing it to partially maintain its directionality between time points over timescales of order τ_p^i. Particles with low persistence times (compared to the observation timescale) move in a diffusive fashion reminiscent of Brownian motion (the Langevin force corresponds to the limit $\tau_p^i \to 0$), whereas cells with high persistence times can move in a quasi-directed manner during timescales of order τ_p^i, as illustrated on Fig. 5.2(a-b). The Ornstein-Uhlenbeck process can create a rich range of trajectories in aggregates by allowing several populations with different standard deviations and correlation timescales coexist. This approach effectively balances the persistent and random aspects of cell motility observed in various biological contexts, such as during embryogenesis and wound healing.

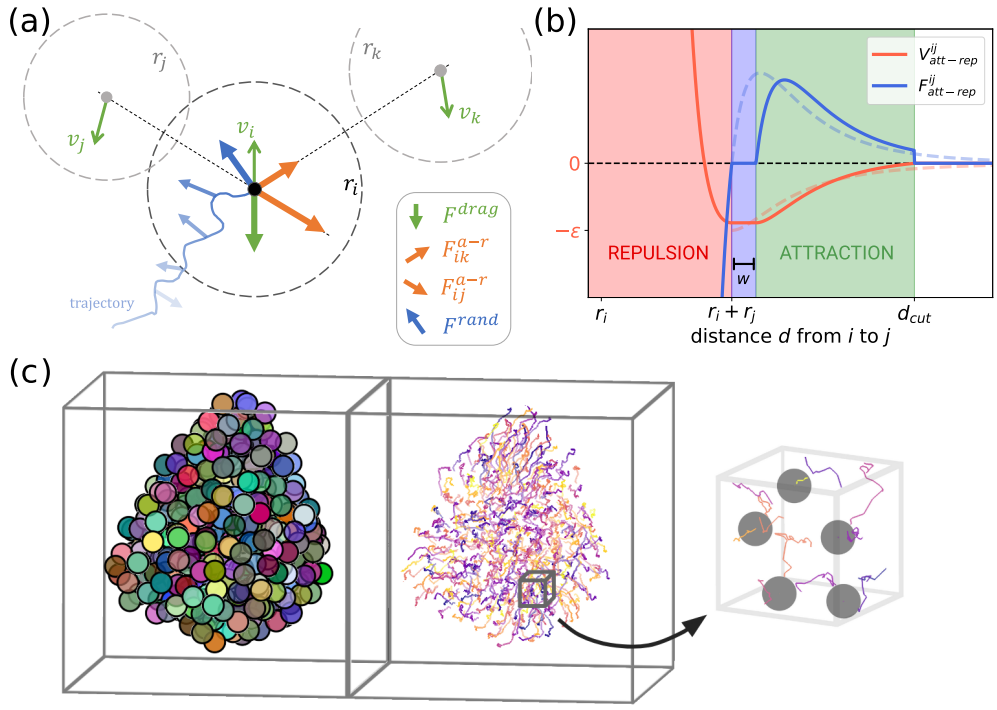

Figure 5.3: **A simple particle simulation framework allows mimicking typical nuclear dynamics observed in gastruloids. (a)** In the simulation, particles are subject to three forces: drag due to dissipation at cell-cell interfaces, attraction/repulsion due to non-interpenetrability of the membranes and nuclei, and stochastic Langevin-like forces creating active motion and heterogeneous dynamics. **(b)** The attraction/repulsion force derives from a typical Lennard-Jones potential (dashed lines) with an extended stability basin of width w and vanishes above a distance d_{cut} (full lines). **(c)** Simulated particles can form dense aggregates that display complex dynamics (left: sphere rendering, right: trajectories).

Attraction/repulsion to nearest neighbors Cell-cell interactions are modeled using a modified Lennard-Jones potential, which captures multiple aspects of cellular behavior. At short distances, the potential exhibits strong repulsion, preventing nuclei interpenetration and reflecting their incompressibility. As distance increases, an extended basin of stability is introduced, representing the deformability of cell membranes. The width w of this basin quantifies the range over which cells can maintain stable contacts, mimicking the flexibility of the actin cortex and associated adhesion molecules. This promotes the circulation of cells and neighbors exchanges. At medium ranges, the potential features an attractive component, simulating how cells actively pull on each other through mechanisms such as Cadherin-mediated adhesion. The strength and range of this attraction are parameterized by the depth and width of the potential's attractive basin, respectively. Beyond a cutoff distance, the attraction abruptly terminates, preventing unrealistic interactions between non-adjacent cells. Our modified Lennard-Jones potential and the force that derives from it are represented on Fig. 5.3(b). As shown on Fig. 5.2(c-d), adding heterogeneity in particles' radii greatly "fluidizes" the dynamics by allowing particles to not fall in a stable hexagonal lattice configuration. This comprehensive approach allows for biologically realistic modeling of cell spacing, intercellular forces, and tissue cohesion, while maintaining com-

putational efficiency.

Given two cells i and j with radii r^i and r^j, the force exerted by cell j at distance d of i reads

$$\mathbf{F}^{ij}_{inter.} = \mathbf{F}^{ij}_{att-rep} = \begin{cases} 12\varepsilon d^{-1}\left(\left(\frac{r^i+r^j}{d}\right)^6 - \left(\frac{r^i+r^j}{d}\right)^{12}\right) & \text{if } d \in [0, r^i + r^j] \\ 0 & \text{if } d \in [r^i + r^j, r^i + r^j + w] \\ 12\varepsilon d^{-1}\left(\left(\frac{r^i+r^j}{d-w}\right)^6 - \left(\frac{r^i+r^j}{d-w}\right)^{12}\right) & \text{if } d \in [r^i + r^j + w, d_{\text{cut}}] \\ 0 & \text{if } d > d_{\text{cut}}, \end{cases} \tag{5.5}$$

where ε is the characteristic energy associated with the interaction. When the first term in the equation dominates the second, the force is attractive, and the converse leads to repulsion. Importantly, the force vanishes at the distance $r_i + r_j$ between these two regimes.

Drag force The model incorporates a viscous drag force to account for the dissipative effects arising from cell-cell interactions and adhesion dynamics, as well as interaction with the medium. This force is crucial for capturing the complex rheological properties of cellular aggregates and tissues. The drag force is formulated in two terms based on Stokes' law: the first one is proportional to the difference in velocity between a cell and that of its nearest neighbors, reflecting the local nature of cell-cell friction, and the second term is proportional to the cell velocity

$$\mathbf{F}^i_{drag} = -\gamma^i\left(\dot{\mathbf{x}}^i - \langle\dot{\mathbf{x}}\rangle^{\aleph_i}\right) - \gamma^i_{med.}\dot{\mathbf{x}}^i, \tag{5.6}$$

where γ^i is an effective friction coefficient between neighboring cells, $\langle\dot{\mathbf{x}}\rangle^{\aleph_i}$ is the ensemble averaged velocity of the nearest neighbors $\aleph_i$ of the cell i, and $\gamma^i_{med.}$ is the friction coefficient with the medium.

Biophysically, the first term represents the energy dissipation occurring at cell-cell junctions due to the constant remodeling of adhesion complexes and the relative motion of adjacent cell membranes. The effective friction coefficient, γ^i, encapsulates various dissipative processes, including the viscous response of the cytoskeleton and the dynamics of transmembrane adhesion proteins. This formulation allows the model to capture emergent viscoelastic behaviors observed in cell aggregates, such as stress relaxation and creep, which are essential for understanding tissue mechanics during morphogenesis and homeostasis (Fig. 5.2(f)).

Finally, cohesive and fluid-like behavior of modeled cell aggregates emerges naturally from the interplay between adhesion/repulsion, propulsion, and friction.

5.4.3.2 Final equation of motion

The combination of all forces is represented on Figure 5.3. Neglecting the inertia term due to the overdamped-regime assumption leads to a first order stochastic differential equation in the positions of the particles

$$\dot{\mathbf{x}}^i = \alpha \left\langle \dot{\mathbf{x}} \right\rangle^{\aleph_i} + \frac{1}{\gamma^i + \gamma^i_{med.}} \left(\sum_{j \in \aleph_i} \mathbf{F}^{ij}_{att-rep} + \mathbf{F}^i_{OU} \right), \tag{5.7}$$

with $\alpha = \gamma^i / \left(\gamma^i + \gamma^i_{med.}\right)$.

5.4.3.3 Implementation

The term $\langle \dot{\mathbf{x}} \rangle_{\aleph_i}$ is computed at t using the velocities at the previous time point $t - \Delta t$ to keep the system of equations separable. This would require us to choose a very low value for Δt, so we use a stochastic version of Heun's integration scheme [200, 201] (of order 2) to compensate for this and keep a reasonable value of the time step. The inclusion of the parameter $\alpha < 1$ in 5.7 also serves to stabilize the equation. To compute the attraction-repulsion forces, we build a k-d tree [202] with the positions of the particles at each time point, which is a data structure that allows for fast lookup of nearest-neighbor relationships and computation of pair-distances. Because these forces vanish beyond a certain distance, the pair-distance matrix is highly-sparse and thus can be computed efficiently. The simulator is implemented in *Python3* and multiprocessing is used with *Numba* to take advantage of the separability of the system of equations at each time point.

Our implementation supports setting stopping boundary conditions (i.e if an object goes beyond the boundaries its position is set to the boundary), which are useful to form dense "*in toto* aggregates (see Fig. 5.3(c)), and periodic boundary conditions (i.e simulation on a torus). The latter is more stable than the former and is thus preferred to study problems linked to particles going in and out of the field of view while keeping the number of objects fixed, and to prevent isolated clusters of cells from forming when the dynamics is highly heterogeneous. It is implemented using modulo arithmetic (see Appendix 6.5).

5.4.4 Simulation of point-like segmentation errors

We produce typical segmentation errors like false negatives (FN), false positives (FP), splits, and merges by corrupting the positions of the simulated particles at each frame, as illustrated on Figure 5.4. To add random FN errors at a given rate ρ_{FN} from a simulation with N particles, we draw $\rho_{FN} N$ particles in a uniform manner (every random choice in this section is done following a uniform distribution) at each frame and remove them. For FP errors at a given rate ρ_{FP}, we choose $\rho_{FP} N$ positions uniformly inside the volume of the particles at which we add fake particles. For merges at a rate ρ_{me}, we compute at each frame all pairs of particles that are immediate neighbors from one another, we take $\rho_{me} N$ pairs, remove the associated particles and add fake particles at their barycenter. For splits

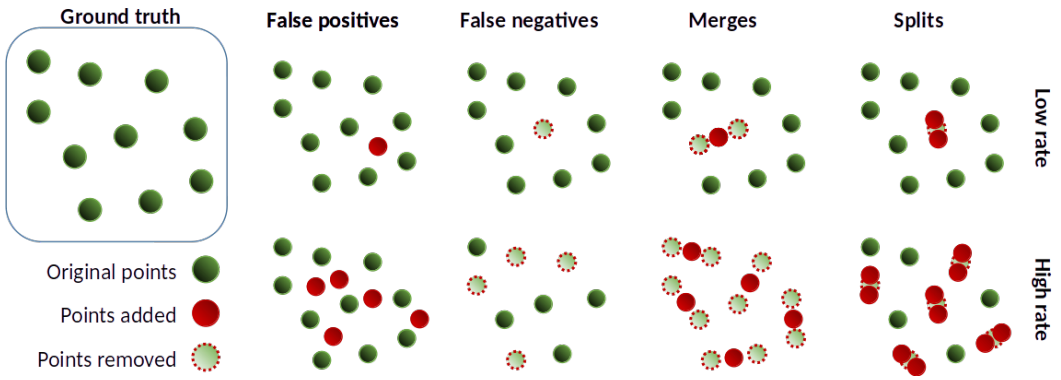

Figure 5.4: **Illustration of the different types of errors introduced in the simulations.** Each error type is shown in a low and a high error rate scenario.

at a rate ρ_{sp}, we choose $\rho_{sp}N$ particles, and for each one, with radius r_i at position x_i, we choose a random 3D vector u of magnitude $r_i/2$. We then remove the particle and add two fake particles at positions $x_i \pm u$.

5.5 Results

5.5.1 Validation using particle simulations

We simulated $N = 400$ particles for a duration $T = 50\Delta t$, with Δt the integration time step, chosen so that the average displacement between frames matches what we observe in our data (illustrated on Figure 5.5). Particle positions were initialized by choosing N random points uniformly in a sphere whose radius is such that the density inside the sphere is below that of random sphere packing. This leads to particles expanding outwards before the aggregate settles to its equilibrium radius, after which particles are on average at distance d_0 from their nearest neighbors (Figure 5.5a). The expansion phase of the dynamics was removed in our experiments. Particles are simulated with infinite lifetimes. To compute our score, we set the search radius SR using our heuristic on the length scales of the different phenomena. Here, the average displacement $d_{dyn} \sim d_0/10$, with d_0 the average nearest-neighbor distance. By design, merge and splits errors produce an average displacements of $d_0/2$ and $d_0/4$ respectively. We thus set

$$SR = \sqrt{(d_0/2) \cdot d_0} \sim 0.7 d_0. \tag{5.8}$$

5.5.1.1 In simulations with a single type of detection error, our score is close to the F1 score.

We first study the behavior of the score in conditions corresponding to adding a single error type and varying its rate. While this is an unrealistic scenario that may not reflect the behavior of the score in experimental conditions, this approach allows us to unbiasedly test

the strength and weaknesses of our method. The results are shown on Figure 5.6a, where we plot the evolution of our score along the centroid-based F1 score $\mathrm{F1}_c\,(\tau_c)$ against error rates. For merge and split conditions, the ground-truth $\mathrm{F1}_c\,(\tau_c)$ is plotted for the complete range of possible values for τ_c, showing that the true performance heavily varies depending on the choice of the matching threshold. The darker $\mathrm{F1}_c$ curve in the middle is obtained for a value τ_c close to the typical scale of the respective events. The dimmer curves correspond to the extremal curves of $\mathrm{F1}_c\,(\tau_c)$ when $\tau_c \to 0$ and $\tau_c \to \infty$. FN and FP events do not have a typical scale associated with them since they will always be detected in those uniform scenarios.

We show that for merges, splits and FP errors, our score follows the trend of the true F1 score, both in its monotonicity and in qualitative proximity to its value. For FP errors, we notice that our score is weakly responsive for rates below 15%. This is coherent with our previous works [150] where tracking performances under a varying rate of FP exhibit a similar regime transition with an inflexion point that decreases as object velocity increase. This suggests that FP detection performs better when the density of errors starts to significantly impact tracking. For FN errors, the score shows little sensitivity. This is unsurprising since an isolated false negative will not be creating dynamic instability. As such, while we do not have an exact match with a given accuracy score, our score is capable of ranking those simulated detection approaches without ground-truth in three scenarios.

5.5.1.2 In simulations with mixed error types, our score varies coherently with respect to the F1 score.

To study conditions closer to realistic cases, where multiple type of error arise jointly, we plot on Figure 5.6b the evolutions of our score and of $\mathrm{F1}_c$ for different rates of both FN and FP errors. When corrupting point positions, we first added FN errors before FP to ensure that no FP error added would be canceled by a FN error. We use $\tau_c = SR$ to compute $\mathrm{F1}_c$. We observe that the F1 score decreases significantly with the increase of the FN and FP rates, which is consistent with an increase in the number of errors. Again, we observe that our score is less sensitive to FN than FP errors, as it significantly decreases when increasing the FP rate at a given FN rate. On the other hand, Figure 5.6c shows that fixing the FP rate and varying the FN rate yields small variations of the score.

For small or intermediate values of the FP rate, Spearman's coefficient (defined as the Pearson correlation coefficient applied to the set of rank of each values, instead of the values themselves) between the score and true F1 indicates that the ranking property is conserved, but at high FP rates (more than 40%), the score increases when the F1 score decreases (negative Spearman's coefficient). Our score reaches its descriptive limit in this extreme scenario because no trajectories can be formed, even in-between a few time points. As such, no ambiguities in trajectory-to-detections can be measured. Considering the large percentage of FP and FN required to reach its breakpoint, we are confident that our score can properly rank detectors, especially since such a large rate of false negatives or false positives can be detected by mere visual inspection.

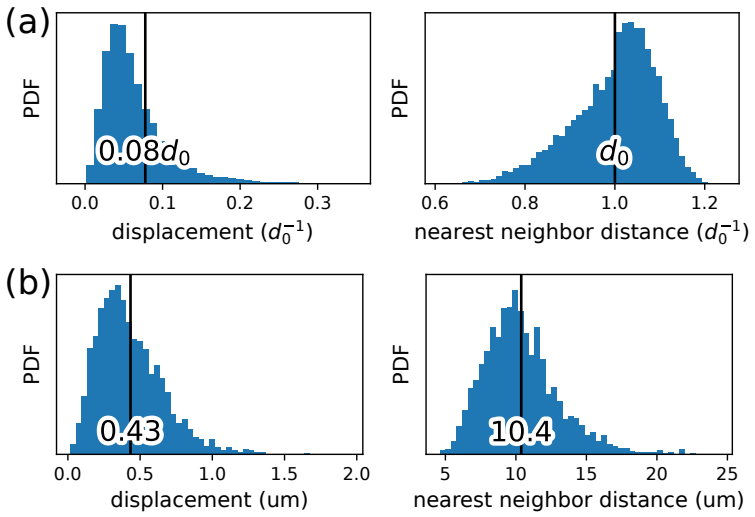

Figure 5.5: Distribution of frame-to-frame displacement and nearest neighbor distance in our simulation **(a)** and in real experimental data **(b)**. Vertical black lines are centered and annotated with the average value of the distributions.

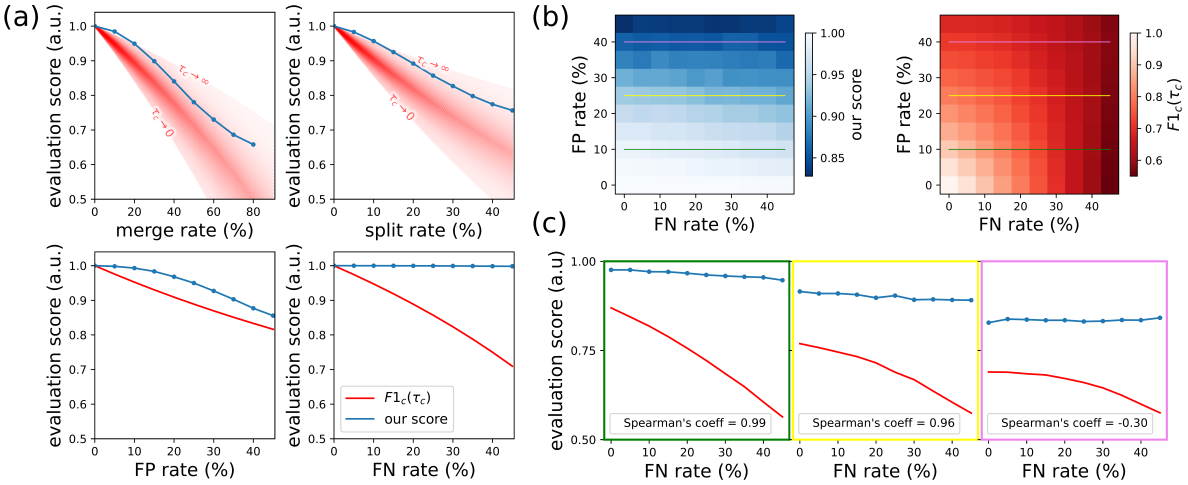

Figure 5.6: Our evaluation score responds strongly to simulated merging, splitting and false positive errors and is sensitive to false negatives in scenarios with mixed error types. **(a)** When a single error type is introduced in simulated data, our score is well correlated with the F1 score except for false negatives, for which it is unresponsive by design. For merge and split conditions, the threshold distance is varied to represent the possible range of F1 score profiles. **(b)** In mixed FN/FP conditions, our score shows a higher sensitivity to false positives than false negatives, and **(c)** for high density of false positives, its ranking property can fail. Plots in (c) display the range of values covered by the colored lines in (b).

5.5.2 Application to experimental data

We demonstrate the usefulness of our approach on two live volumetric imaging datasets. The first one shows fluorescent nuclei of mouse embryonic stem cells forming an organoid which we imaged with two-photon microscopy, and is our primary dataset for the ap-

plication of our works. It exhibits high cellular density and heterogeneities in both cells dynamics and appearance. The second one is a reference dataset from the Cell Tracking Challenge [192] and shows human breast carcinoma cells embedded in collagen. It exhibits sparser cellular density but higher heterogeneities in both cells dynamics and appearance, with notably elongated cells.

5.5.2.1 Our score effectively ranks Cellpose over StarDist3D on a Carcinoma cells dataset.

The carcinoma cells dataset shows MDA231 human breast carcinoma cells infected with a pMSCV vector including the GFP sequence, embedded in a collagen matrix. For further informations, we refer the reader to the Cell Tracking Challenge [192]. This dataset was chosen as a complementary to the organoid cells dataset, as it is much sparser but with higher shape heterogeneity. A frame is illustrated on Figure 5.7(a).

Stardist3D [91] and Cellpose [95] are two open-source state of the art segmentation tools based on neural networks. Stardist3D is presented in greater details in 1.3.2.1. Stardist3D is usually more appropriate for ellipsoidal objects in dense environments, and Cellpose often performs better for complex shapes and intensity heterogeneities [203]. We tested the pre-trained Stardist3D model and two pre-trained Cellpose models (models "*cyto*" and "*cyto2*"), which are compared to the ground truth on Figure 5.7(b). Notably, the Stardist3D model does not handle elongated cells well due to their non-star-convex shapes and also likely due to a lack of similar training data. Thus, it produces many splits and false positives errors, while the two Cellpose models seem to produce accurate segmentations.

The centroid-based F1 metric of Figure 5.7(a) reflects the much better performances of the Cellpose models compared to Stardist3D, but is not enough to separate the former two. From the IoU-based F1 score, we see that Cellpose model *cyto* has improved segmentation performances compared to model *cyto2*, which itself is marginally superior to Stardist3D. We note that despite visually identifying Cellpose models as accurate on Figure 5.7(b), all models have rather mediocre IoU-based F1 scores. This is easily explained by visually comparing the ground truth and Cellpose segmentations (Figure 5.8(b)), the ground truth being limited to the cell body and excluding protrusions while CellPose is typically more precise. Overall, the *cyto* model performs better than *cyto2*.

Finally, we show the performances given by our score on Figure 5.8(b). The search radius SR was still chosen using the previous heuristic and the score is again averaged on all frames. We see that the score accurately ranks the three detectors, with a relatively low value for Stardist3D compared to the two Cellpose models. The latter are also correctly separated, with *cyto* ranking higher than *cyto2*.

5.5.2.2 Our score ranks StardDist3D over Cellpose for a sparse organoid dataset.

Here we show that our approach can be applied successfully to another dataset with different characteristics, in this case a sparse organoid dataset. To decrease the effective visible cell-density, we made the organoids from a mixture of two mouse embryonic cell lines, with

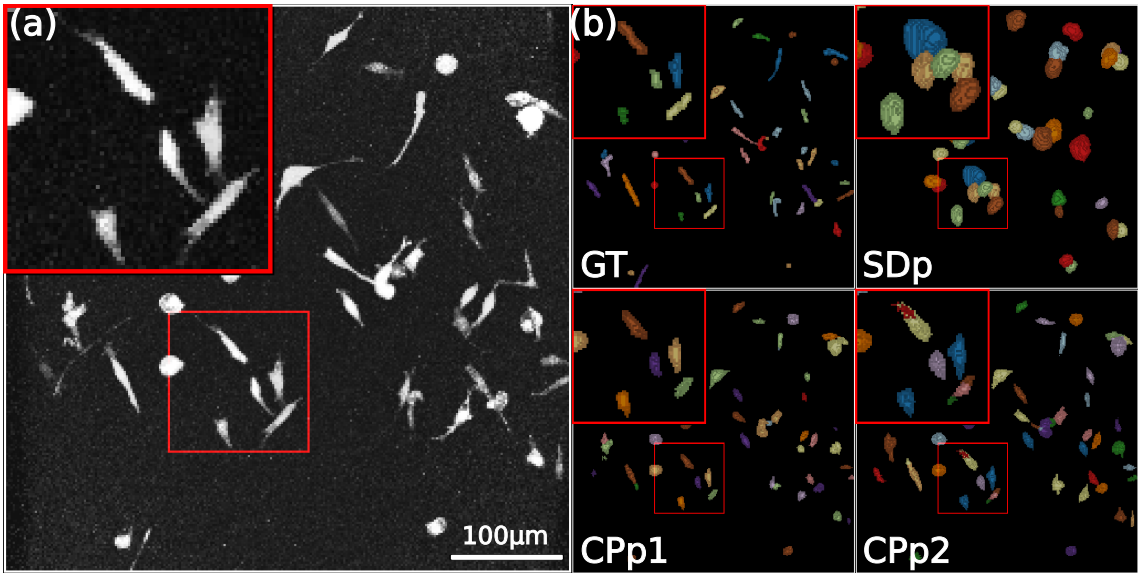

Figure 5.7: Visual example of typical Stardist3D and Cellpose response on 3D live carcinoma cells data. (a) Here, a single frame of the dataset is rendered as a maximum intensity projection along the Z axis. The dataset exhibits sparse cell density but with high heterogeneity in cellular dynamics and appearance. (b) Models tested (SDp is the same Stardist3D pre-trained model as in Figure 5.9 and CPp1 and CPp2 are the "*cyto*" and "*cyto2*" pre-trained models of Cellpose) are compared to the ground-truth segmentation (GT). The inset displays an illustrative ROI.

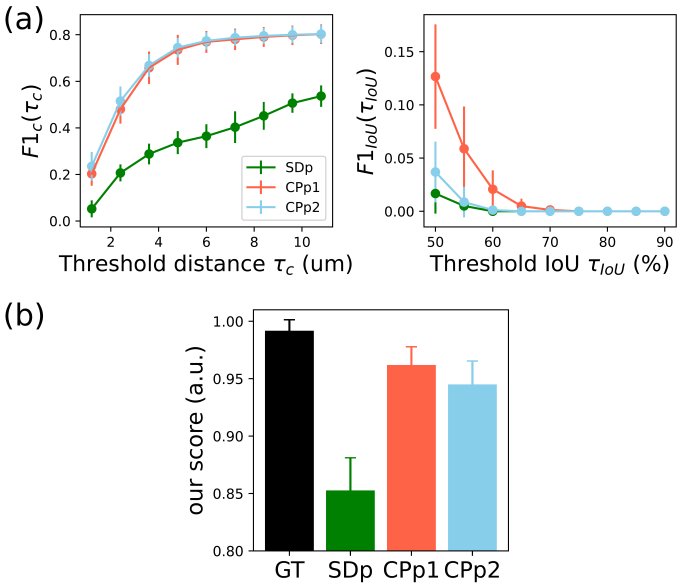

Figure 5.8: Our evaluation score can accurately rank all models studied on the carcinoma cells dataset. (a) We use the F1 scores based on centroid detection and on instance segmentation to rank the models using the ground-truth segmentation. (b) Our score provides a similar ranking.

initially 50% of non-fluorescent cells (cell line ES-E14TG2a) and 50% of cells endowed with a fluorescent reporter of the gene Brachyury (cell line T-Bra-GFP/NE-mKate2), which makes

the cell nucleus fluorescent when the gene is expressed. Biphoton imaging of organoids was performed in a chamber maintained at 37°C, 5% CO2 with a humidifier using a Zeiss 510 NLO (Inverse - LSM) with a femtosecond laser (Laser Mai Tai DeepSee HP, 900 nm) with a 40 x/1.2 C Apochromat objective. A z-stack acquisition of a 50 microns thickness is performed every 5 minutes and every 1 micron. Lateral pixel size is 0.62 microns. A volume from the movie is shown on Figure 5.9a.

Using a custom ground-truth segmentation of this dataset, we first compare Stardist3D and Cellpose using the F1 scores on centroid and on IoU as defined in the Metrics Section. For the sake of emulating a real-life use of segmentation tools in the context of biology, we tested off-the-shelf pre-trained neural networks models for each tool (several pre-trained models are made available by the Cellpose maintainers, we selected the "*nuclei*" pre-trained model). We also trained our own version using only the first frame of the movie, the manual segmentation of which is already cumbersome and takes approximately 6 hours using Napari [69] with a dedicated plugin. Examples of qualitative performances of all models can be observed on Figure 5.9b. Ground-truth-based metrics, shown on Figure 5.10a, are computed using all other frames. We noticed many FN errors due to objects cut at the frame borders in most models (particularly Stardist3D, Figure 5.9b). Since those errors are typically linked to implementation details rather than a fundamental limit of the network model, we sought to fairly compare the performances of Cellpose and Stardist3D using cropped predictions. For the centroid-based F1 score, we observe that a hypothetical ranking of the algorithms based on this metric would be independent from the choice of the threshold distance τ_c. Stardist3D performs better than Cellpose for pre-trained models, while the two reach similar values of $\mathrm{F1}_c$ above their pre-trained counterparts once trained. Secondly, the IoU-based F1 score provides a similar overall ranking, but with much better performance for the pre-trained Stardist3D model compared to the pre-trained Cellpose one for the lower values of the IoU threshold. Trained models are again close to each other for high levels of the IoU threshold, but Cellpose lacks in performance for lower values of IoU. In order to evaluate our score on this dataset, we compute our heuristic on the search radius SR by visually inferring d_{err} and d_0 on a few slices.

We then analyze the performance of our score applied to this dataset. Results are shown on Figure 5.10b, where we present the average scores (averaged on each frame) obtained with each segmentation tool along with the average score obtained by applying our method on the ground truth segmentation. The score obtained by the ground-truth is smaller than the theoretically maximal value 1 due to ambiguities induced not by errors in segmentation but by the dynamics itself (e.g. *bona fides* ambiguities due to large displacement combined with high density, or a real division event that would be interpreted as a split error). First, we observe correct overall ranking for the pre-trained versions of Cellpose and Stardist3D models. Our score predicts correctly the clear difference in performance measured in both $\mathrm{F1}_c$ and $\mathrm{F1}_{IoU}$. Second, the scores for the two trained models are close which reflect their proximity in centroid-based F1 score, but the overall ranking predicts that Stardist3D perform better, which is coherent with the difference in $\mathrm{F1}_{IoU}$.

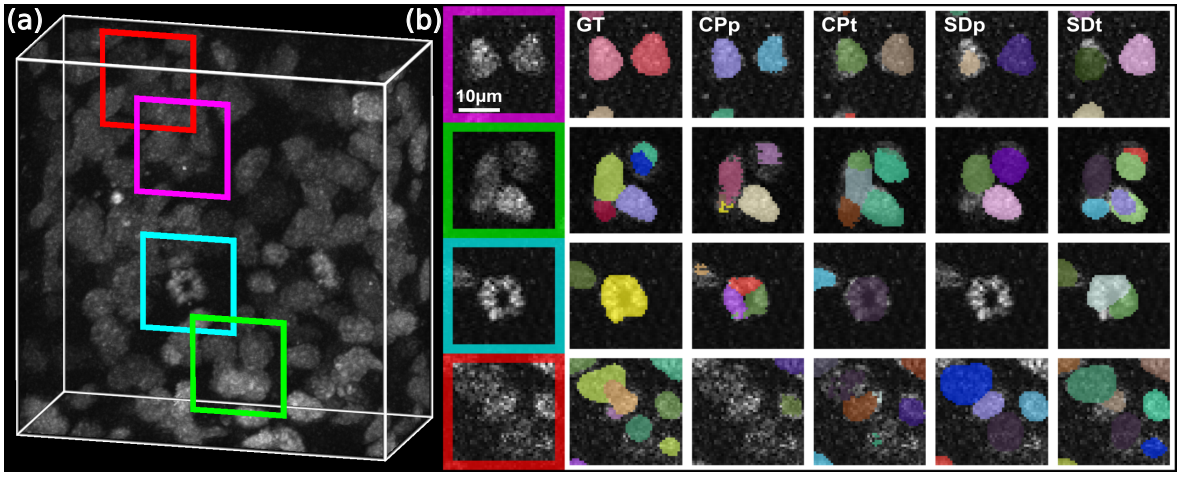

Figure 5.9: Visual example of typical Stardist3D and Cellpose response on challenging 3D live organoid cells data. (a) Here, a single frame of the dataset is rendered in 3D using Napari. The dataset exhibits high cell density with heterogeneity in cellular dynamics and appearance. (b) Models tested (CP,SD: Cellpose, Stardist3D; p,t: pre-trained, trained) are compared to the ground-truth segmentation (GT) in four illustrative ROIs: two highly visible nuclei, a cluster of nuclei, a dividing nucleus, and a cluster with bad signal-to-noise ratio at the border of the field of view.

5.5.2.3 Discrepancies in the value of the score suggest that it is sensitive to cellular environment.

Finally, we observe that the score associated with the pre-trained Stardist3D model reached the average score measured with the ground-truth. As such, it does not reflect the rank established by the true F1 score as opposed to Cellpose-pre-trained, Cellpose-trained and Stardist3D-trained. Further measurement and visual inspection performed to understand this discrepancy (Figure 5.10 and data not shown) suggest that Cellpose and Stardist3D exhibit similar rates but possibly different subtypes of false negatives. Indeed, on the one hand, cell clusters appear to be systematically missed by Cellpose but correctly detected by Stardist3D. On the other hand, isolated cells with unusual shapes seem to produce the reverse performance. The recent benchmarking study on both algorithms by Kleinberg and colleagues confirm that intuition [203]. As such, the discrepancy might stem not only from the mix of false negatives and false positives rates, but also from the cellular environment: whether it is in adherent cellular clusters or a freely moving cell in the aggregate. This could explain the overestimation of our quality score on the pre-trained version of Stardist3D since cell tracking in a relatively static cluster may present less variability in their trajectory-to-measurement association than freely moving cells. We discuss the implication of this hypothesis in our concluding Section.

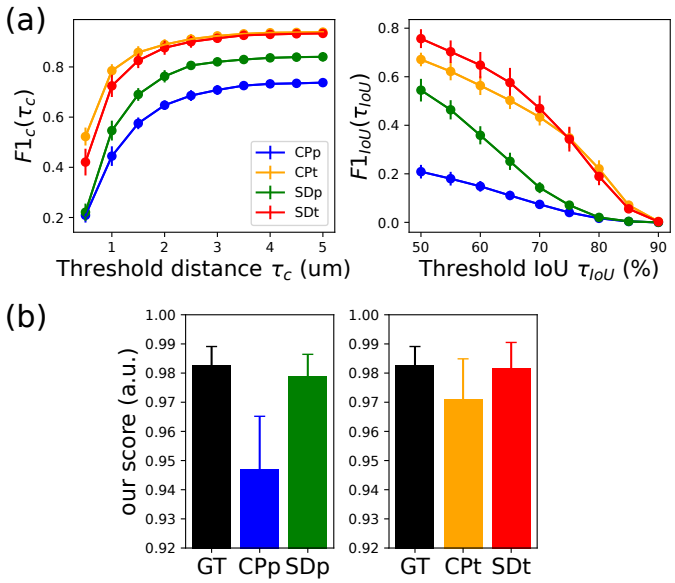

Figure 5.10: Our evaluation score can accurately rank three of the four models studied on the organoid dataset. (a) We use the F1 scores based on centroid detection and on instance segmentation to rank the models using the ground-truth segmentation. We observe better performance from trained models. (b) Our score provides a similar ranking except for the pre-trained model of Stardist3D.

5.6 Discussion

Our results demonstrate how using dynamical information measured through live cell experiments can be a powerful approach for the ranking of multiple detectors independently of their design. Indeed, our data shows that a score that merely described ambiguities in trajectory-to-detection assignment responds strongly to merging, splitting and false positive errors. Specifically, we show on simulations that our score approximates the F1 metric, a score commonly used for segmentation evaluation [77]. While dynamic ambiguities cannot detect individual FN, our score nevertheless responds well to false negatives in more realistic scenarios with mixed error types. In those mixed scenarios, the ranking between two detectors with different FN rates is preserved as long as the rate of FP does not exceed 40%. In practice, a detector with a 40% FP rate is performing particularly poorly and can be easily dismissed through visual inspection. For comparison, the worst false positive rate we can measure across all models (some of them pre-trained on different datasets by other labs) using the centroid-based F1 is 15% on our challenging dataset. The scripts for the analysis of the score are available on Github at github.com/bioimage-mining-group/disco-wight. Our score also shows very promising results when tested on experimental datasets with limited acquisition frequency and signal-to-noise ratio. We are indeed able to correctly rank the performances of deep-learning-based detectors, across pre-trained models and across trained models taken separately. However, we also show that trackability may be overestimated depending on the area of performance of a given detector. This indicates that a deeper analysis of the dynamics of segmentation candidates may help in improving ranking performances in those specific scenarios. As such, our work shows the potential for

using dynamics to detect segmentation errors, but it also opens a variety of paths for future work and exciting perspectives for truly unsupervised validation of cell segmentation. The remainder of this discussion focuses on both of those aspects.

A first path of improvement is linked to sensitivity toward false negatives. This can be improved by injecting more *a priori* on the lifetime of our target, which is here virtually infinite in our organoid scenarios. This approach would help reduce the likelihood of the trajectory-to-death association and increase instability. A second challenge that we did not explore lies in the distinction between split events and cell divisions. This could be addressed using an *a priori* on instantaneous phase transition similar to previous works [168, 204]. Finally, we will further test the impact of the heterogeneity of the underlying biophysical processes and the sparse nature of mosaic imaging techniques. This will require new methods for the simulation of cellular aggregates with a mixture of clusters and freely flowing cells as well as new imaging experiments imaged in non-mosaic conditions.

To conclude, this work paves the way toward the automated exploration of detection errors at the single cell level and the routine, annotation-free, evaluation of detection approaches that typically work as black boxes. An unsupervised score may also help in evaluating the need for retraining, further annotation, or even be part of the loss function used to train deep neural networks. It is our hope that those results will popularize the combination of live cell imaging and stochastic inferences to facilitate the validation of image-based measurements carried out at large scale. Building upon this work, the next chapter will further leverage trackability to perform unsupervised hyperparameter calibration of Bayesian tracking algorithms, addressing another critical bottleneck for practical cell lineage reconstruction in gastruloids.

Chapter 6

Application: Trackability allows unsupervised calibration of Bayesian tracking algorithms

6.1 Chapter overview

This chapter explores the practical application of the trackability metric for unsupervised calibration of Bayesian tracking algorithms. We demonstrate that incorporating a simple prior on trajectory lifetime significantly improves the utility of trackability as an optimization tool, facilitating the automatic and unsupervised selection of tracking parameters. Validation of this approach using simulation data confirms its effectiveness, and application to real gastruloid datasets yield promising results that highlight its practical utility.

My contributions:

1. benchmarked the use of the trackability score for the unsupervised calibration of a key hyperparameter in the cell tracking algorithm
2. augmented the classical trackability score to create a robust optimization metric for hyperparameter calibration
3. validated the metric using Brownian dynamic simulations generated with my custom simulator
4. validated the method qualitatively on real sparse and dense gastruloid datasets

6.2 Motivations

Bayesian tracking approaches, while offering interpretable and physically meaningful parameters, necessitate careful calibration. Traditionally, these parameters can be adjusted through visual inspection of tracks or by analyzing statistical distributions related to the filtering process. However, visualization itself presents a significant challenge in the context of dense gastruloid images. It is difficult to follow individual nuclei for more than a few frames, and even more challenging to disentangle errors stemming from tracker mis-

calibration from those caused by faulty segmentations. Furthermore, statistical analysis is non-trivial with highly heterogeneous and non-stationary data

Automatic calibration methods exist for Kalman filtering, and are traditionally used to provide robust and adaptive filtering by calibrating the parameters during the filtering process. Methods based on the Expectation-Minimization scheme calibrate the hyperparameters by maximizing the likelihood of the observed data given the current value of the parameters [205, 206]. However, these data usually require many data points to converge, and they focus on maximizing the likelihood of a single trajectory without considering the estimation of the optimal trajectory-to-measurement association. In other words, an optimal hyperparameter for the filter might still result in low tracking quality due to local trajectory density, which increases trajectory-to-measurement ambiguity.

Building on the trackability metric presented in Chapters 4 and 5, we aim to provide a proof of concept evaluation of the potential of trackability for tracker optimization. Specifically, we seek to (i) validate these claims within the simulation framework introduced in Chapter 5 and (ii) determine whether trackability-based optimization extends to the complex cellular dynamics observed in gastruloids. Notably, we will show that trackability exhibits a maximum that correlates with improved tracking performance. A simple penalization of the raw trackability score will be proposed to improve the globality of this maximum. This approach was first implemented by Laura Neschen, a former intern in the team.

6.3 Methods

6.3.1 Particle simulations

To assess the effectiveness of trackability as an optimization criterion, we employed our simulation framework to compare whether the maximum value of the trackability score with respect to a given hyperparameter range could be linked to the maximum of a supervised metric, the link Jaccard coefficient (see Appendix 6.5). The primary focus of this analysis was on Brownian dynamics, which was implemented by setting the Lennard-Jones interaction term to zero ($\epsilon = 0$) in our simulator. Trackability was averaged over entire movies to provide a robust global measure of tracking performance.

6.3.2 Penalized trackability as a robust score for hyperparameter optimization

Like in the previous chapter, we use the globally averaged trackability score as a metric, in this case to serve as a target to be maximized under variations of key parameters of the tracking algorithm.

During the course of this study, a fundamental limitation of the original trackability

definition became apparent. Specifically, when the number of estimated track initiation and termination events was increasingly high, indicating probable miscalibration of the tracking algorithm, trackability could be high. To address this issue, we add a conditioning with respect to biological priors (e.g track starts/ends are not expected to be frequent) to the original definition of trackability in equation (4.17).

As a reminder, the trackability score $T^{i,j}$ approximates the likelihood of the hypothesis $\eta^{i,j}$ given enough Monte-Carlo resamplings $T^{i,j} \underset{N_{\text{MC}} \to \infty}{\sim} p\left(\eta^{i,j} | \mathbf{x_t}\right)$.
We write

$$T^{i,j}_{\text{pen.}} \sim p\left(\eta^{i,j} \mid \mathbf{x_t}, B\right) = \frac{p(B \mid \eta^{i,j}, \mathbf{x}_t)\, p(\eta^{i,j} \mid \mathbf{x}_t)}{p(B \mid \mathbf{x}_t)} = \mathcal{P}^{i,j}\, p(\eta^{i,j} \mid \mathbf{x}_t), \tag{6.1}$$

where we introduced B as biological prior information about the cells lifetimes, used Bayes theorem to write the first equality, and defined the penalization factor $\mathcal{P}^{i,j}$

$$\mathcal{P}^{i,j} = \frac{p(B \mid \eta^{i,j}, \mathbf{x}_t)}{p(B \mid \mathbf{x}_t)}. \tag{6.2}$$

This finally leads to the definition of the penalized trackability score $T^{i,j}_{\text{pen.}}$ with respect to the classical trackability score $T^{i,j}_{\text{pen.}} = \mathcal{P}^{i,j}\, T^{i,j}$.

This modified formulation adjusts the hypothesis likelihood by incorporating a prior probability term that reflects biological information about tracked object.

The biological knowledge we wish to enforce here is that tracks are not expected to start or end "frequently", in order to penalize tracking solutions from miscalibrated trackers in which e.g the search radii are chosen too small and tracks fail to persist. In the context of this work, which serves as a proof of concept, we aim for a minimal-prior implementation of the penalized trackability score, and we thus set

$$\mathcal{P}^{i,j} = \begin{cases} 0, & \text{if } \eta^{i,j} \text{ involves starting/ending a track} \\ 1, & \text{otherwise.} \end{cases} \tag{6.3}$$

Consequently, when the tracker assigns high confidence to excessive track starts or deaths, the globally averaged trackability score is effectively penalized. This modification ensures that trackability remains a meaningful metric even in scenarios with high track starts/ends rates.

This naive implementation penalizes every track start/end equivalently, without considering whether the position of the object suggests it could be close to the image boundary. The penalization could be refined by properly modeling the likelihood of track starts/ends close to boundaries, as is done e.g. in btrack [74].
Similarly, when objects are expected to appear or disappear mid-track, as is the case in molecule tracking problems, in which particle blink, Bayesian priors have been designed to model the likelihood of appearance/disappearance [160, 207].

6.4 Results

6.4.1 Penalized trackability guides optimal process noise in Brownian dynamics simulations

Trackability is not an adequate optimization target to find optimal parameters due to the lack of priors on trajectory lifetimes To evaluate the optimization of a single tracking parameter using trackability, we first conducted simulations of Brownian motion. The process noise covariance in the Kalman filter was parameterized as an isotropic matrix, $Q = qI$, where q was optimized and I is the identity matrix. Simulations consisted of $N = 100$ particles undergoing Brownian motion, each characterized by a different diffusion coefficient. A small measurement noise, r, was manually added to the data. Particles had no limit on lifetimes expect if they came out of the field of view, in which case a new trajectory was started from the opposite boundary (periodic boundary conditions). We used adaptive search radii as specified in Section 4.2.5.2.

For each simulation, tracking was performed across a range of q values using the Bayesian tracker. As shown in Fig. 6.1(a), for a simulation with an intermediate diffusion coefficient, the Jaccard metric exhibited a single peak. As q approached zero, the Jaccard metric tended to zero, indicating that overly small search radii (as obtained when specifying a small q in the adaptive search radii scheme) prevented the tracker from maintaining continuous trajectories, leading to repeated track deaths and reinitializations. Conversely, for larger q values, the Jaccard metric saturated, suggesting that increasingly large search radii did not significantly influence assignments. Indeed, beyond a certain value of the search radius, independent trajectories get fused, which results in a fully-connected tracking graph on which new links cannot be added or changed. This results in a stable Jaccard coefficient.

The trackability score, averaged over the full simulation, displayed a local maximum near the Jaccard peak. However, as q decreased, trackability approached one, an unintended behavior stemming from the tracker's overconfidence in track start/end events when alternative hypotheses were absent due to small search radii. In this case, the trajectories estimated are too short, but the trackability score confirms that the set of trajectories is possible considering our a priori model (here particularly through the definition of the adaptive search radii). For large values of q, trackability decreased, as the spread of the Monte Carlo resampling in the trackability estimation depends on the magnitude of the predicted covariance, which grows with Q (see equation (4.8)), which preserve the (erroneously) fully-connected nature of the tracking graph, but results in uncertain assignments.

A simple prior on trajectory lifetime enable the estimation of the optimal process noise To mitigate the undesired increase in trackability at low q, a penalized trackability score was introduced. This modification involved multiplying the link-scale trackability score by a factor that equaled one for standard tracking hypotheses but was set to zero for track start/end events. This adjustment effectively discouraged tracking solutions that excessively favored track initiation or termination. The resulting penalized trackability displayed a peak near the Jaccard maximum and tended toward zero for extreme values of

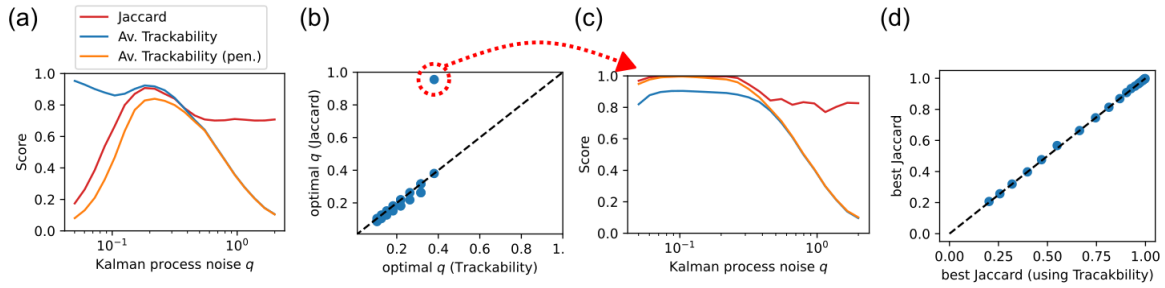

Figure 6.1: **In Brownian simulations, trackability can be used to optimize for process noise without ground-truths. (a)** For one simulation condition, q is varied while link Jaccard coefficient and trackability scores are being computed. The penalized trackability score displays a global maximum, like the Jaccard coefficient. **(b)** The optimal q obtained by optimizing the penalized trackability is very close to the true optimal q. **(c)** Outliers in (b) are due to plateaus in scores when difficulty is too low. **(d)** Maximizing the penalized trackability systematically gives optimal performances.

q, ensuring a well-defined global optimum (Fig. 6.1(a)).

Across a range of diffusion coefficients, Fig. 6.1(b) compares the optimal q obtained by maximizing trackability against the true optimal q (in terms of Jaccard). Most values aligned with the identity line $y = x$, demonstrating excellent parameter estimation. A single outlier corresponded to a condition with a small diffusion coefficient, where both the Jaccard metric and penalized trackability plateaued (see Fig. 6.1(c)), allowing for variations in optimal q without affecting final Jaccard scores. Fig. 6.1(d) confirms this by plotting the best Jaccard obtained through penalized trackability maximization against the best possible Jaccard coefficient. The data points aligned perfectly with the $y = x$ line, indicating that maximizing penalized trackability reliably ensured optimal tracking performance.

6.4.2 Penalized trackability shows promising results in real gastruloid datasets

The methodology was first applied to a dense gastruloid dataset obtained from a fusion experiment, for which manually curated segmentations were available. Using the Bayesian tracker with Brownian motion and isotropic process noise assumptions ($Q = qI$), the parameter q was optimized while all other parameters were calibrated visually.

Despite highly heterogeneous motion, both the original and penalized trackability scores exhibited similar trends to those observed in the simulations (Fig. 6.2(a)). In the absence of available ground-truth trajectories for direct validation, and due to the challenge of visual validation, we were not able to provide a rigorous proof of the optimality of the value of q obtained by optimizing the corrected trackability score. However, qualitative inspection of the tracks suggested that the optimal q value closely matched the expected true optimum.

To reduce the difficulty associated to the visualization challenge, the methodology was extended to a sparse gastruloid dataset, where density was reduced by selectively labeling

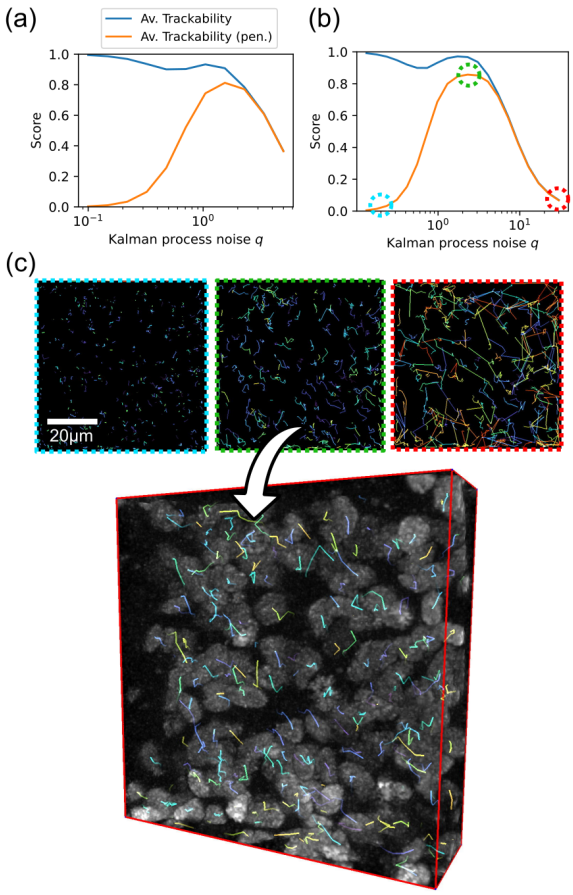

Figure 6.2: **Optimization of process noise with trackability is qualitatively validated on real gastruloid datasets. (a)** On a dense dataset with heterogeneous motion, the penalized trackability still exhibits a global maximum, which leads to a qualitatively satisfying tracking result. **(b)** On a sparse dataset with heterogeneous motion, the same behavior is observed. **(c)** Trajectories from three regions of (b) are displayed. The trajectories associated to the optimal value given by the trackability are shown below.

a fraction of nuclei to facilitate visualization.

Fig. 6.2(b) presents the trackability and penalized trackability scores, both of which exhibited trends consistent with the simulation results. Again, optimal tracking results were confirmed through visual inspection, demonstrating that tracking quality was maximized at the q value corresponding to the peak of penalized trackability. Examples of tracking outcomes for optimal and suboptimal q values are illustrated on Fig. 6.2(c). As discussed in the simulations, suboptimal q values lead to erroneously large links when q is chosen too large, in which case the search radii allow for spurious links instead of track starts/ends, and to repetitive starting and ending of tracks in the case where the radii are so small that no matches are possible. Confirmation of the optimality of the best tracking results should be confirmed through careful inspection by multiple human experts to draw further conclusions. Alternatively, one could design tailored experiments with by pushing the framerate as high as possible and imaging for short periods of time until phototoxicity and bleaching prevent further identification of relevant objects.

6.5 Discussion

This study provides strong evidence that trackability can serve as an effective criterion for optimizing tracker parameters. Through simulations, trackability-based optimization was quantitatively validated, and its applicability was qualitatively confirmed in gastruloid datasets. A modified version of trackability, penalizing track start/end events, was introduced to better align with biological expectations. In cell imaging, track initiation and termination events are primarily expected near image boundaries or during cell divisions, but the bulk of tracking data is dominated by persistent trajectories. This modification ensures that tracking performance is not artificially inflated by an overabundance of track start/end hypotheses.

Despite its heuristic nature, the penalization scheme we have used for this study successfully enforces our biological prior information. This could fail in scenarios where the field of view is small, in which case track starts/ends cannot all be penalized equally. To further refine trackability as a physically informed metric, future work could incorporate explicit dependencies on proximity to image boundaries. Computing the probability that a predicted movement would lead an object outside the field of view could prevent bias in penalized trackability for small imaging regions.

Additionally, future efforts should explore trackability optimization in more complex simulation environments, including heterogeneous motion dynamics, to assess whether multiple parameters can be optimized jointly. Advanced optimization techniques may be required to avoid exhaustive grid searches in high-dimensional parameter spaces. Applying this methodology to benchmark datasets, such as those from the Cell Tracking Challenge [208], will further validate its generalizability to real-world biological data.

Overall, this study demonstrates that trackability, particularly in its penalized form, is a promising tool for optimizing tracking parameters. The results establish a foundation for unsupervised parameter tuning, applicable across different biological contexts, and provide a path forward for improving automated tracking in dense, dynamic cell populations. Collectively, Chapters 4, 5, and 6 form a methodological framework that enables robust, uncertainty-aware lineage reconstruction, paving the way toward reliable and scalable cell lineage reconstruction in gastruloids.

Conclusion

This thesis has focused on laying technical landmarks towards the automated exploration of complex organoid datasets in their native three-dimensional state. In particular, contributions were made to advancing the quantification of cell-scale mechanics and genetics in gastruloids, focusing on both methodological developments and computational tools to facilitate the study of deep, dense three-dimensional embryonic organoids despite stochastic dynamics and noisy measurements.

The first contribution of this research was the creation of a framework for multiscale analysis of mechanical and gene expression properties in deep tissues. A protocol was developed for the segmentation of nuclei in deep tissues, incorporating tailored preprocessing and deep-learning-based segmentation strategies. To assess the performance of the algorithm, a novel validation technique based on 2D annotations was introduced, significantly reducing annotation time by a factor of five, which allowed fully leveraging 3D annotated datasets to train the segmentation model and validation of the performance across different modalities. The model achieves an overall F1 score at 50% IoU of 85% reliably at depths exceeding 200 μm. Furthermore, through the spatial registration and fusion of dual-view datasets, the model was successfully used to segment nuclei in wholemount gastruloids, providing a comprehensive view of tissue architecture. Collaborators successfully applied this same algorithm to segment nuclei in images of drosophila wing disc and in salivary glands organoids, demonstrating its generalizability.

While segmentation errors remain due to the scarce amount of training data and inherent complexity of the images, this algorithm was demonstrated to be robust for the quantification of morphological deformations, which served in a collaboration focused on developing imaging and computational protocols to investigate mechanical and genetic properties in deep gastruloids. In this context, a methodology was designed to measure gene expression in an unbiased manner by leveraging ubiquitous signals for intensity normalization. This approach was rendered robust against wavelength-dependent scattering effects through the integration of a physical model accounting for scattering-induced depth decay. To enable intuitive data exploration, a software package was developed, prioritizing visual interactivity and user-friendliness. In addition to core segmentation functionalities, the software package incorporates advanced analytical tools, including spatial correlation histograms, multiscale instance-aware coarse-graining, and true strain analysis for deformation quantification. Fixed dataset analyses conducted using this framework revealed key morphological and mechanical insights, such as the deformation of nuclei along stress lines and the quantification of local cell packing heterogeneities. The software has been adopted by researchers within the team, supporting ongoing studies in organoid research, including the multiscale spatial analysis of cell division patterns. Moreover, its integration of macro-recording functionalities fosters reproducible analysis, ensuring long-term usability. The platform established in this work serves as a foundation for future extensions incorporating dynamic datasets.

Beyond cell tracking, methodological bottlenecks in accurate lineage reconstruction for dense organoid datasets were systematically examined through a comprehensive review of state-of-the-art tracking algorithms. The strengths and limitations of these methods were assessed within the constraints of gastruloid imaging. A major challenge in this domain remains the calibration and validation of tracking methodologies, largely due to the scarcity of ground-truth annotated trajectories and the difficulty of directly visualizing trajectories from two-dimensional slices of the data. To address the latter limitation, a visualization software was developed to facilitate data exploration in both three- and two-dimensional regions of interest. Implemented as a plugin for the open-source visualization platform Napari, the software leverages Napari's extensive and interactive customization capabilities to tailor the display to specific analysis needs.

To address the uncertainties inherent to cell tracking in deep tissues, a tracking algorithm based on Bayesian filtering and hidden Markov models were implemented. This framework allows quantifying tracking ambiguity in a robust and unsupervised manner. This approach was demonstrated in previous studies to provide a direct readout of tracking quality and to facilitate the validation of tracking quality across scales, mitigating the challenges associated with manual validation and exploratory analyses. To tackle the challenge of using imperfect segmentation results in the tracking workflow which was implemented, this tracking ambiguity metric was adapted for the global ranking of segmentation algorithm quality. A simulation framework was constructed to replicate typical features of nuclear dynamics, providing a controlled environment for evaluating detection performance. This framework enabled the validation of the method on simulations designed to mimic gastruloid datasets in which typical segmentation errors were added, demonstrating fidelity in measuring segmentation errors, even under conditions with high false positive and false negative rates. Further validation was performed using a benchmark dataset with challenging segmentation conditions, as well as a sparse organoid datasets. The results highlighted the ability of the proposed approach to reliably rank state-of-the-art segmentation algorithms while revealing a sensitivity to local error distributions, suggesting potential applications in localized segmentation validation.
To tackle the challenge of tracker calibration, the tracking ambiguity metric was once again adapted to create a well-grounded biology-inspired score which can serve as a target for black-box optimization of key tracker parameters. The simulation framework was used to validate the approach, demonstrating accordance with a supervised score of tracking quality. The approach was applied to real gastruloid datasets and showed promising results in allowing blind calibration on complex datasets, suggesting proper validation on benchmark datasets and further exploration of the limits in future works.

In summary, significant advancements were made in segmentation, tracking, and quantitative analysis methodologies for dense embryonic organoid datasets. While substantial progress was achieved, several challenges remain, particularly in lineage reconstruction, where validation constraints persist. Future work will benefit from the integration of more extensive validation strategies, the refinement of tracking algorithms, and the extension of the biophysical analysis pipeline to dynamic datasets. Nevertheless, the methodologies and computational tools introduced in this thesis provide a robust foundation which paves the way for further exploration of cellular dynamics in complex gastruloid datasets.

Perspectives

Extending the use of trackability

As we have explored in this work how trackability is sensitive to segmentation and tracking quality, and how it can be used to optimize the parameters of our Bayesian tracker continuously, a promising avenue for extending trackability lies in its use as a metric to calibrate the postprocessing parameters of segmentation and tracking algorithms. In segmentation algorithms such as Stardist3D, a postprocessing step involving thresholds is used on the network outputs to transform them into segmentations. These thresholds may require fine-tuning to yield robust results. By employing trackability as an objective function, these thresholds could be optimized to maximize the global trackability.
For tracking algorithms, trackability could be estimated on precomputed tracks by applying our Bayesian filtering framework normally while forcing the association step to choose the links given by the input tracks. This could be applied for instance to alleviate the need for manual calibration of birth/death costs in assignment frameworks. This is particularly relevant as new state-of-the-art tracking methodologies rely on solving network flow ILP to compute optimal associations, which requires many user-defined costs parameters.

A direct continuation of the work we have performed on using trackability as a metric for global segmentation quality is to investigate its properties at the single-link scale. This could help pinpoint problematic areas in segmentation or to identify particular events such as cell divisions. By directing expert attention to localized inconsistencies, the overall efficiency and accuracy of segmentation and track correction processes could be substantially improved.

Another potential direction arises from incorporating trackability into deep-learning pipelines as a means of influencing model training, even though it is not trivially differentiable. Reinforcement learning frameworks offer one possible route, since trackability can function as a reward signal reflecting how successfully a model maintains coherent tracks. The training process would iteratively adjust the model's parameters to maximize this reward, thereby encouraging segmentations and associations that minimize uncertainties in object association. Although integrating such a metric within a standard convolutional neural network may be non-trivial, the growing adoption of reinforcement learning methods in computer vision underscores the feasibility of exploring this approach. This prospect will be explored by a postdoc in the team.

Finally, the deployment of trackability in real-time or near-real-time pipelines presents a valuable opportunity for advancing the field of smart microscopy. As imaging data is acquired, trackability can serve as an unsupervised metric to guide adaptive decisions in microscope control, such as adjusting imaging parameters (e.g laser gain or framerate) or suggesting different acquisition modes when low trackability suggests ambiguous cellular motion or poor segmentation fidelity. This could allow better resource allocation, minimizing phototoxicity and bleaching when high or rapid illumination is unnecessary, and

designing experiments that capture rare biological events with greater precision [209].

Towards a tracking strategy for gastruloid datasets

From our review on current strategies for tracking in developmental biology, we note that to our knowledge, no existing method incorporates (i) ILP-based global optimization of lineage constraints, (ii) hierarchy-based segmentation using temporal context, and (iii) advanced modeling of individual cellular dynamics (e.g with deep learning or Bayesian methods).
In this prospect, Ultrack [163] provides an extensive software that solves points (i) and (ii) on top of being computationally and memory-efficient, fast, modular, well-maintained and well-documented. If no modeling knowledge is available or if many ground-truth tracks can be annotated, deep-learning based algorithms like Trackastra [165] can be used to compute accurate linking likelihoods. On the contrary, Bayesian methods can be leveraged for physics-enriched likelihoods modeling in the ILP framework when calibrating priors is easier than annotating ground-truths. In particular, we have shown that trackability can be used to aid calibration in an unsupervised manner.

Such a method could be greatly influential in the problem of cell tracking in gastruloids, as it would fit the constraints introduced in Section 3.2.2. Notably, Chenouard et al. [160] already laid the ground for Bayesian methods in multiframe association algorithms, Turetken et al. [162] integrated hierarchy-based segmentation during tracking while in a Bayesian framework, and Bragantini et al. [163] released a scalable implementation of hierarchical methods combined with ILP-based global optimization. Because they often concentrate on low-dimensional representations of the state of the tracked objects (instead of e.g the images of the objects), Bayesian methods are inherently computation-efficient and scale extremely well to large datasets [74, 75], which suggests that they could pair well within the framework of Bragantini et al. [163].

In the context of cell tracking, correctly modeling likelihoods for biological events like cell division is fundamental. As this can be complex, inspiration can be taken from btrack [74], which computes a cell division likelihoods based on the estimated progression of cells in the mitotic cycle. This allows preventing high division likelihoods from arising when cells are detected to be in a state that biologically prevents it.

Through collaborations with the team of Jean-François Rupprecht (CPT, Univ. Aix-Marseille), we have participated in the design of a deep-learning method for division detection based on temporal context. Successfully applied in 2D datasets in the past [135], the method is based on a first 4D U-net [92] which analyzes local spatiotemporal windows of the data (the network uses time as the fourth dimension) and outputs a pixelwise map of division. After inference on the full volume of a movie, local maxima in division probability are detected, and spatiotemporal windows are division candidates are given to a second 4D U-net which outputs an estimate of the orientation of the daughter-daughter axis. This algorithm could be used to infer cell division likelihoods by extracting the local maximum likelihood output by the first network, and applying Platt's scaling to calibrate the result

[210].

Appendices

Derivation of maximum-likelihood solution for multiple Brownian particles tracking

For the specific case of a particle whose dynamics is described by equation (3.1) with a stationary and homogeneous drift term $\boldsymbol{\mu}$, the Fokker-Planck equation for the PDF $p(\mathbf{x}, t|\mathbf{x}_0, t_0)$ reads

$$\frac{\partial}{\partial t}p(\mathbf{x}, t|\mathbf{x}_0, t_0) = \frac{1}{\gamma}\nabla \cdot (\boldsymbol{\mu} p(\mathbf{x}, t|\mathbf{x}_0, t_0)) + \frac{\sigma^2}{2\gamma^2}\Delta p(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{6.4}$$

This is a standard diffusion equation with drift, where we can identify the diffusion coefficient $D = \sigma^2/2\gamma^2$ and the drift velocity $\mathbf{v}_d = \boldsymbol{\mu}/\gamma$. This equation can be solved by the method of separation of variables, which yields the following solution in d dimensions

$$p(\mathbf{x}, t|\mathbf{x}_0, t_0) = \frac{1}{(4\pi D(t-t_0))^{d/2}} \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}_0 - \mathbf{v}_d(t-t_0)\|^2}{4D(t-t_0)}\right] \tag{6.5}$$

The PDF of the particle is Gaussian, its center moves at velocity $\mathbf{v}_d$ on the line of equation $\mathbf{x} = \mathbf{x}_0 + \mathbf{v}_d(t-t_0)$, and the spread of the distribution, given by its variance $2D(t-t_0)$, grows linearly with the elapsed time. Interestingly, this PDF has a spherical symmetry around the central position $\mathbf{x}_0 + \mathbf{v}_d(t-t_0)$, so that it only depends on the distance

$$\delta = \|\mathbf{x} - \mathbf{x}_0 - \mathbf{v}_d(t-t_0)\| \tag{6.6}$$

to the center (provided t is given).

Note that from a purely mathematical standpoint, this solution holds if the particle evolves freely in an infinitely large space so that vanishing boundary conditions can be used. In practice, we assume that that the typical dimension L of the d-dimensional space is much bigger than the typical particle displacement so that the particle remains at all time far away from the space's boundaries, i.e

$$L \gg \|\mathbf{v}_d\|(t-t_0) + \sqrt{2dD(t-t_0)} \tag{6.7}$$

To reconstitute the motion of N particles undergoing Brownian dynamics from a set of images, we alleviate the challenges that come with detecting objects in images and assume that the first detection step has been done already and yielded perfect results. We thus leverage the time-tested approach of "detect-then-track" [73, 75, 139, 144–146], which consists in treating the detection and association tasks independently. This paradigm is traditionally used when temporal context is not necessary to obtain perfect detection results. This approach is consistent with the pioneering work of Bao et al. [144] who performed

automated lineage tracking in *Caenorhabditis elegans* embryos imaged with a confocal microscope. Due to the small size (relative to the size of the cells) and the sparsity of nuclei in their images, they obtained a quasi-perfect detection result using a simple local-maximum algorithm. However, access to perfect detections/segmentations is usually not a marginal assumption in developmental biology studies, as detecting cells or nuclei in embryos is a challenging task. Most methods adopt a different paradigm that will be further addressed later.

To link detections in one frame to their counterparts in the next frame, we wish to leverage the probabilistic description of Brownian dynamics we obtained before to define the most likely assignment. To simplify this case study to the maximum, we make four assumptions to proceed with the demonstration

- Though cells in dense tissues are by definition constantly interacting to maintain the integrity of the tissue, we make the assumption that the motion of the particles between two consecutive frames is not dominated by the inter-particle interactions, and that particles are thus independent time scales comparable to the framerate.
- The objects are virtually point-like (feature-less), and thus perfectly indistinguishable, so that only proximity of a particle from one frame to the particles in the next frame will be used to compute the most likely assignment.
- Particles have no drift velocity.
- Particles have the same diffusion coefficient. Its value does not need to be known at this stage.

We have established in equation (6.5) that the likelihood of a Brownian particle moving with a displacement $\boldsymbol{\delta} = \mathbf{x}_{t+\tau} - \mathbf{x}_t$ between two consecutive frames separated by a time τ is

$$p(\boldsymbol{\delta}|\tau) = \frac{1}{(4\pi D\tau)^{d/2}} \exp\left[-\frac{\delta^2}{4D\tau}\right] \tag{6.8}$$

Now if we postulate a given one-to-one assignment (each particle in a frame is linked to a single particle in the next frame) which results in observing N individual displacements $\{\boldsymbol{\delta}^l\}_{l\in[1;N]}$ for the N noninteracting particles, the associated likelihood of observing $\{\boldsymbol{\delta}^l\}$ is the product of the individual likelihoods

$$p(\{\boldsymbol{\delta}^l\}|\tau) = \prod_{l=1}^{N} p(\boldsymbol{\delta}^l|\tau) = \frac{1}{(4\pi D\tau)^{Nd/2}} \exp\left[-\sum_{l=1}^{N} \frac{(\delta^l)^2}{4D\tau}\right] \tag{6.9}$$

The assignment associated with the maximum likelihood maximizes $p(\{\boldsymbol{\delta}^l\}|\tau)$, or equivalently minimizes the negative log-likelihood

$$-\log\left(p(\{\boldsymbol{\delta}^l\}|\tau)\right) = \sum_{l=1}^{N} \frac{(\delta^l)^2}{4D\tau} + \frac{Nd}{2}\log\left(4\pi D\tau\right) \tag{6.10}$$

In theory, the most likely set of trajectories could thus be constructed by iterating over all consecutive pairs of frames and finding the optimal assignment resulting in optimal displacements $\{\boldsymbol{\delta}^l_*\}$ maximizing $p(\{\boldsymbol{\delta}^l\}|\tau)$ among all possible $\{\boldsymbol{\delta}^l\}$ of the pair.
This formalizes as

$$\begin{aligned}\{\boldsymbol{\delta}^l_*\} &= \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmax}}\, p(\{\boldsymbol{\delta}^l\}|\tau) \\ &= \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmin}} \left[-\log\left(p(\{\boldsymbol{\delta}^l\}|\tau)\right)\right] \\ &= \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmin}} \sum_{l=1}^{N} \frac{(\delta^l)^2}{4D\tau} + \frac{Nd}{2}\log\left(4\pi D\tau\right)\end{aligned} \tag{6.11}$$

Due to the assumption that all particles have equal D, this finally reduces to a problem of minimizing the sum of squared distances

$$\{\boldsymbol{\delta}^l_*\} = \underset{\{\boldsymbol{\delta}^l\}}{\operatorname{argmin}} \sum_{l=1}^{N} (\delta^l)^2 \tag{6.12}$$

Performance metrics

Detection and segmentation

When comparing detections, having a ground-truth allows one to precisely quantify which true objects have been detected (TP), and which ones have been missed (FN) or correspond to false detections (FP). The F1 score, also called the Dice score, is a popular choice for a universal score taking all error types into consideration[77]. It is defined as the harmonic mean of precision and recall, and reads

$$F1 = \frac{2TP}{2TP + FN + FP}. \tag{6.13}$$

Note that it is monotonously related to the accuracy metric. If an object can be represented by a single point, e.g if it is point-like by nature or if its centroid gathers most of its positional information, these quantities are computed in a one-to-one assignment framework, in which a detection is consider a TP if it is sufficiently close (up to a threshold distance τ_c) to a point in the ground-truth. We will call this metric the centroid-based F1 score $\mathrm{F1}_c\left(\tau_c\right)$. Points that are not matched in the detected and ground-truth datasets are considered FPs and TPs respectively. When objects are detected as volumes as opposed to points, they can be made of several pixels/voxels or be represented by enclosing volumes (e.g bounding boxes). The correspondence between two individual objects can be assessed by computing the Intersection over Union score from their volume

$$IoU = \frac{V_D \cap V_{GT}}{V_D \cup V_{GT}} \tag{6.14}$$

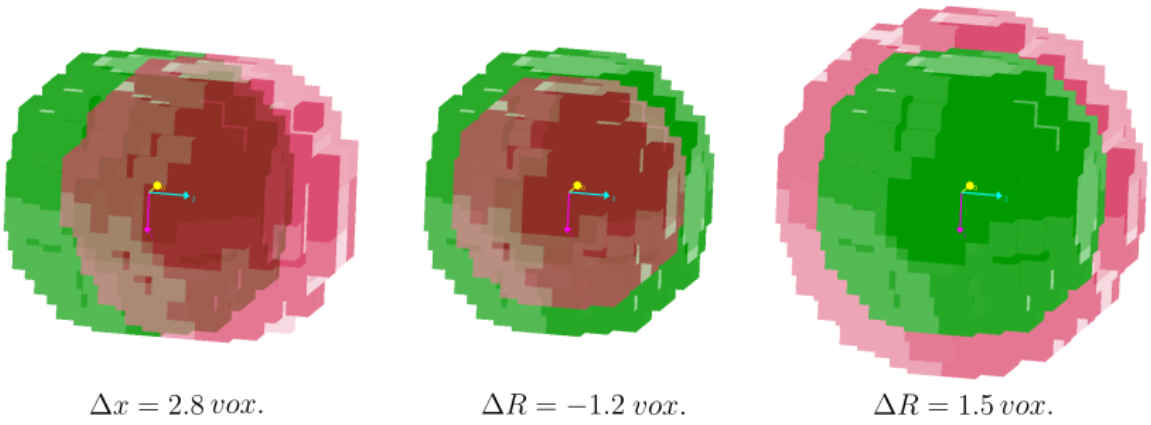

Figure 6.3: **Visualization of difference between ground-truth and candidate segmentation at 50% IoU.** We show how far a ground-truth spherical 3D segment of radius $R_0 = 6$ voxels (in green) can differ from a candidate spherical segment (in red) either shifted (left) or centered but stretched down/up (middle and right respectively) such that their IoU remains above 50%.

with V_D and V_E the volumes defined by a detected and a ground-truth object respectively. The F1 score can then be computed, in which case a detected object is considered matched to a ground-truth object (and thus a TP) if their IoU is above a certain threshold τ_{IoU} (usually above 0.5). This is again determined in a one-to-one assignment framework. We will call this metric the IoU-based F1 score $\text{F1}_{IoU}\,(\tau_{IoU})$. In Chapters 1 and 2, τ_{IoU} is set to $0,5$ unless specified otherwise, and the metric is simply called F1 score. On Fig. 6.3, we illustrate using spherical 3D segments what a F1_{IoU} score of $0,5$ represents in terms of pure translations or dilations.

Tracking

Over the past years, numerous metrics have been developed to evaluate the performance of multiple object tracking algorithms, enabling precise and standardized comparisons within the bioimage analysis community [208].

Tracking performance can be assessed through a range of approaches, from simple methods that count false positives (FP) and false negatives (FN) to more sophisticated high-level metrics. Among the latter, metrics such as the the Tracking Accuracy (TRA) score [208], the Higher-Order Tracking Accuracy (HOTA) [211], or the Cell Lineage-based Object Tracking Accuracy (CHOTA) [212] provide increasingly more comprehensive evaluations by considering detection, accuracy, association quality, and cell lineage reconstruction accuracy. These advanced measures offer deeper insights into the structural consistency of tracking outputs, moving beyond basic error counting to assess how well cell trajectories are reconstructed over time.

In this work, we rely exclusively on the link Jaccard coefficient, also referred to as the link Jaccard index, as our primary tracking performance metric. This measure is derived

from the classical Jaccard formula commonly used in segmentation tasks (see equation (6.13)), but it is specifically adapted to evaluate the correctness of links between detected objects. In this context, true positives (TP), false positives (FP), and false negatives (FN) are defined based on the presence or absence of links between already precomputed detections or segmentations. Consequently, this metric is suitable for assessing tracking algorithms that operate on fixed segmentation results, as it inherently assumes that segmentation provides the spatial anchors for evaluating link correctness. However, it is not applicable to tracking approaches that simultaneously refine or perform segmentation during the tracking process, as the definition of ground-truth links becomes ambiguous when segmentations are not fixed.

Depth-dependant correlation of intensity between immunostained and reporter T-bra signals

To assess whether the observed T-Bra intensity gradient results from differential antibody penetration or reflects a genuine expression pattern, we compared immunostained T-Bra signal with a transgenic T-Bra reporter (H2B-GFP expressed under a T-Bra promoter) in gastruloids fixed at 72 h. Fig. 6.4(a) presents a representative sample, showing a strong correlation between the immunostained (magenta) and endogenous reporter signals (green), except in mitotic cells where the reporter signal displays is visible in a rounded region.

Fig. 6.4(b) displays XZ views of Hoechst, endogenous T-Bra, and immunostained T-Bra signals before and after intensity normalization using Hoechst as a reference. This normalization accounts for potential wavelength-dependent variations in signal attenuation.

To quantify the correlation, we computed intensity correlation heatmaps between the immunostained and endogenous signals across the full sample (Fig. 6.4(c)). After normalization, the two signals are nearly identical apart from a small additive bias. Further analysis in three depth-resolved slices (Fig. 6.4(d)) confirms a consistent correlation across all depths, indicating that antibody penetration heterogeneities do not contribute significantly to the observed T-Bra gradient.

Piecewise-stationary motion smoothing for multiple dynamics hypothesis and dynamical switching

Piecewise-stationary motion modeling

A single Kalman filter, as presented in 4.2.3, cannot adapt to unpredictable changes in particle dynamics [204] due to the motion model being fully characterized by the dynamical model matrix F. Thus, it cannot accurately model the heterogeneity of motion in cell aggregates requires handling multiple motion types.

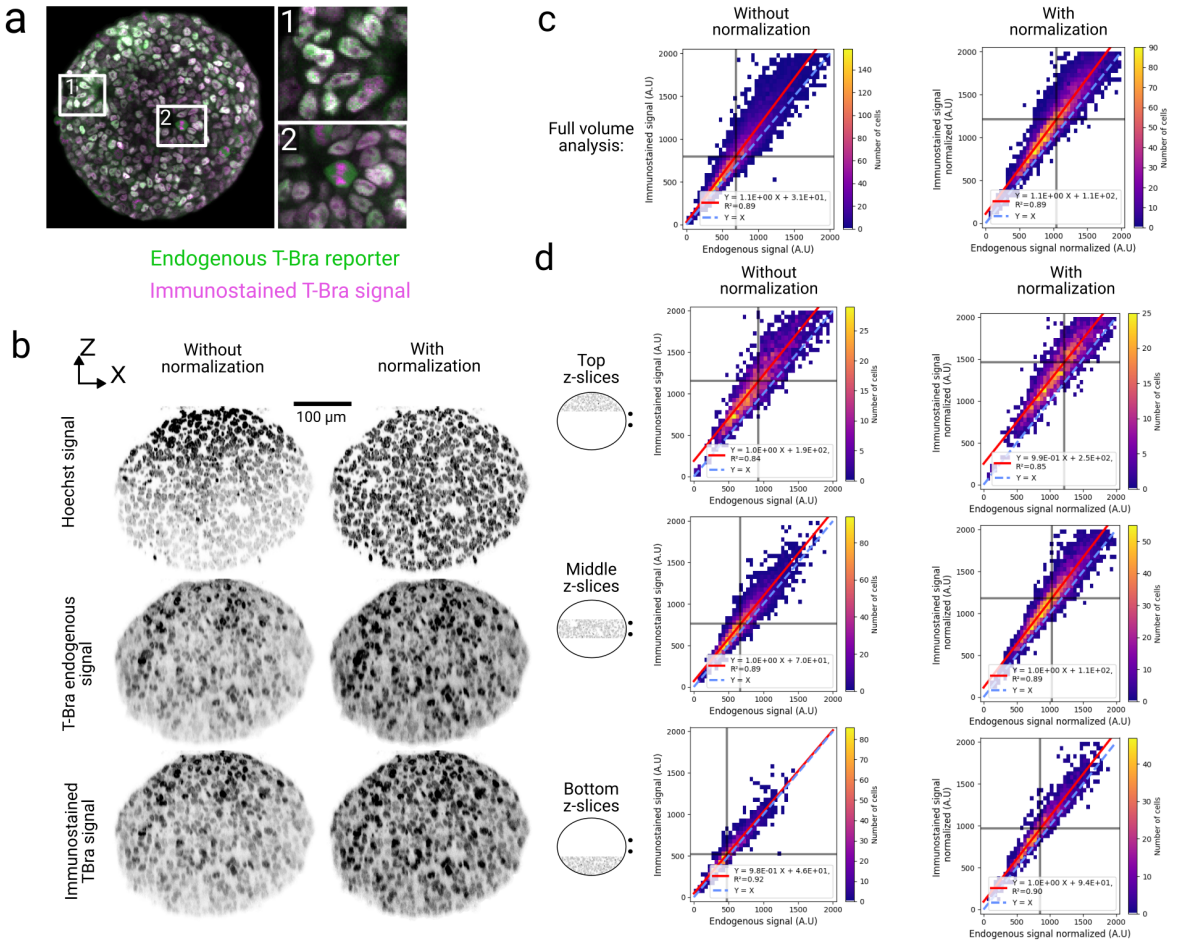

Figure 6.4: **Comparison of immunostained T-Bra and endogenous T-Bra reporter signals in gastruloids fixed at 72 h. (a)** Representative sample showing strong signal correlation, except in mitotic cells. **(b)** XZ views of Hoechst, endogenous T-Bra, and immunostained T-Bra signals before and after normalization. **(c)** Correlation heatmaps across the full sample, demonstrating near-equality of signals post-normalization. **(d)** Correlation heatmaps in three depth-resolved slices, confirming homogeneity in the correlation relationship.

Particle tracking in challenging scenarios involving rapid motion changes or high particle densities is effectively addressed by combining stochastic smoothing with piecewise-stationary motion modeling. This approach has been shown to mitigate issues of incomplete trajectory reconstruction. The method uses a two-filter smoothing technique that integrates both past and future measurements to predict particle displacements, reducing the likelihood of early track termination while avoiding reliance on large temporal windows. In addition, the piecewise-stationary motion model enables the dynamic handling of abrupt transitions between motion regimes. This is achieved through the adaptive reinitialization of Kalman filter variables during iterative tracking rounds. Originally implemented within the *u-track* framework, the method enhances the frame-to-frame linking step without introducing additional user-defined parameters.

To address the challenge of handling multiple motion types, tracking methods [151, 152] use N_m Kalman filters in parallel, each associated with a distinct dynamical model defined by its transition matrix $\mathbf{F}^{\theta_{t+1}}$, where $\theta_t \in \Lambda = [\![1; N_m]\!]$. The filtering process accounts for the sequence of model indices $\theta_{1:t}$ with transitions between models at times $\mathbb{T} = \{T_1, T_2, \dots T_{last}\}$ where each $T_i \in [\![1; T_{tot}]\!]$. The model indices are added to the Markovian scheme such that the motion model at time $t+1$ is independent of the state at time t, as illustrated on Fig. 6.5. This allows defining and estimating separate parameters for each motion model, which better captures the heterogeneity of motion and thus refines the Bayesian estimates.

Between two transitions at times T_i and T_{i+1}, the system is assumed to follow a homogeneous and stationary Markov process:

$$\mathbf{x}_{t+1} = \mathbf{F}^{\theta_{t+1}}\mathbf{x}_t + \mathbf{w}_{t+1}^{\theta_{t+1}}, \tag{6.15}$$

$$\mathbf{z}_{t+1} = \mathbf{H}^{\theta_{t+1}}\mathbf{x}_{t+1} + \mathbf{v}_{t+1}^{\theta_{t+1}}, \tag{6.16}$$

with $t \in [\![T_i; T_{i+1} - 1]\!]$ and where the Kalman filter variables follow the same indexed notation.

In practice, changes in the dynamical model are detected on-the-fly by testing if the new measurement contradicts the stationarity assumption. A time point $t+1$ is considered part of the transition time set $\mathbb{T}$ if:

$$\underset{\theta_i \in \Lambda}{\operatorname{argmax}}\, p(\mathbf{z}_{t+1}|\mathbf{z}_{T_i:t}, \theta_i) \neq \underset{\theta_i \in \Lambda}{\operatorname{argmax}}\, p(\theta_i|\mathbf{z}_{T_i:t}). \tag{6.17}$$

Mode stationarity is tested for each time point using this condition, resulting in a transition time set $\mathbb{T}_\theta$.

After a transition has been detected, say to the model θ_{t+1}, the filter is reinitialized using the parameters from the last time point $s < t+1$ such that $\theta_s = \theta_{t+1}$. Though theses parameters were appropriately learned to model the movement in a previous stationary regime of the dynamics, the first estimate will *a priori* not be representative of the new dynamics, which prevents accurate estimation of (i) transition times and (ii) motion parameters.

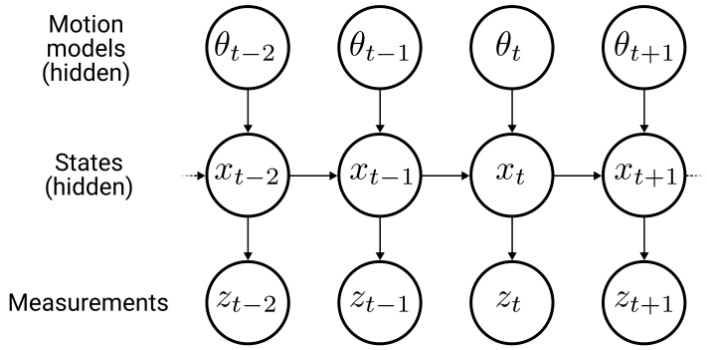

Figure 6.5: **Hierarchical schematics of the hidden Markov model used in the PMMS approach.** In classical Kalman filtering, the hidden state evolves based only on the value of the state at the previous frame. The measurements are generated from the states, and are conditionally independent from each other given the state. In the PMMS approach, a hidden motion model index is used to generate a hidden state based on the state at the previous frame.

Two-filter smoothing improves the estimation of transition times and motion parameters

To solve this problem, the PMMS framework implements a two-filter smoothing framework [213] which combines forward and backward Kalman filtering passes.
Three rounds of tracking in forward and backward directions are performed by reversing the temporal order of the image sequence. After the first round, the filter can be reinitialized using the parameter from the last round at time $t+1$ instead of the parameters at time s. Information from the previous round is also used to initialize the Kalman filter at the very beginning of the sequence, which alleviates having to find proper initialization parameters.

Additionally, the forward-backward smoothing can be used to find new transitions candidates by checking if two consecutive rounds converge on the same model

$$\underset{\theta_i \in \Lambda}{\operatorname{argmax}}\, p(\mathbf{z}_{t+1}|\mathbf{z}_{T_i:t}, \theta_i) = \underset{\theta_i \in \Lambda}{\operatorname{argmax}}\, p(\mathbf{z}_t|\mathbf{z}_{t+1:T_{i+1}}, \theta_i). \tag{6.18}$$

Time points that fail this condition define boundaries of independent motion regimes, contributing to the set $\mathbb{T}_{FB}$.

During the last round, the complete set of transitions is obtained by combining $\mathbb{T}_\theta$ and $\mathbb{T}_{FB}$. This integrated filtering and smoothing approach offers a cost-effective and robust solution for detecting transitions and filtering time sequences from trajectories that involve switches between dynamical models.

PMMS in the assignment framework

In the context of multiple object tracking, the PMMS algorithm does not handle the problem of conflicts between forward and backward passes in terms of measurement selection, i.e

if the backward filter pass suggests an assignment $j \to i$ which is inconsistent with the assignment $i \to j'$ (time direction is reversed between passes) suggested by the the prior forward filtering pass.

To apply the PMMS methodology in a MHT framework, the optimal measurement at each time point is selected based on information from both the forward and backward filters. When both filters select the same sequence, measurements are defined as:

$$\mathbf{z}_{t+1} = \underset{\mathbf{z} \in D_{t+1}}{\operatorname{argmax}}\, p(\mathbf{z}|\mathbf{z}_{T_i:t}, \hat{\theta}_i^f) \tag{6.19}$$

$$\mathbf{z}_t = \underset{\mathbf{z} \in D_t}{\operatorname{argmax}}\, p(\mathbf{z}|\mathbf{z}_{t+1:T_{i+1}}, \hat{\theta}_i^b) \tag{6.20}$$

where D_{t+1} is the set of particle detections at time $t+1$. If this condition is not met, the measurement with the highest likelihood is selected, and $t+1$ is treated as a boundary between stationary motion regimes. Even if modes are identical, differing measurements justify separate regimes. The transition times obtained this way are stored in the set $\mathrm{T}_{\mathbf{z}}$. When no measurement is selected, the state covariance matrix is set to infinity. In the last filtering round, the total transition time set is thus defined as $\mathbb{T} = \mathbb{T}_\theta \cup \mathbb{T}_{FB} \cup \mathbb{T}_{\mathbf{z}}$, with each regime smoothed independently.

Euclidean distance on domains with periodic boundary conditions

Our particle simulator allow simulation on a n-dimensional volume $\mathcal{V} = \prod_{i=1}^{n} [0, L^i]$ with periodic boundary conditions (PBC), i.e $\mathcal{V}$ is embed on a n-dimensional torus. Practically, this means that a particle moving beyond L_i in dimension i wraps back through position 0.

Using PBC requires modifying the computation of the euclidean distance between two particles i and j at positions $\mathbf{x}^i$ and $\mathbf{x}^j$, as it is now possible to wrap connections between two points beyond the limits of the volume. Euclidean distance in PBC can be computed efficiently using modulo arithmetic [214].
For a one-dimensional domain $[a, b]$, with $a < b$, we define $L = b - a$ and the boundary wrapping operator, which wraps arbitrary coordinates back to their value inside the interval, as:

$$[\cdot] : \begin{cases} \mathbb{R} \to [a, b] \\ x \mapsto ((x + L/2) \bmod L) - L/2 \end{cases} \tag{6.21}$$

For n-dimensional vector positions, the n-dimensional boundary wrapping operator $[\cdot]$ simply applies the one-dimensional wrapping operator to each component of the vector using their respective value of the interval length L^i.

The wrapping operator provides a concise way to right and compute the shortest vector

going from i to j, which we shall note $\mathbf{v}^{i\to j}$, as

$$\mathbf{v}^{i\to j} = \mathbf{x}^j - \mathbf{x}^i \quad \longrightarrow \quad \mathbf{v}^{i\to j}_{PBC} = \left[\mathbf{x}^j - \mathbf{x}^i\right] \tag{6.22}$$

Similarly, the scalar euclidean distance $d^{i,j}$ between i and j becomes

$$d^{i,j} = \|\mathbf{x}^j - \mathbf{x}^i\| \quad \longrightarrow \quad d^{i,j}_{PBC} = \|\left[\mathbf{x}^j - \mathbf{x}^i\right]\| \tag{6.23}$$

References

[1] Claudio Collinet et al. “Programmed and self-organized flow of information during morphogenesis”. In: *Nature Reviews Molecular Cell Biology* 22.4 (Jan. 2021), pp. 245–265. ISSN: 1471-0080. DOI: 10.1038/s41580-020-00318-6.

[2] Randy D. Mehlenbacher et al. “Nanomaterials for in vivo imaging of mechanical forces and electrical fields”. In: *Nature Reviews Materials* 3.2 (Dec. 2017). ISSN: 2058-8437. DOI: 10.1038/natrevmats.2017.80.

[3] Michael B. Elowitz et al. “Stochastic Gene Expression in a Single Cell”. In: *Science* 297.5584 (Aug. 2002), pp. 1183–1186. ISSN: 1095-9203. DOI: 10.1126/science.1070919.

[4] Hiroshi Ochiai et al. “Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells”. In: *Scientific Reports* 4.1 (Nov. 2014). ISSN: 2045-2322. DOI: 10.1038/srep07125.

[5] Quinton Smith et al. “Stochasticity and Spatial Interaction Govern Stem Cell Differentiation Dynamics”. In: *Scientific Reports* 5.1 (July 2015). ISSN: 2045-2322. DOI: 10.1038/srep12617.

[6] Thomas Lecuit et al. “Orchestrating size and shape during morphogenesis”. In: *Nature* 450.7167 (Nov. 2007), pp. 189–192. ISSN: 1476-4687. DOI: 10.1038/nature06304.

[7] X. Chen et al. “Transcriptional Regulatory Networks in Embryonic Stem Cells”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 73.0 (Jan. 2008), pp. 203–209. ISSN: 1943-4456. DOI: 10.1101/sqb.2008.73.026.

[8] Maria-Elena Torres-Padilla et al. “Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage”. In: *Development* 141.11 (June 2014), pp. 2173–2181. ISSN: 0950-1991. DOI: 10.1242/dev.102624.

[9] Peter S. Swain et al. “Intrinsic and extrinsic contributions to stochasticity in gene expression”. In: *Proceedings of the National Academy of Sciences* 99.20 (Sept. 2002), pp. 12795–12800. ISSN: 1091-6490. DOI: 10.1073/pnas.162041399.

[10] Karin E. de Visser et al. “The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth”. In: *Cancer Cell* 41.3 (Mar. 2023), pp. 374–403. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2023.02.016.

[11] David B. Brückner et al. “Stochastic nonlinear dynamics of confined cell migration in two-state systems”. In: *Nature Physics* 15.6 (Mar. 2019), pp. 595–601. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0445-4.

[12] Qi Chen et al. “Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo”. In: *Nature Communications* 9.1 (May 2018). ISSN: 2041-1723. DOI: 10.1038/s41467-018-04155-2.

[13] Shinya Inoué et al. “Force Generation by Microtubule Assembly/Disassembly in Mitosis and Related Movements”. In: *Molecular Biology of the Cell* 6.12 (Dec. 1995), pp. 1619–1640. ISSN: 1939-4586. DOI: 10.1091/mbc.6.12.1619.

[14] Ido Lavi et al. “Deterministic patterns in cell motility”. In: *Nature Physics* 12.12 (Aug. 2016), pp. 1146–1152. ISSN: 1745-2481. DOI: 10.1038/nphys3836.

[15] Jun Yan et al. “Stochastic variations of migration speed between cells in clonal populations”. In: *TECHNOLOGY* 02.03 (Sept. 2014), pp. 185–188. ISSN: 2345-7740. DOI: 10.1142/s2339547814200027.

[16] Aline Xavier da Silveira dos Santos et al. “From single cells to tissue self-organization”. In: *The FEBS Journal* 286.8 (Nov. 2018), pp. 1495–1513. ISSN: 1742-4658. DOI: 10.1111/febs.14694.

[17] James Briscoe et al. “Morphogen rules: design principles of gradient-mediated embryo patterning”. In: *Development* 142.23 (Dec. 2015), pp. 3996–4009. ISSN: 0950-1991. DOI: 10.1242/dev.129452.

[18] Alan Mathison Turing. “The chemical basis of morphogenesis”. In: *Bulletin of mathematical biology* 52 (1990), pp. 153–197.

[19] Jeremy B. A. Green et al. “Positional information and reaction-diffusion: two big ideas in developmental biology combine”. In: *Development* 142.7 (Apr. 2015), pp. 1203–1211. ISSN: 0950-1991. DOI: 10.1242/dev.114991.

[20] J. Raspopovic et al. “Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients”. In: *Science* 345.6196 (Aug. 2014), pp. 566–570. ISSN: 1095-9203. DOI: 10.1126/science.1252960.

[21] Alicia M Salvi et al. “Mechanisms linking mechanotransduction and cell metabolism”. In: *Current Opinion in Cell Biology* 54 (Oct. 2018), pp. 114–120. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2018.05.004.

[22] Timothy Fulton et al. “The unappreciated generative role of cell movements in pattern formation”. In: *Royal Society Open Science* 9.4 (Apr. 2022). ISSN: 2054-5703. DOI: 10.1098/rsos.211293.

[23] Marion Louveaux et al. “Cell division plane orientation based on tensile stress inArabidopsis thaliana”. In: *Proceedings of the National Academy of Sciences* 113.30 (July 2016). ISSN: 1091-6490. DOI: 10.1073/pnas.1600677113.

[24] Raphaël Etournay et al. “Interplay of cell dynamics and epithelial tension during morphogenesis of the Drosophila pupal wing”. In: *eLife* 4 (June 2015). ISSN: 2050-084X. DOI: 10.7554/elife.07090.

[25] Mehdi Saadaoui et al. “A tensile ring drives tissue flows to shape the gastrulating amniote embryo”. In: (Sept. 2018). DOI: 10.1101/412767.

[26] Shao-Zhen Lin et al. “Structure and Rheology in Vertex Models under Cell-Shape-Dependent Active Stresses”. In: *Physical Review Letters* 130.5 (Feb. 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.058202.

[27] Simon Gsell et al. “Marangoni-like tissue flows enhance symmetry breaking of embryonic organoids”. In: *Nature Physics* 21.4 (Mar. 2025), pp. 644–653. ISSN: 1745-2481. DOI: 10.1038/s41567-025-02802-2.

[28] Diana Pinheiro et al. “Morphogen gradient orchestrates pattern-preserving tissue morphogenesis via motility-driven unjamming”. In: *Nature Physics* 18.12 (Oct. 2022), pp. 1482–1493. ISSN: 1745-2481. DOI: 10.1038/s41567-022-01787-6.

[29] Prachiti Moghe et al. “Coupling of cell shape, matrix and tissue dynamics ensures embryonic patterning robustness”. In: *Nature Cell Biology* 27.3 (Feb. 2025), pp. 408–423. ISSN: 1476-4679. DOI: 10.1038/s41556-025-01618-9.

[30] Lilianna Solnica-Krezel. *Gastrulation: From embryonic pattern to form*. Vol. 136. Academic Press, 2020.

[31] Pierre-François Lenne et al. “Learning the mechanobiology of development from gastruloids”. In: *Emerging Topics in Life Sciences* 7.4 (Dec. 2023), pp. 417–422. ISSN: 2397-8562. DOI: 10.1042/etls20230081.

[32] Yasmine el Azhar et al. “Development in a Dish—In Vitro Models of Mammalian Embryonic Development”. In: *Frontiers in Cell and Developmental Biology* 9 (May 2021). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.655993.

[33] Yunhee Kim et al. “A new era of stem cell and developmental biology: from blastoids to synthetic embryos and beyond”. In: *Experimental & Molecular Medicine* 55.10 (Oct. 2023), pp. 2127–2137. ISSN: 2092-6413. DOI: 10.1038/s12276-023-01097-8.

[34] Leonardo Beccari et al. “Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids”. In: *Nature* 562.7726 (Oct. 2018), pp. 272–276. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0578-0.

[35] Martin Estermann. “Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites”. In: (Apr. 2020). DOI: 10.1242/prelights.18906.

[36] Ali Hashmi et al. “Cell-state transitions and collective cell movement generate an endoderm-like region in gastruloids”. In: *eLife* 11 (Apr. 2022). ISSN: 2050-084X. DOI: 10.7554/elife.59371.

[37] Laurent Argiro et al. “Gastruloids are competent to specify both cardiac and skeletal muscle lineages”. In: *Nature Communications* 15.1 (Nov. 2024). ISSN: 2041-1723. DOI: 10.1038/s41467-024-54466-w.

[38] Susanne C. van den Brink et al. “3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation”. In: *Trends in Cell Biology* 31.9 (Sept. 2021), pp. 747–759. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2021.06.007.

[39] Ravian L van Ineveld et al. “3D imaging for driving cancer discovery”. In: *The EMBO Journal* 41.10 (Apr. 2022). ISSN: 1460-2075. DOI: 10.15252/embj.2021109675.

[40] Bechara Saykali et al. “Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo”. In: (Oct. 2018). DOI: 10.1101/438721.

[41] Martin H. Dominguez et al. “Graded mesoderm assembly governs cell fate and morphogenesis of the early mammalian heart”. In: *Cell* 186.3 (Feb. 2023), 479–496.e23. ISSN: 0092-8674. DOI: 10.1016/j.cell.2023.01.001.

[42] Anna Labernadie et al. “A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion”. In: *Nature Cell Biology* 19.3 (Feb. 2017), pp. 224–237. ISSN: 1476-4679. DOI: 10.1038/ncb3478.

[43] Anna Bigas et al. “Using Pluripotent Stem Cells to Understand Normal and Leukemic Hematopoietic Development”. In: *Stem Cells Translational Medicine* 11.11 (Oct. 2022), pp. 1123–1134. ISSN: 2157-6580. DOI: 10.1093/stcltm/szac071.

[44] Eleonora Stronati et al. “YAP1 regulates the self-organized fate patterning of hESC-derived gastruloids”. In: *Stem Cell Reports* 17.2 (Feb. 2022), pp. 211–220. ISSN: 2213-6711. DOI: 10.1016/j.stemcr.2021.12.012.

[45] Denise Ragusa et al. “Dissecting infant leukemia developmental origins with a hemogenic gastruloid model”. In: (Feb. 2025). DOI: 10.7554/elife.102324.1.

[46] Fernando Amat et al. “Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data”. In: *Nature Methods* 11.9 (July 2014), pp. 951–958. ISSN: 1548-7105. DOI: 10.1038/nmeth.3036.

[47] Katie McDole et al. “In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level”. In: *Cell* 175.3 (Oct. 2018), 859–876.e33. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.09.031.

[48] Merlin Lange et al. “Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors”. In: (Mar. 2023). DOI: 10.1101/2023.03.06.531398.

[49] Franziska Moos et al. “Open-top multisample dual-view light-sheet microscope for live imaging of large multicellular systems”. In: *Nature Methods* 21.5 (Mar. 2024), pp. 798–803. ISSN: 1548-7105. DOI: 10.1038/s41592-024-02213-w.

[50] Mélody Merle et al. “Precise and scalable self-organization in mammalian pseudo-embryos”. In: *Nature Structural & Molecular Biology* 31.6 (Mar. 2024), pp. 896–902. ISSN: 1545-9985. DOI: 10.1038/s41594-024-01251-4.

[51] David Oriola et al. “Arrested coalescence of multicellular aggregates”. In: *Soft Matter* 18.19 (2022), pp. 3771–3780. ISSN: 1744-6848. DOI: 10.1039/d2sm00063f.

[52] Tai Ma et al. “PIViT: Large Deformation Image Registration with Pyramid-Iterative Vision Transformer”. In: (2023), pp. 602–612. ISSN: 1611-3349. DOI: 10.1007/978-3-031-43999-5_57.

[53] Xavier Trepat et al. “Mesoscale physical principles of collective cell organization”. In: *Nature Physics* 14.7 (July 2018), pp. 671–682. ISSN: 1745-2481. DOI: 10.1038/s41567-018-0194-9.

[54] Sebastian J Streichan et al. “Global morphogenetic flow is accurately predicted by the spatial distribution of myosin motors”. In: *eLife* 7 (Feb. 2018). ISSN: 2050-084X. DOI: 10.7554/elife.27454.

[55] Brian DePasquale et al. “Neural population dynamics underlying evidence accumulation in multiple rat brain regions”. In: *eLife* 13 (Aug. 2024). ISSN: 2050-084X. DOI: 10.7554/elife.84955.

[56] David B. Brückner et al. “Inferring the Dynamics of Underdamped Stochastic Systems”. In: *Physical Review Letters* 125.5 (July 2020). ISSN: 1079-7114. DOI: 10.1103/physrevlett.125.058103.

[57] David B. Brückner et al. “Learning the dynamics of cell–cell interactions in confined cell migration”. In: *Proceedings of the National Academy of Sciences* 118.7 (Feb. 2021). ISSN: 1091-6490. DOI: 10.1073/pnas.2016602118.

[58] Elizabeth M.C. Hillman et al. “Light-Sheet Microscopy in Neuroscience”. In: *Annual Review of Neuroscience* 42.1 (July 2019), pp. 295–313. ISSN: 1545-4126. DOI: 10.1146/annurev-neuro-070918-050357.

[59] Varun Mannam et al. “Machine learning for faster and smarter fluorescence lifetime imaging microscopy”. In: *Journal of Physics: Photonics* 2.4 (Sept. 2020), p. 042005. ISSN: 2515-7647. DOI: 10.1088/2515-7647/abac1a.

[60] Xuan Zheng et al. “Organoid cell fate dynamics in space and time”. In: *Science Advances* 9.33 (Aug. 2023). ISSN: 2375-2548. DOI: 10.1126/sciadv.add6480.

[61] Gustavo de Medeiros et al. “Multiscale light-sheet organoid imaging framework”. In: *Nature Communications* 13.1 (Aug. 2022). ISSN: 2041-1723. DOI: 10.1038/s41467-022-32465-z.

[62] Anne Beghin et al. “Automated high-speed 3D imaging of organoid cultures with multi-scale phenotypic quantification”. In: *Nature Methods* 19.7 (June 2022), pp. 881–892. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01508-0.

[63] Ina Maria Schießl et al. “Deep insights: intravital imaging with two-photon microscopy”. In: *Pflügers Archiv-European Journal of Physiology* 468 (2016), pp. 1505–1516.

[64] Fritjof Helmchen et al. “Deep tissue two-photon microscopy”. In: *Nature Methods* 2.12 (Nov. 2005), pp. 932–940. ISSN: 1548-7105. DOI: 10.1038/nmeth818.

[65] Chad M. Hobson et al. “Practical considerations for quantitative light sheet fluorescence microscopy”. In: *Nature Methods* 19.12 (Oct. 2022), pp. 1538–1549. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01632-x.

[66] Lina Streich et al. “High-resolution structural and functional deep brain imaging using adaptive optics three-photon microscopy”. In: BRAIN (2021), BM1B.3. DOI: 10.1364/brain.2021.bm1b.3.

[67] Pierre Mahou et al. “Multicolor two-photon light-sheet microscopy”. In: *Nature Methods* 11.6 (May 2014), pp. 600–601. ISSN: 1548-7105. DOI: 10.1038/nmeth.2963.

[68] Mitsutoshi Ataka et al. “Multibeam continuous axial scanning two-photon microscopy for in vivo volumetric imaging in mouse brain”. In: *Biomedical Optics Express* 15.2 (Jan. 2024), p. 1089. ISSN: 2156-7085. DOI: 10.1364/boe.514826.

[69] Chi-Li Chiu et al. “napari: a Python Multi-Dimensional Image Viewer Platform for the Research Community”. In: *Microscopy and Microanalysis* 28.S1 (Aug. 2022), pp. 1576–1577. ISSN: 1431-9276. DOI: 10.1017/s1431927622006328.

[70] Alice Gros et al. “A quantitative pipeline for whole-mount deep imaging and multiscale analysis of gastruloids”. In: (Aug. 2024). DOI: 10.1101/2024.08.13.607832.

[71] Benjamin Seelbinder et al. “Nuclear deformation guides chromatin reorganization in cardiac development and disease”. In: *Nature Biomedical Engineering* 5.12 (Dec. 2021), pp. 1500–1516. ISSN: 2157-846X. DOI: 10.1038/s41551-021-00823-9.

[72] Marie Versaevel et al. “Spatial coordination between cell and nuclear shape within micropatterned endothelial cells”. In: *Nature Communications* 3.1 (Feb. 2012). ISSN: 2041-1723. DOI: 10.1038/ncomms1668.

[73] Khuloud Jaqaman et al. “Robust single-particle tracking in live-cell time-lapse sequences”. In: *Nature Methods* 5.8 (July 2008), pp. 695–702. ISSN: 1548-7105. DOI: 10.1038/nmeth.1237.

[74] Kristina Ulicna et al. “Automated deep lineage tree analysis using a Bayesian single cell tracking approach”. In: (Sept. 2020). DOI: 10.1101/2020.09.10.276980.

[75] Philippe Roudot et al. “u-track3D: Measuring, navigating, and validating dense particle trajectories in three dimensions”. In: *Cell Reports Methods* 3.12 (Dec. 2023), p. 100655. ISSN: 2667-2375. DOI: 10.1016/j.crmeth.2023.100655.

[76] Norberto Malpica et al. “Applying watershed algorithms to the segmentation of clustered nuclei”. In: *Cytometry* 28.4 (Dec. 1998), pp. 289–297. ISSN: 0196-4763. DOI: 10.1002/(sici)1097-0320(19970801)28:4<289::aid-cyto3>3.0.co;2-7.

[77] Juan C. Caicedo et al. “Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images”. In: *Cytometry Part A* 95.9 (July 2019), pp. 952–965. ISSN: 1552-4930. DOI: 10.1002/cyto.a.23863.

[78] Cyril Achard et al. “CellSeg3D: self-supervised 3D cell segmentation for microscopy”. In: (Sept. 2024). DOI: 10.7554/elife.99848.1.

[79] Xinpeng Xie et al. “Instance-Aware Self-supervised Learning for Nuclei Segmentation”. In: (2020), pp. 341–350. ISSN: 1611-3349. DOI: 10.1007/978-3-030-59722-1_33.

[80] Michael C. Robitaille et al. “Self-supervised machine learning for live cell imagery segmentation”. In: *Communications Biology* 5.1 (Nov. 2022). ISSN: 2399-3642. DOI: 10.1038/s42003-022-04117-x.

[81] Alexander Kirillov et al. “Segment Anything”. In: (Oct. 2023). DOI: 10.1109/iccv51070.2023.00371.

[82] Anwai Archit et al. “Segment Anything for Microscopy”. In: *Nature Methods* 22.3 (Feb. 2025), pp. 579–591. ISSN: 1548-7105. DOI: 10.1038/s41592-024-02580-4.

[83] Stephan Preibisch et al. “Globally optimal stitching of tiled 3D microscopic image acquisitions”. In: *Bioinformatics* 25.11 (2009).

[84] David Horl et al. “BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples”. In: *Nature methods* 16.9 (2019).

[85] Stefan Klein et al. “Elastix: a toolbox for intensity-based medical image registration”. In: *IEEE transactions on medical imaging* 29.1 (2009).

[86] Christian Tischer. “ElastixWrapper: Fiji plugin for 3D image registration with elastix”. In: *Zenodo* 10 (2019).

[87] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature methods* 9.7 (2012).

[88] Sébastien Ourselin et al. “Block matching: A general framework to improve robustness of rigid registration of medical images”. In: (2000), pp. 557–566.

[89] Léo Guignard et al. “Contact area–dependent cell communication and the morphological invariance of ascidian embryogenesis”. In: *Science* 369.6500 (July 2020). ISSN: 1095-9203. DOI: 10.1126/science.aar5663.

[90] Friedrich Preusser et al. “FRC-QE: a robust and comparable 3D microscopy image quality metric for cleared organoids”. In: *Bioinformatics* 37.18 (Mar. 2021). Ed. by Peter Robinson, pp. 3088–3090. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btab160.

[91] Martin Weigert et al. “Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy”. In: (Mar. 2020), pp. 3655–3662. DOI: 10.1109/wacv45572.2020.9093435.

[92] Olaf Ronneberger et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: (2015), pp. 234–241. ISSN: 1611-3349. DOI: 10.1007/978-3-319-24574-4_28.

[93] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: (June 2016). DOI: 10.1109/cvpr.2016.90.

[94] Liuhao Ge et al. “3D Convolutional Neural Networks for Efficient and Robust Hand Pose Estimation from Single Depth Images”. In: (July 2017), pp. 5679–5688. DOI: 10.1109/cvpr.2017.602.

[95] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. In: (Feb. 2020). DOI: 10.1101/2020.02.02.931238.

[96] *stardist/augmend*. 2023. URL: https://github.com/stardist/augmend (visited on 04/30/2025).

[97] Jules Vanaret et al. “A Detector-Independent Quality Score for Cell Segmentation Without Ground Truth in 3D Live Fluorescence Microscopy”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 29.4: Biophotonics (July 2023), pp. 1–12. ISSN: 1558-4542. DOI: 10.1109/jstqe.2023.3275108.

[98] Amin Zehtabian et al. “Precise measurement of nanoscopic septin ring structures in deep learning-assisted quantitative superresolution microscopy”. In: (Dec. 2021). DOI: 10.1101/2021.12.28.474382.

[99] Justin Sonneck et al. “On the risk of manual annotations in 3D confocal microscopy image segmentation”. In: (Oct. 2023), pp. 3896–3904. DOI: 10.1109/iccvw60793.2023.00421.

[100] Taryn A. Donovan et al. “Mitotic Figures—Normal, Atypical, and Imposters: A Guide to Identification”. In: *Veterinary Pathology* 58.2 (Dec. 2020), pp. 243–257. ISSN: 1544-2217. DOI: 10.1177/0300985820980049.

[101] Ko Sugawara. “Training deep learning models for cell image segmentation with sparse annotations”. In: *BioRxiv* (June 2023). DOI: 10.1101/2023.06.13.544786.

[102] Thibaut Goldsborough et al. “InstanSeg: an embedding-based instance segmentation algorithm optimized for accurate, efficient and portable cell segmentation”. In: *arXiv preprint arXiv:2408.15954* (2024).

[103] Manan Lalit et al. “EmbedSeg: Embedding-based Instance Segmentation for Biomedical Microscopy Data”. In: *Medical Image Analysis* 81 (Oct. 2022), p. 102523. ISSN: 1361-8415. DOI: 10.1016/j.media.2022.102523.

[104] Felix Yuran Zhou et al. “Universal consensus 3D segmentation of cells from 2D segmented stacks”. In: (May 2024). DOI: 10.1101/2024.05.03.592249.

[105] Alessandro Mongera et al. “Mechanics of the cellular microenvironment as probed by cells in vivo during zebrafish presomitic mesoderm differentiation”. In: *Nature Materials* 22.1 (Dec. 2022), pp. 135–143. ISSN: 1476-4660. DOI: 10.1038/s41563-022-01433-9.

[106] M.J. BISSELL et al. “Microenvironmental Regulators of Tissue Structure and Function Also Regulate Tumor Induction and Progression: The Role of Extracellular Matrix and Its Degrading Enzymes”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 70.0 (Jan. 2005), pp. 343–356. ISSN: 1943-4456. DOI: 10.1101/sqb.2005.70.013.

[107] Kristaps Kairišs et al. “Visualisation of gene expression within the context of tissues using an X-ray computed tomography-based multimodal approach”. In: *Scientific Reports* 14.1 (Apr. 2024). ISSN: 2045-2322. DOI: 10.1038/s41598-024-58766-5.

[108] Mehdi Ebrahimi et al. “Organoids”. In: (2020), pp. 35–52. DOI: 10.1007/978-3-030-41476-4_4.

[109] Yohalie Kalukula et al. “Mechanics and functional consequences of nuclear deformations”. In: *Nature Reviews Molecular Cell Biology* 23.9 (May 2022), pp. 583–602. ISSN: 1471-0080. DOI: 10.1038/s41580-022-00480-z.

[110] Ian D. Estabrook et al. “Calculation of the force field required for nucleus deformation during cell migration through constrictions”. In: (Dec. 2020). DOI: 10.1101/2020.12.17.423200.

[111] Simon Suppinger et al. “Multimodal characterization of murine gastruloid development”. In: *Cell Stem Cell* 30.6 (June 2023), 867–884.e11. ISSN: 1934-5909. DOI: 10.1016/j.stem.2023.04.018.

[112] Alba Villaronga Luque et al. “Integrated Molecular-Phenotypic Profiling Reveals Metabolic Control of Morphological Variation in Stembryos”. In: (Dec. 2023). DOI: 10.1101/2023.12.04.569921.

[113] Susanne C. van den Brink et al. “Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells”. In: *Development* 141.22 (Nov. 2014), pp. 4231–4242. ISSN: 0950-1991. DOI: 10.1242/dev.113001.

[114] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17.3 (Feb. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2.

[115] F. Graner et al. “Discrete rearranging disordered patterns, part I: Robust statistical tools in two or three dimensions”. In: *The European Physical Journal E* 25.4 (Apr. 2008), pp. 349–369. ISSN: 1292-895X. DOI: 10.1140/epje/i2007-10298-8.

[116] Georgina A Stooke-Vaughan et al. “Physical control of tissue morphogenesis across scales”. In: *Current Opinion in Genetics & Development* 51 (Aug. 2018), pp. 111–119. ISSN: 0959-437X. DOI: 10.1016/j.gde.2018.09.002.

[117] Chii J. Chan et al. “Coordination of Morphogenesis and Cell-Fate Specification in Development”. In: *Current Biology* 27.18 (Sept. 2017), R1024–R1035. ISSN: 0960-9822. DOI: 10.1016/j.cub.2017.07.010.

[118] Sham Tlili et al. “Colloquium: Mechanical formalisms for tissue dynamics”. In: *The European Physical Journal E* 38.5 (May 2015). ISSN: 1292-895X. DOI: 10.1140/epje/i2015-15033-4.

[119] S. Tlili et al. “Migrating Epithelial Monolayer Flows Like a Maxwell Viscoelastic Liquid”. In: *Physical Review Letters* 125.8 (Aug. 2020). ISSN: 1079-7114. DOI: 10.1103/physrevlett.125.088102.

[120] Elsa Abranches et al. “Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression”. In: *PLoS ONE* 8.3 (Mar. 2013). Ed. by Qiang Wu, e59928. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0059928.

[121] Nicola Gritti et al. “MOrgAna: accessible quantitative analysis of organoids with machine learning”. In: *Development* 148.18 (Sept. 2021). ISSN: 1477-9129. DOI: 10.1242/dev.199611.

[122] Jesús A. Andrés-San Román et al. “CartoCell, a high-content pipeline for 3D image analysis, unveils cell morphology patterns in epithelia”. In: (Jan. 2023). DOI: 10.1101/2023.01.05.522724.

[123] Lucas von Chamier et al. “Democratising deep learning for microscopy with ZeroCostDL4Mic”. In: *Nature Communications* 12.1 (Apr. 2021). ISSN: 2041-1723. DOI: 10.1038/s41467-021-22518-0.

[124] K Somani Arun et al. “Least-squares fitting of two 3-D point sets”. In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pp. 698–700.

[125] Laura Zigutyte et al. *napari-clusters-plotter*. Aug. 2023. DOI: 10.5281/zenodo.8238460. URL: https://zenodo.org/records/8238460 (visited on 09/19/2024).

[126] Elisabete C. Costa et al. “3D tumor spheroids: an overview on the tools and techniques used for their analysis”. In: *Biotechnology Advances* 34.8 (Dec. 2016), pp. 1427–1441. ISSN: 0734-9750. DOI: 10.1016/j.biotechadv.2016.11.002.

[127] Marie Schott et al. “Open-ST: High-resolution spatial transcriptomics in 3D”. In: *Cell* 187.15 (July 2024), 3953–3972.e26. ISSN: 0092-8674. DOI: 10.1016/j.cell.2024.05.055.

[128] Adrian Wolny et al. “Accurate and Versatile 3D Segmentation of Plant Tissues at Cellular Resolution”. In: (Jan. 2020). DOI: 10.1101/2020.01.17.910562.

[129] In: (2017), pp. 1054–1054. DOI: 10.1007/978-1-4899-7687-1_695.

[130] Matthias Arzt et al. “LABKIT: Labeling and Segmentation Toolkit for Big Image Data”. In: *Frontiers in Computer Science* 4 (Feb. 2022), p. 777728. ISSN: 26249898. DOI: 10.3389/FCOMP.2022.777728/BIBTEX.

[131] Stuart Berg et al. “ilastik: interactive machine learning for (bio)image analysis”. In: *Nature Methods* 16.12 (Sept. 2019), pp. 1226–1232. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0582-9.

[132] Robert Haase et al. *haesleinhuepf/napari-accelerated-pixel-and-object- classification: 0.14.1*. Version 0.14.1. Nov. 2023. DOI: 10.5281/zenodo.10071078. URL: https://doi.org/10.5281/zenodo.10071078.

[133] Anna-Kristina Marel et al. "Flow and Diffusion in Channel-Guided Cell Migration". In: *Biophysical Journal* 107.5 (Sept. 2014), pp. 1054–1064. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2014.07.017.

[134] Michael J. Shannon et al. "cellPLATO: an unsupervised method for identifying cell behaviour in heterogeneous cell trajectory data". In: (Oct. 2023). DOI: 10.1101/2023.10.28.564355.

[135] Marc Karnat et al. "Inferring the location and orientation of cell divisions on time-lapse image sequences". In: (Feb. 2024). DOI: 10.1101/2024.02.05.578987.

[136] Max Hopkins et al. "Realizable Learning is All You Need". In: *TheoretiCS* Volume 3 (Feb. 2024). ISSN: 2751-4838. DOI: 10.46298/theoretics.24.2.

[137] Jeffrey M. Perkel. "Python power-up: new image tool visualizes complex data". In: *Nature* 600.7888 (Dec. 2021), pp. 347–348. ISSN: 1476-4687. DOI: 10.1038/d41586-021-03628-7.

[138] Robert Haase et al. "A Hitchhiker's guide through the bio-image analysis software universe". In: *FEBS Letters* 596.19 (July 2022), pp. 2472–2485. ISSN: 1873-3468. DOI: 10.1002/1873-3468.14451.

[139] Jean-Yves Tinevez et al. "TrackMate: An open and extensible platform for single-particle tracking". In: *Methods* 115 (Feb. 2017), pp. 80–90. ISSN: 1046-2023. DOI: 10.1016/j.ymeth.2016.09.016.

[140] Daniel Franco-Barranco et al. "BiaPy: Accessible deep learning on bioimages". In: (Feb. 2024). DOI: 10.1101/2024.02.03.576026.

[141] Dimitri Fabrèges et al. "Temporal variability and cell mechanics control robustness in mammalian embryogenesis". In: (Jan. 2023). DOI: 10.1101/2023.01.24.525420.

[142] Donald Reid. "An algorithm for tracking multiple targets". In: *IEEE transactions on Automatic Control* 24.6 (1979), pp. 843–854.

[143] John C. Crocker et al. "Methods of Digital Video Microscopy for Colloidal Studies". In: *Journal of Colloid and Interface Science* 179.1 (Apr. 1996), pp. 298–310. ISSN: 0021-9797. DOI: 10.1006/jcis.1996.0217.

[144] Zhirong Bao et al. "Automated cell lineage tracing in Caenorhabditis elegans". In: *Proceedings of the National Academy of Sciences* 103.8 (Feb. 2006), pp. 2707–2712. ISSN: 1091-6490. DOI: 10.1073/pnas.0511111103.

[145] Rutger N.U. Kok et al. "OrganoidTracker: efficient cell tracking using machine learning and manual error correction". In: (Mar. 2020). DOI: 10.1101/2020.03.18.996421.

[146] Tal Ben-Haim et al. “Graph Neural Network for Cell Tracking in Microscopy Videos”. In: (2022), pp. 610–626. ISSN: 1611-3349. DOI: 10.1007/978-3-031-19803-8_36.

[147] Marco Baldovin et al. “Langevin equations from experimental data: The case of rotational diffusion in granular media”. In: *PLOS ONE* 14.2 (Feb. 2019). Ed. by Alberto d’Onofrio, e0212135. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0212135.

[148] Lucy Ham et al. “The chemical Langevin equation for biochemical systems in dynamic environments”. In: *The Journal of Chemical Physics* 157.9 (Sept. 2022). ISSN: 1089-7690. DOI: 10.1063/5.0095840.

[149] Manfred Schienbein et al. “Langevin equation, Fokker-Planck equation and cell migration”. In: *Bulletin of Mathematical Biology* 55.3 (1993), pp. 585–608.

[150] Kevin M. Dean et al. “Diagonally Scanned Light-Sheet Microscopy for Fast Volumetric Imaging of Adherent Cells”. In: *Biophysical Journal* 110.6 (Mar. 2016), pp. 1456–1465. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2016.01.029.

[151] M. Yeddanapudi et al. “IMM estimation for multitarget-multisensor air traffic surveillance”. In: *Proceedings of the IEEE* 85.1 (1997), pp. 80–96. ISSN: 0018-9219. DOI: 10.1109/5.554210.

[152] Yaakov Bar-Shalom et al. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. Wiley, Jan. 2002. ISBN: 9780471221272. DOI: 10.1002/0471221279.

[153] Roy Jonker et al. “A shortest augmenting path algorithm for dense and sparse linear assignment problems”. In: (1988), pp. 622–622.

[154] Jerome H. Friedman et al. “An Algorithm for Finding Best Matches in Logarithmic Expected Time”. In: *ACM Transactions on Mathematical Software* 3.3 (Sept. 1977), pp. 209–226. ISSN: 1557-7295. DOI: 10.1145/355744.355745.

[155] Ivo F Sbalzarini et al. “Feature point tracking and trajectory analysis for video imaging in cell biology”. In: *Journal of structural biology* 151.2 (2005), pp. 182–195.

[156] Ryoma Bise et al. “Reliable cell tracking by global data association”. In: (Mar. 2011). DOI: 10.1109/isbi.2011.5872571.

[157] Léo Guignard et al. “Contact area–dependent cell communication and the morphological invariance of ascidian embryogenesis”. In: *Science* 369.6500 (July 2020). ISSN: 1095-9203. DOI: 10.1126/science.aar5663.

[158] Ihor Smal et al. “Quantitative comparison of multiframe data association techniques for particle tracking in time-lapse fluorescence microscopy”. In: *Medical Image Analysis* 24.1 (Aug. 2015), pp. 163–189. ISSN: 1361-8415. DOI: 10.1016/j.media.2015.06.006.

[159] Linqing Feng et al. “Multiple dense particle tracking in fluorescence microscopy images based on multidimensional assignment”. In: *Journal of Structural Biology* 173.2 (Feb. 2011), pp. 219–228. ISSN: 1047-8477. DOI: 10.1016/j.jsb.2010.11.001.

[160] Nicolas Chenouard et al. “Multiple hypothesis tracking in microscopy images”. In: (June 2009), pp. 1346–1349. DOI: 10.1109/isbi.2009.5193314.

[161] Caroline Malin-Mayor et al. “Automated Reconstruction of Whole-Embryo Cell Lineages by Learning from Sparse Annotations”. In: (July 2021). DOI: 10.1101/2021.07.28.454016.

[162] Engin Turetken et al. “Network Flow Integer Programming to Track Elliptical Cells in Time-Lapse Sequences”. In: *IEEE Transactions on Medical Imaging* 36.4 (Apr. 2017), pp. 942–951. ISSN: 1558-254X. DOI: 10.1109/tmi.2016.2640859.

[163] Jordao Bragantini et al. “Ultrack: pushing the limits of cell tracking across biological scales”. In: (Sept. 2024). DOI: 10.1101/2024.09.02.610652.

[164] Peter Hirsch et al. “Tracking by Weakly-Supervised Learning and Graph Optimization for Whole-Embryo C. elegans lineages”. In: (2022), pp. 25–35. ISSN: 1611-3349. DOI: 10.1007/978-3-031-16440-8_3.

[165] Benjamin Gallusser et al. “TRACKASTRA: Transformer-Based Cell Tracking for Live-Cell Microscopy”. In: (Oct. 2024), pp. 467–484. ISSN: 1611-3349. DOI: 10.1007/978-3-031-73116-7_27.

[166] Jean-Baptiste Lugagne et al. “DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning”. In: (July 2019). DOI: 10.1101/720615.

[167] Deepak R Chittajallu et al. “In vivo cell-cycle profiling in xenograft tumors by quantitative intravital microscopy”. In: *Nature Methods* 12.6 (Apr. 2015), pp. 577–585. ISSN: 1548-7105. DOI: 10.1038/nmeth.3363.

[168] Philippe Roudot et al. “Piecewise-Stationary Motion Modeling and Iterative Smoothing to Track Heterogeneous Particle Motions in Dense Environments”. In: *IEEE Transactions on Image Processing* 26.11 (Nov. 2017), pp. 5395–5410. ISSN: 1941-0042. DOI: 10.1109/tip.2017.2707803.

[169] Tobias Pietzsch et al. “BigDataViewer: visualization and processing for large image data sets”. In: *Nature Methods* 12.6 (May 2015), pp. 481–483. ISSN: 1548-7105. DOI: 10.1038/nmeth.3392.

[170] Anne E Carpenter et al. “CellProfiler: image analysis software for identifying and quantifying cell phenotypes”. In: *Genome Biology* 7.10 (Oct. 2006). ISSN: 1474-760X. DOI: 10.1186/gb-2006-7-10-r100.

[171] Benjamin Schott et al. “EmbryoMiner: A new framework for interactive knowledge discovery in large-scale cell tracking data of developing embryos”. In: *PLOS Computational Biology* 14.4 (Apr. 2018). Ed. by Dina Schneidman, e1006128. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006128.

[172] Carsten Wolff et al. “Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb”. In: *eLife* 7 (Mar. 2018). ISSN: 2050-084X. DOI: 10.7554/elife.34410.

[173] Mohamed El Beheiry et al. “Virtual Reality: Beyond Visualization”. In: *Journal of Molecular Biology* 431.7 (Mar. 2019), pp. 1315–1321. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2019.01.033.

[174] Paul J. Hezel et al. *Head Mounted Displays for Virtual Reality*. Feb. 1993. DOI: 10.21236/ada263498.

[175] Teun A.P.M. Huijben et al. “inTRACKtive — A Web-Based Tool for Interactive Cell Tracking Visualization”. In: (Oct. 2024). DOI: 10.1101/2024.10.18.618998.

[176] Fernando Amat et al. “Efficient processing and analysis of large-scale light-sheet microscopy data”. In: *Nature Protocols* 10.11 (Oct. 2015), pp. 1679–1696. ISSN: 1750-2799. DOI: 10.1038/nprot.2015.111.

[177] Ko Sugawara et al. “Tracking cell lineages in 3D by incremental deep learning”. In: (Feb. 2022). DOI: 10.1101/2021.02.26.432552.

[178] Nicolas Chenouard et al. “Objective comparison of particle tracking methods”. In: *Nature Methods* 11.3 (Jan. 2014), pp. 281–289. ISSN: 1548-7105. DOI: 10.1038/nmeth.2808.

[179] William J Godinez et al. “Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences”. In: *Medical image analysis* 13.2 (2009), pp. 325–342.

[180] I. Smal et al. “Particle Filtering for Multiple Object Tracking in Dynamic Fluorescence Microscopy Images: Application to Microtubule Growth Analysis”. In: *IEEE Transactions on Medical Imaging* 27.6 (June 2008), pp. 789–804. ISSN: 0278-0062. DOI: 10.1109/tmi.2008.916964.

[181] Rudolph Emil Kalman. “A new approach to linear filtering and prediction problems”. In: *Journal of Basic Engineering* 1 (1960). DOI: 10.1115/1.3662552.

[182] Benjamin Favetto et al. “Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck process with biological application”. In: *Scandinavian Journal of Statistics* 37.2 (2010), pp. 200–220.

[183] Oliver Grothe. “A higher order correlation unscented Kalman filter”. In: *Applied Mathematics and Computation* 219.17 (May 2013), pp. 9033–9042. ISSN: 0096-3003. DOI: 10.1016/j.amc.2013.03.019.

[184] Arthur Gelb et al. “Applied optimal estimation”. In: (1974).

[185] Yoshio Kosuge et al. “The gate size estimation method and the optimal gate shape for target tracking”. In: *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 85.5 (Jan. 2002), pp. 10–22. ISSN: 1520-6440. DOI: 10.1002/ecjc.1090.

[186] Moritz Sturm et al. “SynCellFactory: Generative Data Augmentation for Cell Tracking”. In: (2024), pp. 304–313. ISSN: 1611-3349. DOI: 10.1007/978-3-031-72390-2_29.

[187] Steve Runser et al. “SimuCell3D: three-dimensional simulation of tissue mechanics with cell polarization”. In: *Nature Computational Science* 4.4 (Apr. 2024), pp. 299–309. ISSN: 2662-8457. DOI: 10.1038/s43588-024-00620-9.

[188] Philipp Germann et al. “ya||a: GPU-Powered Spheroid Models for Mesenchyme and Epithelium”. In: *Cell Systems* 8.3 (Mar. 2019), 261–266.e3. ISSN: 2405-4712. DOI: 10.1016/j.cels.2019.02.007.

[189] Timo Kuhn et al. “Single molecule tracking and analysis framework including theory-predicted parameter settings”. In: *Scientific Reports* 11.1 (May 2021). ISSN: 2045-2322. DOI: 10.1038/s41598-021-88802-7.

[190] Janick Cardinale et al. “Bayesian image analysis with on-line confidence estimates and its application to microtubule tracking”. In: (2009), pp. 1091–1094.

[191] Alexandru Niculescu-Mizil et al. “Predicting good probabilities with supervised learning”. In: ICML ’05 (2005), pp. 625–632. DOI: 10.1145/1102351.1102430.

[192] Vladimír Ulman et al. “An objective comparison of cell-tracking algorithms”. In: *Nature Methods* 14.12 (Oct. 2017), pp. 1141–1152. ISSN: 1548-7105. DOI: 10.1038/nmeth.4473.

[193] Meghan K. Driscoll et al. “Quantifying Modes of 3D Cell Migration”. In: *Trends in Cell Biology* 25.12 (Dec. 2015), pp. 749–759. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2015.09.010.

[194] Mateusz Koziński et al. “Tracing in 2D to reduce the annotation effort for 3D deep delineation of linear structures”. In: *Medical Image Analysis* 60 (Feb. 2020), p. 101590. ISSN: 1361-8415. DOI: 10.1016/j.media.2019.101590.

[195] Ulrik Günther et al. “Bionic Tracking: Using Eye Tracking to Track Biological Cells in Virtual Reality”. In: (2020), pp. 280–297. ISSN: 1611-3349. DOI: 10.1007/978-3-030-66415-2_18.

[196] Anne Beghin et al. “Localization-based super-resolution imaging meets high-content screening”. In: *Nature Methods* 14.12 (Oct. 2017), pp. 1184–1190. ISSN: 1548-7105. DOI: 10.1038/nmeth.4486.

[197] Hamid Fehri et al. “Bayesian Polytrees With Learned Deep Features for Multi-Class Cell Segmentation”. In: *IEEE Transactions on Image Processing* 28.7 (July 2019), pp. 3246–3260. ISSN: 1941-0042. DOI: 10.1109/tip.2019.2895455.

[198] Benoît Audelan et al. “Unsupervised quality control of segmentations based on a smoothness and intensity probabilistic model”. In: *Medical Image Analysis* 68 (Feb. 2021), p. 101895. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101895.

[199] Debarati Sarkar et al. “A minimal model for structure, dynamics, and tension of monolayered cell colonies”. In: *Communications Physics* 4.1 (Feb. 2021). ISSN: 2399-3650. DOI: 10.1038/s42005-020-00515-x.

[200] Daniel T. Gillespie. “Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral”. In: *Physical Review E* 54.2 (Aug. 1996), pp. 2084–2091. ISSN: 1095-3787. DOI: 10.1103/physreve.54.2084.

[201] Leszek Słomiński. “On reflected Stratonovich stochastic differential equations”. In: *Stochastic Processes and their Applications* 125.2 (Feb. 2015), pp. 759–779. ISSN: 0304-4149. DOI: 10.1016/j.spa.2014.10.003.

[202] Jon Louis Bentley. “Multidimensional binary search trees used for associative searching”. In: *Communications of the ACM* 18.9 (Sept. 1975), pp. 509–517. ISSN: 1557-7317. DOI: 10.1145/361002.361007.

[203] Giona Kleinberg et al. “Usability of deep learning pipelines for 3D nuclei identification with Stardist and Cellpose”. In: *Cells & development* 172 (2022), p. 203806.

[204] Auguste Genovesio et al. “Multiple particle tracking in 3-D+ t microscopy: method and application to the tracking of endocytosed quantum dots”. In: *IEEE Transactions on Image Processing* 15.5 (2006), pp. 1062–1070.

[205] Simo Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, Sept. 2013. ISBN: 9781107619289. DOI: 10.1017/cbo9781139344203.

[206] Christian Robert. “Machine Learning, a Probabilistic Perspective”. In: *CHANCE* 27.2 (Apr. 2014), pp. 62–63. ISSN: 1867-2280. DOI: 10.1080/09332480.2014.914768.

[207] Angel F. Garcia-Fernandez et al. “A comparison between PMBM Bayesian track initiation and labelled RFS adaptive birth”. In: (July 2022), pp. 1–8. DOI: 10.23919/fusion49751.2022.9841338.

[208] Martin Maška et al. “The Cell Tracking Challenge: 10 years of objective benchmarking”. In: *Nature Methods* 20.7 (May 2023), pp. 1010–1020. ISSN: 1548-7105. DOI: 10.1038/s41592-023-01879-y.

[209] Julia Bonnet et al. “The Roboscope: Smart and Fast Microscopy for Generic Event-Driven Acquisition”. In: (Sept. 2024). DOI: 10.1101/2024.09.24.614735.

[210] Chenhao Cui et al. “Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration”. In: *Chemometrics and Intelligent Laboratory Systems* 182 (Nov. 2018), pp. 9–20. ISSN: 0169-7439. DOI: 10.1016/j.chemolab.2018.07.008.

[211] Jonathon Luiten et al. “HOTA: A Higher Order Metric for Evaluating Multi-object Tracking”. In: *International Journal of Computer Vision* 129.2 (Oct. 2020), pp. 548–578. ISSN: 1573-1405. DOI: 10.1007/s11263-020-01375-2.

[212] Lalith Sharan et al. “mvHOTA: A multi-view higher order tracking accuracy metric to measure temporal and spatial associations in multi-point tracking”. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 11.4 (Dec. 2022), pp. 1281–1289. ISSN: 2168-1171. DOI: 10.1080/21681163.2022.2159535.

[213] D Fraser et al. “The optimum linear smoother as a combination of two optimum linear filters”. In: *IEEE Transactions on automatic control* 14.4 (1969), pp. 387–390.

[214] Tom Mohr. *A Practical Guide to Periodic Boundary Conditions*. URL: https://tommohr.dev/pbc/ (visited on 02/16/2025).