Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Elena Buglakova

born in: Korsakov, Russia

Oral examination: 24.07.2025

MULTIMODAL IMAGE ANALYSIS IN BIOLOGY

Referees:
Dr. Ana Martin-Villalba
Dr. Sinem Saka

ABSTRACT

Living organisms are striking in their complexity at all levels of organization. No one single analytical method can capture all the information necessary to provide comprehensive knowledge about the cellular processes. The rapid development of new technologies for biological imaging and single cell analysis gives an opportunity to measure different aspects of the living systems at single-cell resolution. Combining these diverse data types can lead to an insight that would not have been possible if each modality were to be considered separately. However, there is no general recipe for a successful multimodal analysis project and it remains challenging to both keep high data quality and establish a reliable way to find correspondence between the two data types. In this thesis I show two different examples of combining different imaging modalities.

Firstly, I demonstrate how, making use of the stereotypical development of the marine worm *Platynereis dumerilii*, I developed fully automated deep-learning based registration pipeline that allowed to map multiple 3D smFISH datasets to the EM image stack of a whole body 6-days post fertilization larva of the animal with near single-cell accuracy. Automated registration enables systematic study of the connection between cell-type specific gene expression and cell phenotype.

In the second project I combined imaging mass spectrometry for spatially-resolved detection of $^{13}C_6$-glucose-derived fatty acids in cellular lipids with microscopy and computational methods for data integration and analysis. I validated this method on a spatially-heterogeneous normoxia-hypoxia model of liver cancer cells. I demonstrated the single-cell heterogeneity of acetyl-CoA pool labelling degree upon ACLY knockdown that would be impossible to detect with bulk analysis.

Segmentation of matching objects in different modalities is a crucial step in multimodal image analysis. Transferable and easy to train segmentation of large biological images with neural networks remains challenging. In the final part of my thesis I show how feature normalization inside the neural network can lead to tiling artifacts or suboptimal performance, and propose a normalization strategy for successfully eliminating the artifacts while keeping high segmentation accuracy.

ZUSAMMENFASSUNG

Lebende Organismen zeichnen sich durch ihre Komplexität auf allen Organisationsebenen aus. Keine einzelne analytische Methode kann alle Informationen erfassen, die für ein umfassendes Verständnis zellulärer Prozesse erforderlich sind. Die rasante Entwicklung neuer Technologien für biologische Bildgebung und Einzelzellanalyse bietet die Möglichkeit, verschiedene Aspekte lebender Systeme mit Einzelzellauflösung zu messen. Die Kombination dieser unterschiedlichen Datentypen kann zu Erkenntnissen führen, die bei isolierter Betrachtung jeder Modalität nicht möglich wären. Allerdings gibt es kein allgemeines Protokoll für erfolgreiche multimodale Analysen, und es bleibt eine Herausforderung, sowohl eine hohe Datenqualität sicherzustellen als auch eine zuverlässige Methode zur Verknüpfung der Datentypen zu etablieren. In dieser Arbeit zeige ich zwei verschiedene Beispiele für die Kombination unterschiedlicher Bildgebungsmodalitäten.

Zunächst zeige ich, wie ich – unter Ausnutzung der stereotypen Entwicklung des Meereswurms Platynereis dumerilii – eine vollständig automatisierte, auf Deep Learning basierende Registrierungs-Pipeline entwickelt habe. Diese ermöglichte es mir, mehrere 3D-smFISH-Datensätze mit nahezu Einzelzellauflösung auf den EM-Bildstapel einer ganzen, 6 Tage nach der Befruchtung aufgenommenen Larve abzubilden. Die automatisierte Registrierung ermöglicht eine systematische Untersuchung des Zusammenhangs zwischen zelltypspezifischer Genexpression und Zellphänotyp.

Im zweiten Projekt kombinierte ich bildgebende Massenspektrometrie – zur räumlich aufgelösten Detektion von aus 13C6-Glukose stammenden Fettsäuren in Zelllipiden – mit Mikroskopie und rechnergestützten Methoden zur Datenintegration und -analyse. Ich validierte diese Methode an einem räumlich heterogenen Normoxie-Hypoxie-Modell von Leberkrebszellen. Ich konnte die Einzelzell-Heterogenität des Acetyl-CoA-Pools nach ACLY-Knockdown nachweisen – ein Effekt, der mit Durchschnittsanalyse nicht erfassbar wäre.

Die Segmentierung korrespondierender Objekte in verschiedenen Modalitäten ist ein entscheidender Schritt in der multimodalen Bildanalyse. Eine übertragbare und einfach zu trainierende Segmentierung großer biologischer Bilder mit neuronalen Netzwerken bleibt eine Herausforderung. Im letzten Teil meiner Arbeit zeige ich, wie Feature-Normalisierung innerhalb des neuronalen Netzwerks zu Blockartefakten oder suboptimaler Leistung führen kann, und präsentiere eine Normalisierungsstrategie, die diese Artefakte erfolgreich beseitigt und gleichzeitig eine hohe Segmentierungsgenauigkeit beibehält.

Insanity is doing the same thing over and over again
and expecting different results.

— Unknown thinker

ACKNOWLEDGEMENTS

Я бы хотела поблагодарить свою маму Ольгу Иосифовну Буглакову за веру в меня и постоянную поддержку.

Также я бы хотела поблагодарить своих улиток Лапку, Сладкую Булочку и Дружка-Пирожка за мотивацию и вдохновение.

Many thanks to my friends Liza Yudina, Måns Ekelöf and Ricardo Sánchez, this work would have been impossible without you!

I would like to thank my supervisor Anna Kreshuk and all former and present members of our group for creating a friendly, creative environment. I learned a lot and felt very welcome.

Thanks to all the people who supported me on my way: my brother Alexey Buglakov, my friends Galina Beliakova, Anya Harutyunyan, Roma Doronin, Danya Polyakov, Magdalena Schindler, Tomaz Fogaca Vieira, Max Kholmatov, Arina, Ksenia Malysheva, Marina Makharova and all the amazing people I met at EMBL.

Most of the work I have done during my PhD was done in a highly collaborative setting. I'm very grateful to all my collaborators for making an effort to understand my explanations, try out my scripts and, in their turn, explain their work to me. I learned a lot by working with such amazing specialists as Luisa Abreu, Edoardo D'Imprima, Marteinn T. Snaebjornsson, Luca Santangeli, Adam Phillip Oel and many many others.

Learning about project management, negotiation and conflict resolution was as important or even more important in my PhD. I am really grateful to EMBL ombudsperson Martina Peskoller-Fuchs and my therapist Elena Knyazeva for mentorship and professionalism.

CONTENTS

I FIELD OVERVIEW AND RESEARCH CONTEXT 1

1 INTRODUCTION 3
- 1.1 Examples of correlative workflows . . . 3
- 1.2 Examples of multimodal integration . . . 5
- 1.3 Methods for matching and registration . . . 6
- 1.4 Thesis overview . . . 8
  - 1.4.1 Construction of *Platynereis dumerilii* multimodal gene expression atlas . . . 8
  - 1.4.2 Spatial single-cell isotope tracing . . . 9
  - 1.4.3 Challenges of feature normalization in bioimage segmentation . . . 9

II CONSTRUCTION OF *platynereis dumerilii* MULTIMODAL GENE EXPRESSION ATLAS 11

2 BACKGROUND 13
- 2.1 *Platynereis dumerilii* as a model organism . . . 13
- 2.2 Integration of EM and single cell RNA sequencing via spatial transcriptomics . . . 14
- 2.3 Creation of atlas based on stereotypic development . . . 15
- 2.4 Registration pipeline outline . . . 15

3 REGISTRATION PIPELINE VALIDATION 17
- 3.1 Description of input data . . . 17
- 3.2 Pre-alignment: global registration based on SVD . . . 17
- 3.3 View merging . . . 19
- 3.4 Segmentation . . . 22
  - 3.4.1 Problem description . . . 22
  - 3.4.2 Ground truth collection . . . 25
  - 3.4.3 Segmentation quality metrics . . . 26
  - 3.4.4 Merging Stardist and Mutex Watershed predictions . . . 27
  - 3.4.5 Results and limitations . . . 29
- 3.5 Spot detection . . . 30
  - 3.5.1 Problem description . . . 30
  - 3.5.2 Spot detection ground truth annotation . . . 31
  - 3.5.3 Spot detection with Spotiflow . . . 32
- 3.6 Gene signal assignment . . . 32
- 3.7 Preliminary similarity alignment . . . 33
- 3.8 Coherent Point Drift . . . 34
- 3.9 Cell matching . . . 35
  - 3.9.1 Assignment problem formulation . . . 35
  - 3.9.2 Scaling up with a sparse distance matrix . . . 36
- 3.10 Deformable registration with matched nuclei centroids as landmarks . . . 37
- 3.11 Export to MoBIE . . . 37

4 GENE EXPRESSION ANALYSIS 39

4.1 Pipeline overview . . . 39
4.1.1 Preprocessing . . . 39
4.1.2 Registration . . . 41
4.1.3 Data analysis and visualization . . . 42
4.2 Assessment of registration quality . . . 45
4.3 Gene expression analysis . . . 46
4.4 Summary . . . 52
4.5 Limitations . . . 53
4.6 Discussion and outlook . . . 55

III SPATIAL SINGLE-CELL ISOTOPE TRACING 57

5 SINGLE-CELL ISOTOPE TRACING OF HETEROGENEOUS CELL CULTURES 59
5.1 Isotope tracing . . . 59
5.2 Spatial single-cell mass spectrometry . . . 59
5.3 Analysis of isotopologue distributions of fatty acids incorporated in lipids . . . 61
5.4 Fatty acid spectra interpretation . . . 62
5.5 ACLY knockdown experiments . . . 63
5.6 Application to tissue samples . . . 63
5.7 Discussion . . . 63

IV CHALLENGES OF FEATURE NORMALIZATION IN SEGMENTATION OF LARGE BIOLOGICAL IMAGES 69

6 CHALLENGES OF FEATURE NORMALIZATION IN BIOIMAGE SEGMENTATION 71
6.1 Introduction . . . 71
6.2 Methods . . . 74
6.2.1 Sliding window inference with halo . . . 74
6.2.2 Evaluation metrics . . . 75
6.2.3 Normalization strategies . . . 76
6.3 Experiments . . . 78
6.3.1 Datasets . . . 78
6.3.2 Architectures . . . 79
6.3.3 Implementation . . . 79
6.4 Results . . . 80
6.4.1 Receptive field . . . 80
6.4.2 Tile-wise feature normalization with `InstanceNorm` causes tiling artifacts . . . 81
6.4.3 Mismatch in batch statistics causes `train` and `eval` mode performance disparity for `BatchNorm` . . . 82
6.4.4 BatchRenorm corrects for train/eval disparity and provides seamless stitching . . . 85
6.5 Discussion . . . 85
6.6 Conclusion . . . 87

V CONCLUSIONS AND OUTLOOK 91

7 CONCLUSIONS AND OUTLOOK 93
7.1 Summary . . . 93

7.2 Cross-modal registration: outlook 94

BIBLIOGRAPHY 97

LIST OF FIGURES

Figure 1 Dense point correspondence with MASt3R 7
Figure 2 Input and final result of the pre-alignment 18
Figure 3 Rotation of the point cloud based on the principle axes . 19
Figure 4 Merging of the confocal views 20
Figure 5 Comparison of dorsal, ventral and composite intensity profiles 21
Figure 6 Ground truth annotation for segmentatation, ROI1 . . . 22
Figure 7 Ground truth annotation for segmentatation, ROI2 . . . 23
Figure 8 Affinity predictions, ROI1 24
Figure 9 Instance segmentation steps, ROI1 25
Figure 10 Final merged segmentation of a full volume 26
Figure 11 Comparison of spot detection with LoG and Spotiflow for the full volume 30
Figure 12 Comparison of spot detection with LoG and Spotiflow for the cells with high gene expression level 31
Figure 13 Gene signal assignment to cell regions 33
Figure 14 CPD alignment 34
Figure 15 Point matching with ILP 36
Figure 16 Registration pipeline: preprocessing 40
Figure 17 Registration pipeline: EM template 41
Figure 18 Registration pipeline: registration to EM template 43
Figure 19 Registration pipeline: visualization 44
Figure 20 Registration quality control with StMhc staining 46
Figure 21 Correlation of the StMhc signal mapped to EM from different replicates 48
Figure 22 Registration quality control with ROpsin staining 49
Figure 23 Correlation of the ROpsin signal mapped to EM from different replicates 50
Figure 24 Expression of Pkd1 and Arx mapped to EM 51
Figure 25 Pipeline tools 54
Figure 26 Protocol for single-cell metabolomics 60
Figure 27 13C-SpaceM workflow as applied to interrogation of de novo fatty acid synthesis 65
Figure 28 Validating 13C-SpaceM by interrogating de novo fatty acid synthesis in spatially-heterogeneous normoxia-hypoxia model 66
Figure 29 Single-cell quantitative analysis of lipogenic acetyl-CoA production and heterogeneity 67
Figure 30 Illustration of the pipeline for processing images larger than GPU memory 72
Figure 31 Examples of predictions with and without tiling artifacts 74
Figure 32 Explanation of tile mismatch metric 75
Figure 33 Log10 of mean gradient of the central pixel output with respect to input 78

Figure 34 Dependency of performance on tile size 81
Figure 35 Example of train/eval disparity 81
Figure 36 Tiling gets worse in transfer 89

LIST OF TABLES

Table 1 Segmentation quality 28
Table 2 List of confocal samples for quality control 45
Table 3 List of confocal samples for gene expression 47
Table 4 Receptive fields and halo for different architectures . . . 80
Table 5 U-Net results table 83
Table 6 UNETR results table 84
Table 7 U-Net transfer results 86

LISTINGS

ACRONYMS

2D Two-dimensional

3D Three-dimensional

SVD Singular Value Decomposition

CNN Convolutional neural network

EM Electron microscopy

FIB Focused ion beam

DL Deep learning

DNA Deoxyribonucleic acid

k-NN k-Nearest neighbors

MLP Multi-layer perceptron

PCA Principal component analysis

RGB Red-Green-Blue

RNA Ribonucleic acid

UMAP Uniform manifold approximation and projection for dimension reduction

ILP Integer Linear Programming

CPD Coherent Point Drift

TRF Theoretical receptive field

ERF Effective receptive field

Part I

FIELD OVERVIEW AND RESEARCH CONTEXT

1 INTRODUCTION

Understanding the mechanisms underlying biological processes requires careful comparison and analysis of evidence coming from different analytical techniques. Traditionally orthogonal methods are applied independently to different samples and then the results of different experiments are brought together to form a comprehensive answer to the research question. In such setup the synthesis of different modalities happens at the level of conclusions. A different, more challenging approach to the analysis of complementary data is to correlate the modalities at the level of raw data, creating a richer representation of the studied objects before drawing final conclusions about them.

Multimodal and correlative workflows consist of two parts: performing the experiment with each of the analytical techniques, which often have contradictory requirements for sample preparation, and establishing correspondence between the entities across the modalities. Due to the necessity to register images or find matching objects the correlative workflows almost always have an indispensable computational component. Recent advances in experimental techniques enabled multimodal and correlative workflows in bioimaging and single cell domains, however, the particular methods of registration and aggregation of different data types vary vastly depending on the specific limitations imposed by the handling of the biological sample, compatibility of different experimental methods and the research question. Let us look at the examples of correlative workflows to see how registration and integration methods are tailored to the demands of each particular project.

1.1 EXAMPLES OF CORRELATIVE WORKFLOWS

One of the widespread examples of the correlative workflows is correlative light and electron microscopy which combines the ability of the fluorescence microscopy to specifically image the molecules of interest with the high resolution of electron microscopy, making it an crucial approach for cell biology. A lot of effort has been made to develop sample preparation protocols that would allow to perform light and electron microscopy on the same sample, for example, [2–4].

Fluorescence microscopy can help to find the rare event in the sample for further high-resolution imaging with EM. For example, Focused Ion Beam Scanning Electron Microscopy (FIB-SEM) produces datasets with near-isotropic resolution down to 2 nm per pixel and the sizes of imaged volumes at the order of magnitude of 100 μ m3. Even though it is a relatively large field of view, it still constitutes only a part of the biological sample, so precise targeting for the biologically interesting regions of interest (ROI) is crucial. Such targeting can be performed using fluorescence microscopy following a number of correlative light and electron microscopy (CLEM) workflows, for example [5]. In this kind of workflows the whole sample is first imaged comparatively quickly at

a lower spatial resolution with some type of fluorescence microscopy and then ROIs for slower and higher-resolution FIB-SEM imaging are chosen based on the fluorescent markers. As a result, the high-resolution FIB-SEM volume image and low-resolution fluorescence volume image of the same biological sample are acquired. After rigid registration the fluorescent signal can be viewed as rough, inexact ground truth for segmentation of the structures of interest in the vEM data.

Using both modalities together requires placing the fluorescence image and electron microscopy in the same coordinates. A simple overlay of the data can be done in many different software, such as ecCLEM [6]. Unfortunately, in many cases rough affine registration might not be enough to fully align two modalities, so in some cases the location of the light signal does not correspond to the location of the object of interest in EM. If minimizing warping at the sample handling level is not possible, deformable registration can be performed to create more accurate overlay.

The appearance of light and electron microscopy images is very different, therefore registration requires detection of features which match between modalities. There are several options for finding the matching points.

Firstly, registration can be done based on manual landmarks. For example, in [7] manual landmarks and calculating thin plate spline transformation in BigWarp [8] allowed to accurately combine cryogenic super-resolution fluorescence microscopy and focused ion beam milling scanning electron microscopy to visualize protein-ultrastructure relationships in three dimensions (3D) across whole cells. However, manual landmark detection can be slow and subjective.

To avoid the time-consuming process of manual registration, special fiducial nanoparticles which are both fluorescent and visible in electron microscopy can be added to the sample to enable registration, such as in [9]. Although artificial fiducials provide a fast way to detect poits for alignment, there are two causes for concern. Optimal registration quality is reached when the fiducials are uniformly distributed throughout the sample, as the deformation field can only be accurately estimated in the area close to the control points. The penetration of the nanoparticles in the sample can be uneven, resulting in areas where no features are available for registration. Another consideration against this method, which is much harder to quantify, is that adding nanoparticles can interfere with the process or cellular structure of interest.

Finally, natural features of the sample can be used to find correspondence between light and electron microscopy. For instance, [10] uses lipid droplets for alignment as they are easily identifiable in both modalities. For complicated approach involves automated segmentation as a source of landmarks. In [11] mitochondria segmentation was used to perform automated deformable registration in Elastix [12]. CLEMreg [13], a tool for automated CLEM, also uses automated segmentation of mitochondria but instead of aligning light and EM segmentations it utilizes point cloud based registration. This is a promising method as it does not require adding any modifications to the sample but the segmentation can be a bottleneck as acquiring ground truth and training a segmentation model can be more tedious and time-consuming than providing manual registration landmarks. In addition, finding the feature which is

visible in EM, can be labeled in fluorescent microscopy and is spread evenly throughout the sample to ensure uniform registration quality can be challenging.

1.2 EXAMPLES OF MULTIMODAL INTEGRATION

Another field in which using different modalities together has been investigated in depth is single-cell data analysis. Various single-cell data modalities can be roughly divided into spatial and non-spatial.

In non-spatial modalities called "omics" the sample is first dissociated into separate cells after which the measurements are conducted for each cell independently, providing a vector of values. The most widespread single-cell data type is scRNAseq which measures gene expression. A large variety of single-cell protocols has been developed to address different research questions and handle different types of samples, enabling researches to measure chromatin accessibility, DNA methylation, proteomics profiles, immune receptor repertoire and other characteristics of the cells [14].

In spatial modalities ("spatial omics") [15] the measurements are made at single-cell or near single-cell resolution while keeping the spatial context of the sample. This can be achieved either by probing a thin slice of the sample in a grid fashion, such as with Visium platform for spatial transcriptomics, or by microscopy, for example, with MERFISH [16], a transcriptomics method based on consecutive staining and washing cycles and fluorescence in situ hybridization.

All of these data types share a common feature: the measurements at single-cell level are very sparse and noisy. Each separate cell contains very small amount of material, making it difficult to detect signal even with methods that involve amplification. Sparse matrices with a lot of zeros are characteristic of single-cell data in all modalities. Another reason for high noise level is stochastic differences between cells. Even in a very homogeneous cell culture all cells do not have exactly the same level of gene expression [17]. Usually the goal of single-cell analysis is to discover groups of cells which would otherwise be lost in the bulk experiment, therefore single-cell expression vectors are used for automated clustering. Averaging expression vectors over a group of cells reduces the noise and allows to compare cell types.

Multi-omics data integration [18] refers to joint analysis of single-cell data acquired using different modalities applied to different samples with the goal of mapping all datasets and modalities into a shared space to enrich the single-cell representations and interpret data in a larger context.

One example of multi-omics integration task is the integration of spatial transcriptomics with non-spatial scRNAseq [19]. Spatial transcriptomics provides information about the proximity of cells and tissue organization but measures only a subset of genes while the scRNAseq atlas of the same tissue has more robust verified cell type annotations. In this task both modalities share the variable space, meaning that the correspondence between features is known, making it is possible to calculate the probability that each spatial measurement comes from a certain scRNAseq cell type. There exists a vari-

ety of computational methods solving this task by using correlation analysis, Bayesian statistics or deep learning.

An important difference of multimodal integration from correlative imaging workflows is that in most cases complementary data does not come from the exact same sample, so the goal of the analysis is to find similarity in the structure of the datasets in a shared feature space or discover correspondence between features that would allow to match cell types.

1.3 METHODS FOR MATCHING AND REGISTRATION

Multimodal and correlative workflows rely on computational methods for the alignment of data from different modalities.

Image registration is a diverse group of algorithms which aim to overlay two input images in a way that maximizes their similarity. Usually one image remains fixed. It is used as a template for registration of the moving image. The algorithm finds a transformation that would match the images.

Transformation can be parametrized in different ways depending on the expected difference between images. Affine transformation includes translation, rotation, scaling ans sheer. In many cases rigid transform (translation + rotation, 6 parameters in 3D space) or similarity transform (translation + rotation + scaling, 9 parameters in 3D space) is enough to roughly align the images. These transformations have small number of parameters so in principle finding just a few matching landmarks is enough to estimate the transformation parameters.

If parts of the sample moved with respect to each other between the imaging rounds, affine or rigid transforms become insufficient for accurate alignment. For example, if the sample was imaged with light microscopy and then underwent sample preparation for electron microscopy, it becomes warped from interaction with buffer and physical handling, so the electron microscopy does not perfectly match the light microscopy. In such cases deformable registration is applied to compensate for relative movement. Deformable registration finds displacement of different points in the sample. A widely used and general way to parametrize a deformable transformation is splines [20]. With splines the displacements are estimated for a set of so-called control points in the image and the displacement between control points is approximated by the spline function. With B-Splines control points are arranged in a regular grid while with thin-plate splines control points can have arbitrary locations. Varying the number of control points and applying different constraints on the splines allows to regulate the flexibility and number of parameters of the transform, making splines an adaptable parametrization for deformable alignment.

The obvious way to register to images is to slide them with respect to each other, checking the match quality at each displacement to find the best relative orientation. Such procedure would provide global registration, however, it is very slow and not feasible for most practical image sizes. Fourier-based registration [21] makes use of the convolution theorem and Fast Fourier Transform to speed up normalized correlation coefficient. Modifications of this algorithm allow to take the rotation angle into account [22].

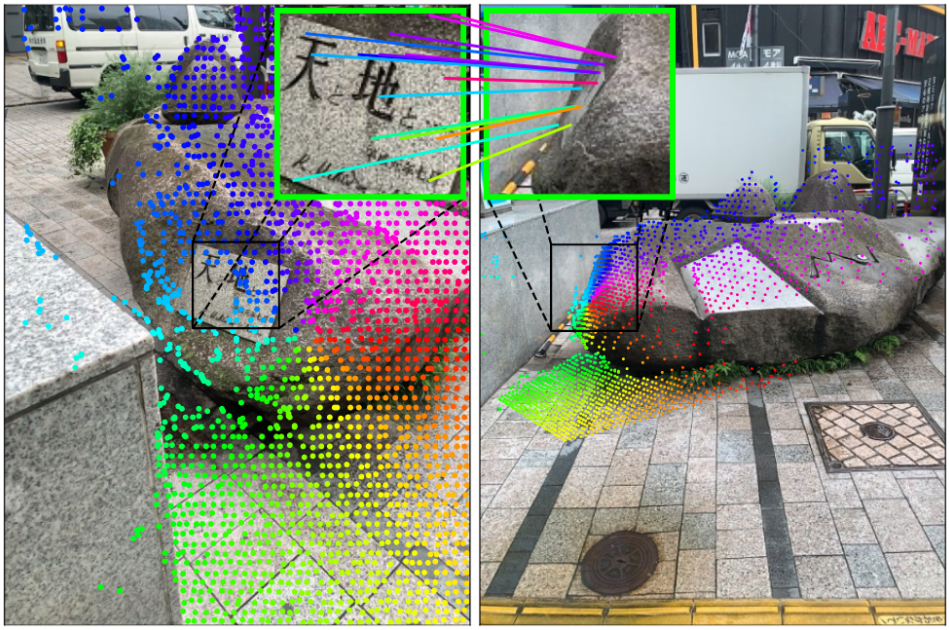


Figure 1: MASt3R predicts dense correspondences, even in regions where camera motion significantly degrades the visual similarity. Reproduced from [24].

Even with FFT, direct calculation of the exact correlation is not feasible for large images, especially for 3D data. Popular framework for intensity-based registration Elastix [12] solves this problem by estimating the correlation from a fixed number of sampled positions instead of calculating full transformation of the moving image for every possible transformation. In Elastix the transformation parameters are initialized as a center of mass alignment or identity transform, after which gradient descent is used to find the transformation which would maximize correlation between images. Unlike registration with cross-correlation, registration in Elastix is local, meaning that the final result depends on the initialization.

Regardless of how the transformation parameters are found, registration methods based on direct comparison of the pixel values share common disadvantages. Individual pixel values are often noisy which means that correlation of the correctly registered images might not be very different from the background. Cross-correlation is known to have many peaks if the images have repeating objects and choosing the correct peak can be non-trivial. Multiresolution registration and image smoothing are used to address this problem but do not fully solve it [23]. Another problem is that intensity-based registration has limited applicability in multimodal setting. Other metrics, such as Mutual Information, were proposed in cases when modalities are relatively close, such as in medical imaging [23].

Finding informative landmarks and matching them between images is an alternative to intensity-based registration that has multiple advantages. Firstly, even with a large number of landmarks, the matching is much faster than resampling full images and calculating correlation coefficient. Using landmarks allows to make use of prior information about the objects and is less prone to getting stuck in local minima. The registration pipeline consists of 3 parts: identifying the landmarks, calculating point descriptors (vectors, which describe

the properties of the landmarks and allow to determine if two landmarks look similar) and finding a matching between points. In the end the transformation parameters are estimated using the landmark pairs. Different algorithms can be used for each of the three steps, giving rise to a large variety of registration pipelines.

A classic example of landmark-based registration is Scale-invariant Feature Transform (SIFT) [25]. In this algorithm landmarks and their descriptors are found using a special kind of classic image filters. After that the correspondence between points is found using Random Sample Consensus (RANSAC).

Despite many attempts Deep Learning has not been widely used in practice to directly estimate the transform parameters or deformation fields [26]. This can be attributed to the local nature of CNN features as well as the gradient descent being a local optimization method while registration is inherently a global optimization problem. However, neural networks are well-suited for finding the informative landmarks and calculating meaningful descriptors [24, 27–30] which, followed by combinatorial optimization methods for finding point matching, yield impressive results in challenging registration tasks, such as finding correspondence between photos of the same scene taken from different camera angles (Figure 1).

This incomplete review shows that registration is a very diverse group of concepts, algorithms, heuristics and practical tricks which draws inspiration from all areas of mathematics and computer science. There is no single method that can solve any image alignment problem and methods are often developed for each particular problem based on general principles of computer vision and optimization. The approach I consider the most promising, used throughout this thesis, is to exploit excellent local prediction properties of DL and combine them with global optimization for registration itself. While it is hard to formulate the registration/point matching problem in a fully differentiable way, detection of local landmarks - as an example of object detection - is a classic problem of computer vision and can be successfully solved with DL.

1.4 THESIS OVERVIEW

In my PhD I worked on two projects, both of which involve multimodal image analysis but deal with completely different modalities and make contribution to different parts of the workflow.

1.4.1 *Construction of Platynereis dumerilii multimodal gene expression atlas*

The first part of the thesis describes the automated registration pipeline for mapping gene expression data to the EM volume of *Platinereis dumerilii* to enable systematic study of the connection between cell-type specific gene expression based on scRNAseq and cell phenotype. In this project I make use of the stereotypic development of the organism to accurately correlate images acquired for different individual animals.

1.4.2 *Spatial single-cell isotope tracing*

In my second project I extended an existing spatial single-cell metabolomics method to study the utilization of different carbon sources at single-cell resolution. Developed technique relies on the all-ion fragmentation of the cell's lipids to overcome low sensitivity and specificity of the mass spectrometry imaging and provide robust, easily interpretable readout of the cell's metabolism.

1.4.3 *Challenges of feature normalization in bioimage segmentation*

As I show in the previous two chapters, automated segmentation is a crucial step of the single-cell multimodal image analysis pipelines. Making deep learning-based segmentation more reliable and easier to use would make automated cross-modal registration more easily available. In the last project I investigated the causes of the tiling artifacts which appear in the predictions of many publicly available segmentation methods. I discovered that the most prominent reason for the artifacts is the feature normalization inside the neural network and proposed an alternative normalization strategy which takes into account the differences between the biomedical image analysis domain and natural images.

Part II

CONSTRUCTION OF *PLATYNEREIS DUMERILII* MULTIMODAL GENE EXPRESSION ATLAS

Contributions

- I developed registration pipeline and manually annotated ground truth for spot detection
- Luca Santangeli tested the protocol for confocal data acquisition and acquired part of the samples
- Samuel Haury Parra acquired part of the confocal volumes
- Zülfiye Gülce Serka manually annotated ground truth for training nuclei segmentation model
- Albert Dominguez Mantes trained a model for spot detection
- Qin Yu trained Stardist model for nuclei segmentation
- Adam Phillip Oel worked on scRNAseq atlas of 6 dpf *P. dumerilii*

2 BACKGROUND

2.1 *platynereis dumerilii* AS A MODEL ORGANISM

Platynereis dumerilii is a sea annelid which can be kept in the lab conditions. It serves as a convenient model organism for evolutionary and developmental biology [31]. Larval development of *P. dumerilii* is synchronized between individual animals until the late nectochaete stage (5-7 days post fertilization) [32], making this time point especially interesting to study. At this stage the larva already has about 11000 cells as well as many differentiated tissues and cell types while remaining relatively small (300 µm by 100 µm). 6 days post fertilization (6 dpf) embryos are amenable to various imaging techniques and developmental synchronization enables researches to image different individual animals with different techniques and correlate the results.

Significant work has been done to characterize spatial gene expression in different developmental stages of the *P. dumerilii*. In [33] the image analysis pipeline called Profiling by Image Registration was used to create an average template of the axonal scaffold of the Platynereis larval brain at 48 hpf using 36 different individual animals. Then samples with gene staining were registered to the template. Multiple replicates were acquired for each gene, then the gene signal was normalized and averaged across replicates to decrease the influence of registration errors and individual variability on the final gene expression probability map. Similarly, in [34] a reference template of nuclei staining was created using 40 individual image stacks for 48 and 72 hpf larvae.

[35] introduced Profiling by Signal Probability Mapping (ProSPr) to create an expression atlas for 6 dpf developmental stage. It was observed that "at this more variable later stage with complex body features, the variability in position of a single, uniquely identifiable cell is larger than the average cell radius, in conflict with cellular resolution" making the registration significantly more challenging than in [33, 34]. The average template of nuclei staining was generated using 153 individual animals and at least 12 samples per gene were necessary to create the gene expression probability maps with the reliable overlap between replicates. At least 20 replicates were imaged per gene, then DAPI staining of each sample was aligned to the template DAPI using automated registration procedure similar to [33, 34]. After that the gene signal was thresholded semi-manually and the outliers whose expression did not sufficiently match the average over 20 samples were manually removed, retaining at least 12 samples. This shows that intensity-based registration failed in about 40% of cases. Necessity to introduce manual input compared to [33, 34] once again highlights that at 6 dpf registration becomes very challenging. Another disadvantage of this approach is that manual thresholding introduces bias into gene expression quantification, meaning that areas with less active expression or areas which have worse registration quality might be omitted in the final binary expression mask.

In [36] full 6 pdf larva of *P. dumerilii* was imaged using SBEM at $10nm \times 10nm \times 25nm$ resolution. Nuclei and cells were automatically segmented and then the ProSPr nuclei template was registered to the EM volume using nuclei segmentation. Registration was done in Elastix, using a progression of registration steps highly tailored for this particular averaged nuclei probability map. This work resulted in a multimodal atlas complementing the high resolution cell morphology uncovered by electron microscopy with the gene expression probability maps.

Further analysis of this EM volume in [37] led to the creation of MorphoFeatures: unsupervised method for the characterization of cell morphology. Multimodal atlas proved to be a valuable resource for the interpretation and validation of unsupervised analysis results, allowing to identify genetic markers of morphological clusters.

2.2 INTEGRATION OF EM AND SINGLE CELL RNA SEQUENCING VIA SPATIAL TRANSCRIPTOMICS

Single-cell RNA sequencing (scRNAseq) is a technique that allows to interrogate the gene expression of individual cells, providing insight into heterogeneity of cells and allowing to discover new cell types. scRNAseq was applied to annelids [38–40] and other trochophore larva [41] to study evolutionary and developmental history of various cell types. scRNAseq measures the expression profiles of individual cells, however, this technique is not spatial. Spatial expression of anchor genes acquired using Single Molecule Fluorescence in Situ Hybridization (smFISH) [42] can serve as a bridge between scRNAseq and electron microscopy, linking rich ultrustructural information with full expression profiles.

If spatial distribution of anchor genes' expression could be accurately mapped to cells in EM, it would be possible to construct expression profiles of cells in EM and then use these truncated profiles based only on marker genes to determine to which cluster in scRNAseq cluster this EM cell belongs, allowing to complement marker gene expression with the full transcriptional profile of the cluster. Such connection would allow to integrate gene expression information with MorphoFeatures, creating a unique multimodal atlas for 6dpf larva of *Platynereis dumerilii*.

ProSPr gene expression atlas contains many genes and was already registered with EM volume, however, it has some drawbacks. Firstly, manual thresholding makes the expression probability maps incomplete and somewhat subjective. Keeping signal in the areas with lower gene expression and areas which are more challenging for registration would facilitate more accurate automated scRNAseq cluster assignment. Secondly, newer staining protocols and advances in confocal imaging allow to acquire spatial transcriptomics data of better quality, making it possible to match individual nuclei avoiding the averaging. Lastly, with the manual input and 20 replicates necessary to achieve near-single-cell resolution, adding new genes to the atlas is very tedious.

All these factors led to the development of new registration pipeline, aiming to fully automatically map smFISH datasets to EM with single-cell resolution.

2.3 CREATION OF ATLAS BASED ON STEREOTYPIC DEVELOPMENT

The most widely used stereotypic model organism is *C elegans*. The *C. elegans* nervous system consists of 302 neurons located in 11 ganglia throughout the body of the worm. Whole-brain imaging and molecular profiling are used to study nervous system development and brain function. Exact identification of each neuron's identity is a major challenge for correct interpretation of all the various data types. Despite being highly stereotypical, the neurons have some variability in location [43] making cell identification based solely on the spatial location hard or, in some cases, impossible. Experimentally, many methods were designed to aid matching neurons to the atlas, for example, [44]. Computationally, cell matching pipeline starts with instance segmentation of the nuclei, after which the nuclei from different samples are matched either to the common atlas or to each other, such as in [45, 46]. Matching can be based only on the relative distances or use other features such as nuclei shapes and fluorescence signal.

2.4 REGISTRATION PIPELINE OUTLINE

Aim:

Automatically map gene expression data to the EM volume to enable systematic study of the connection between cell-type specific gene expression based on scRNAseq and cell phenotype

Requirements:

- Up to 5 replicates per gene
- Single-cell registration accuracy
- No manual input

To address these challenging requirements, instead of the averaging used in [33–35] I chose the strategy similar to the neuron identification in *C elegans*. Firstly, I perform automated instance segmentation of nuclei in confocal microscopy. Then segmented nuclei from each of the confocal volumes are matched to the instance nuclei segmentation of EM volume which plays a role of the registration template.

In Chapter 3 all parts of the registration pipeline are discussed in detail. Registration quality assessment and showcase of the mapped gene signal are presented in Chapter 4

3 REGISTRATION PIPELINE VALIDATION

3.1 DESCRIPTION OF INPUT DATA

- `Electron microscopy volume`

  Resolution: $10 \times 10 \times 25$ nm. The raw EM data is only used for final visualization and does not take part in registration.

- `Nuclei segmentation in electron microscopy`

  Nuclei segmentation of EM data was downsampled to $0.32 \times 0.32 \times 0.32$ µm and used as a template for the registration of confocal volumes. Final registered volumes were

- `Confocal volumes: dorsal and ventral views`

  Resolution: $0.5 \times 0.13 \times 0.13$ µm.

  Channels: DAPI (nuclei staining) and 1 or 2 channels with gene statining.

  Only DAPI channel was used for registration. Final transformation was applied to the gene channels.

  Due to significant anisotropy and large size of the images after spot detection and segmentation further registration to EM and visualization is done at $0.32 \times 0.32 \times 0.32$ µm.

- `Metadata table`

  Metadata includes gene names, resolution and channel order.

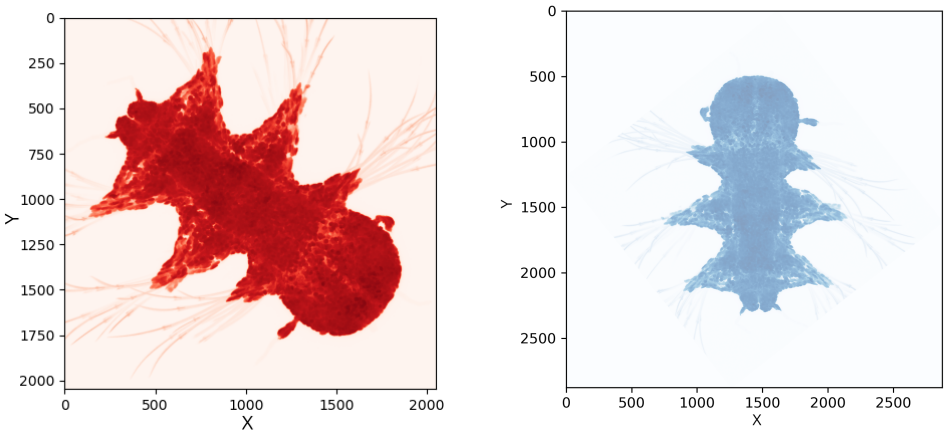
3.2 PRE-ALIGNMENT: GLOBAL REGISTRATION BASED ON SVD

Input samples can be oriented randomly, for example, the sample in the ventral view can be rotated by 180° degrees in XY plane which poses significant difficulty for intensity-based registration algorithms. The first step of the registration pipeline aims to perform rough global registration to make sure that all samples have enough overlap to ensure consistent convergence of intensity-based registration using Elastix [12]. The elongated body of the worm has clearly distinguishable anteroposterior, dorsoventral and left-right axes, making it possible to use Singular Value Decomposition to find the body axes and orient them along the orthogonal basis of the common coordinate space.

SVD is often used for image alignment, for example, alignment of principal axes of two point clouds underlies the classic Iterative Closest Point algorithm [47]. In this case a modified version of the SVD-based alignment was used to avoid resampling the images.

1. `Foreground segmentation`

   DAPI channel was normalized using quantile normalization between 0 and 0.99 quantiles. Then the volumes were downsampled by the factor



(a) Maximum projection of the input sample

(b) Maximum projection of the pre-aligned sample

Figure 2: Input and final result of the pre-alignment.

of 10 and smoothed with a Gaussian filter with $\sigma = 3$ and thresholded to obtain an approximate foreground segmentation. $n = 10000$ points were sampled from the foreground mask resulting in the point cloud representing the sample. Coordinates of pixels belonging to the segmented region in the image:

$$p_i = (x_i, y_i, z_i) : 1 = 1, ..., n \tag{1}$$

2. `SVD`

   First the point cloud was centered:

$$P_i = (x_i - \sum_{i=1}^{n} x_i, y_i - \sum_{i=1}^{n} y_i, z_i - \sum_{i=1}^{n} z_i) : 1 = 1, ..., n \tag{2}$$

   Then SVD was applied to the resulting point cloud:

$$SVD(P) = U\Sigma V^T \tag{3}$$

3. `Translation and rotation of the image in XY plane`

   Rotation angle α in XY plane around the center of mass was determined using matrix V^T:

$$\alpha = 90 - arctan2(V^T[0,1], V^T[0,0]) \tag{4}$$

   Here $arctan2$ is the signed angle between the ray from the origin and passing through the point (1,0), and the ray from the origin through the point (x2, x1).

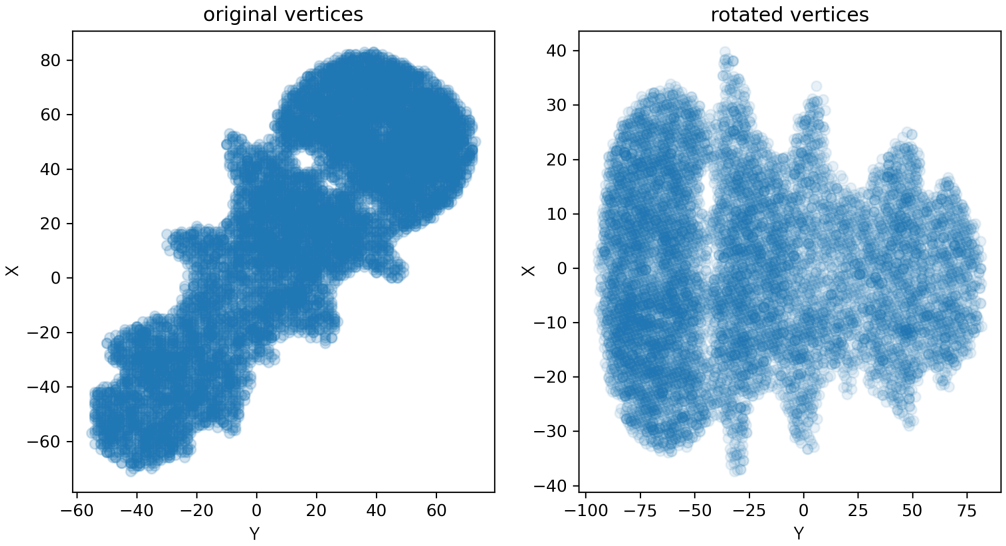


Figure 3: Rotation of the point cloud based on the principle axes. Left: maximum projection of the point cloud corresponding to the input image. Right: the same point cloud, rotated with the matrix V.

4. `Orienting the head of the animal with maximum intensity projection`

   To ensure that the head of the animal is always facing towards the negative direction of the Y axis, the maximum projection on Y axis was taken. The head of the animal has higher cell density than the rest of the body, so if the maximum of intensity was further from the origin than the center of the image, the image was flipped along Y.

This heuristic procedure ensured that all samples are pre-aligned enough both for the registration of the confocal views and for further registration with the EM template while having the orientation that is convenient for quality control within the pipeline: dorsoventral axis along Z, anteroposterior axis along Y and left-right axis along X.

3.3 VIEW MERGING

In fluorescence microscopy, the illumination as well as the detected light is subject to scattering in biological tissues, limiting the penetration depth of the imaging. Decreasing signal intensity makes automated segmentation and spot detection much more challenging and leads to the loss of gene expression information. In addition, in Platynereis embryo samples the scattering depends on the orientation of the embryo with respect to imaging plane.

Various techniques such as multiphoton imaging [48] and tissue clearing [49] have been developed to circumvent this limitation. In this project taking into account the available equipment and protocol multiview confocal imaging was used to compensate for the loss of signal. In multiview imaging the sample is imaged from different angles to increase the signal to noise ratio and achieve isotropic resolution [50].

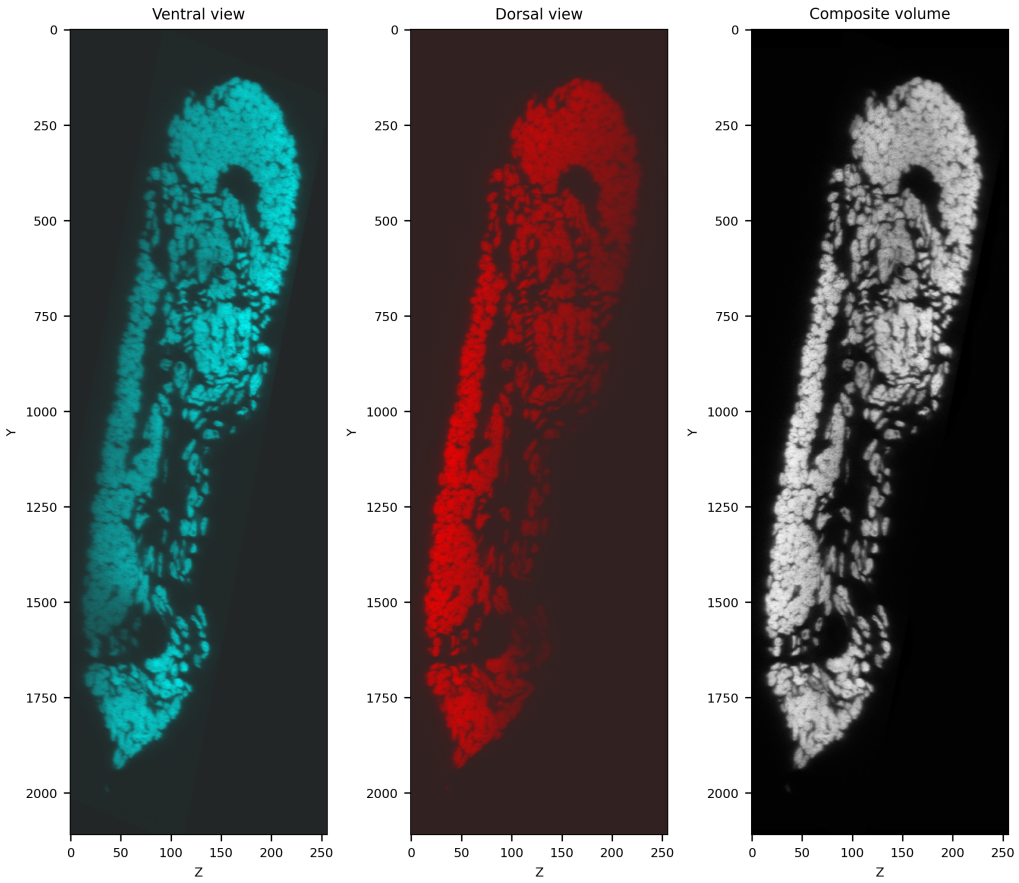


Figure 4: Merging of the confocal views. Ventral view (in cyan) and dorsal view (in red) are shown after rigid and deformable registration. Composite volume shows the final result of merging.

The larvae was mounted between two coverslips which allowed to scan it from one side, then physically rotate the slide and image the larvae again, producing dorsal and ventral views (Fig. 4). After that the views were registered in two steps and merged:

- `Rigid alignment`

  Pre-aligned images were registered using Euler transform in Elastix [12]. Firstly, the registration was done using mutual information metric with 5 levels in the image pyramid schedule (64, 32, 8, 4 and 1) and 20000 spatial samples. In the second step rigid registration was done using normalized correlation coefficient using 5 levels in the image pyramid schedule (16, 8, 4, 2 and 1) and 50000 spatial samples for metric and gradient estimation.

- `Deformable alignment`

  Due to small deformations resulting from the sample handling between the acquisition of the views it was not possible to fully register two views using only rigid registration. Elastix B-Spline deformable registration was used as the last registration step. The metric consisted of normalized correlation, mutual information and rigidity penalty with the corresponding weights of 1, 1 and 100. Registration was performed in one step at the original resolution with the final grid spacing of 64 voxels

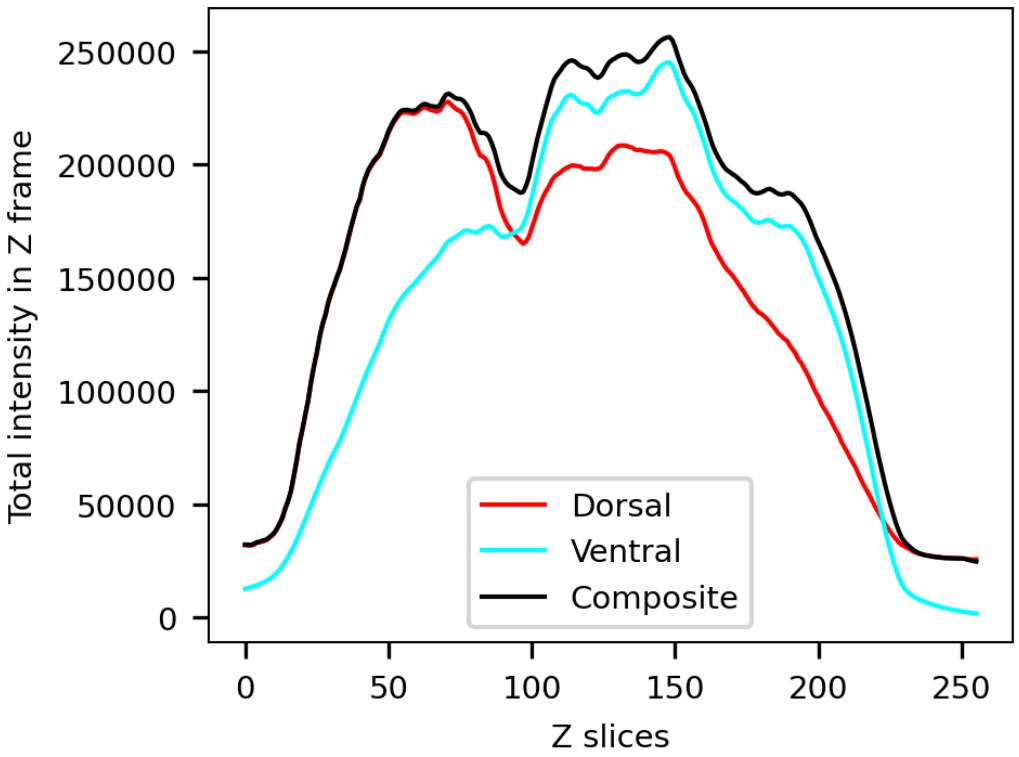


Figure 5: Comparison of intensity profiles of DAPI signal in ventral (cyan) and dorsal (red) views and in the composite volume (black). Each line shows the sum of the signal over the whole Z frame. In some frames the composite volume has higher total signal than either view, demonstrating that the merging utilizes the areas of higher signal rather than performing merging based purely on the coordinate.

in X and Y directions and 16 voxels in Z direction using 100000 spatial samples for metric and gradient estimation.

- `Creating composite volume`

  Small inaccuracy in registration of two views can cause duplication of the spots and make nuclei segmentation more difficult, therefore instead of directly averaging the registered views, the composite volume was constructed. DAPI channel of both registered images was smoothed with gaussian with kernel size of (10, 20, 20) (ZYX). Then the volume was split into the areas where one of the smoothed views had higher intensity than the other and this mask was used to create a weighted average of the views (Fig. 4). This merging procedure allowed to use images even in case they did not match perfectly while preserving as much of the original signal as possible without introducing merging artifacts, as seen in the intensity profiles of the samples in Fig. 5.

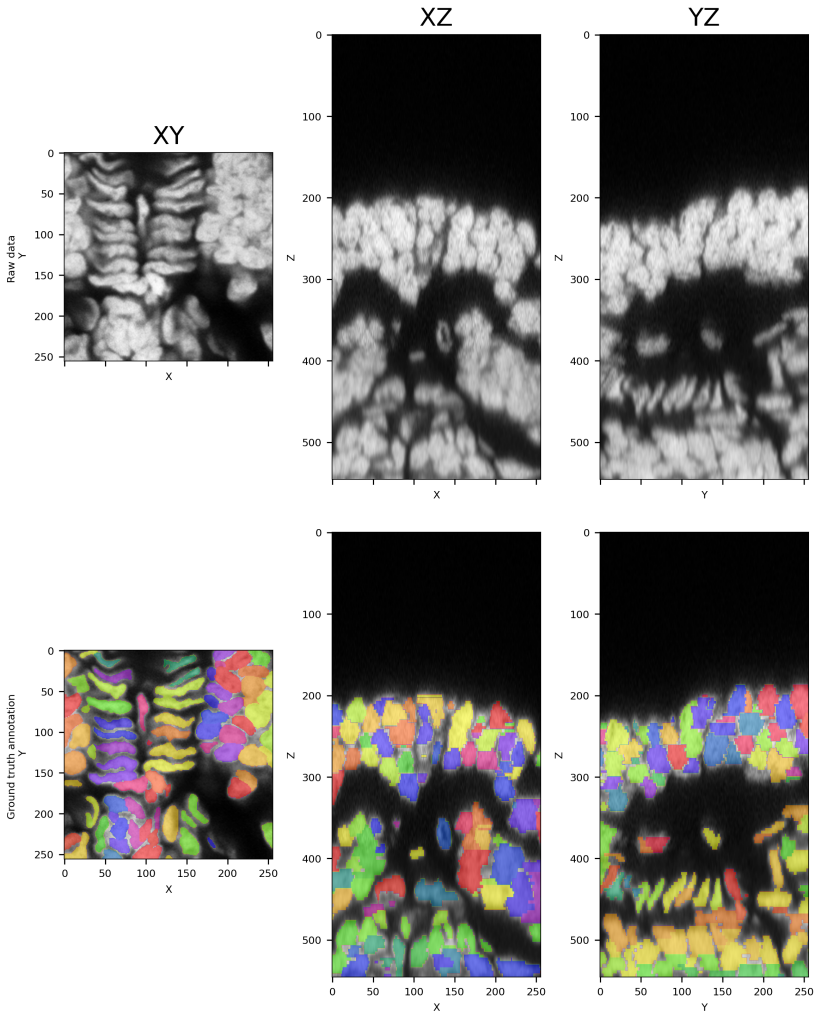


Figure 6: Ground truth annotation for segmentatation, ROI1.

3.4 SEGMENTATION

3.4.1 *Problem description*

Finding matching landmarks between images is a crucial step of cross-modal registration, especially if the modalities have as different appearance as confocal and electron microscopy. Nuclei segmentation is an obvious choice for recognizing matching objects these modalities.

Nuclei are clearly visible in electron microscopy and for the EM volume of *Platynereis* instance segmentation was performed earlier [36] using a combination of deep learning semantic segmentation and graph partitioning method called multicut [51]. In confocal microscopy DAPI staining is a reliable and widely used technique for imaging nuclei. However, registration of raw DAPI signal with semantic segmentation from EM is challenging due to variations in intensity and the fact that individual animals do not necessarily match cell to cell, making pixel-wise correspondence of intensities not possible.

To achieve accurate cell-to-cell matching it is necessary to perform accurate instance nuclei segmentation in confocal volumes. Nuclei segmentation in fluorescence microscopy is a well-studied problem with a multitude of methods

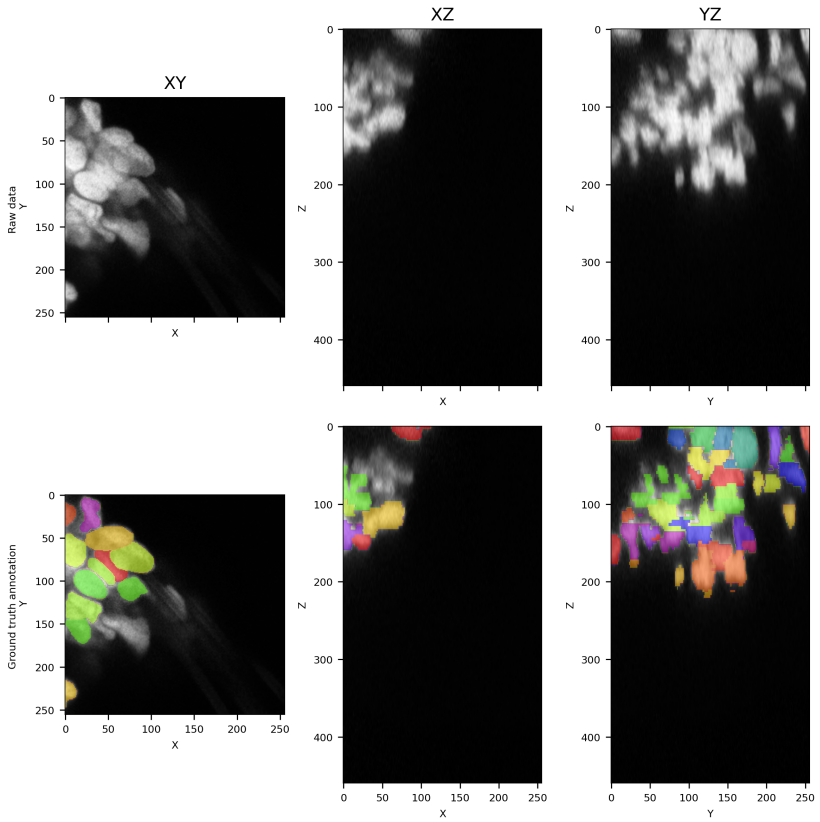


Figure 7: Ground truth annotation for segmentatation, ROI2.

already developed, for example, Cellpose [52] and Stardist [53] are popular tools distributed together with pretrained models. However, segmentation of this data posed multiple challenges:

- `3D segmentation`

  Due to lack of publicly available annotated 3D datasets, a lot of methods focus on segmentation of 2D images. 3D segmentation is achieved through heuristic-based methods. For example, Cellpose 3D approximates 3D flows using predictions of a 2D network. Another option is to perform segmentation in each Z slice individually and then use hungarian matching or tracking to connect instances between slices [54]. However, in this data the main difficulty is correctly splitting objects in Z which is impossible without taking into account 3D context. For this reason only tools which employ 3D networks and perform postprocessing in 3D were considered.

- `Variety of nuclei shapes within one dataset`

  Instance segmentation methods have a postprocessing step which converts network predictions to separate objects. This postprocessing assumes that the objects should have a particular size or shape. For example, Stardist uses star-convex objects, which means that this method can only recognize blob-like objects. Multicut is more flexible and can be adjusted to segment different types of objects, however, only one type can be found at a time. *Platynereis* at this stage has tissues which consist of

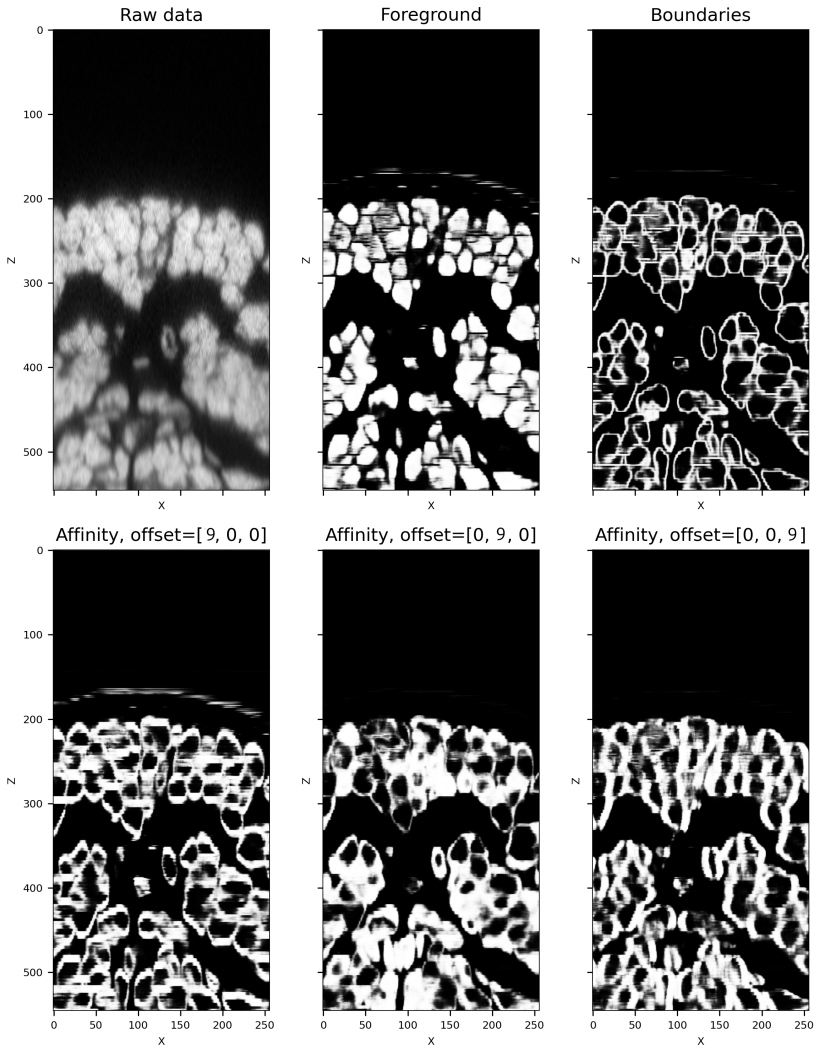


Figure 8: Affinity predictions, ROI1.

cells with very different nucleus shapes: the brain of the animal is made up of tightly packed cells with round nuclei and barely any cytoplasm while muscle cells can have flat, narrow and curved nuclei.

- `Resolution in Z`

  Confocal microscopy has lower resolution in the direction perpendicular to the imaging plane, making it harder to tell where one nucleus ends and another starts even if it is relatively easy to perform segmentation in XY plane. Even manual segmentation is hard and somewhat arbitrary (see the shape of the nuclei in XZ and YZ slices in Figure 6 and Figure 7), especially in the brain of the animal where the density of the objects is high.

- `Ground truth annotation in 3D`

  Training a 3D model and assessing instance segmentation quality in 3D requires 3D ground truth annotations. It is very time-consuming and tedious to do these annotations.

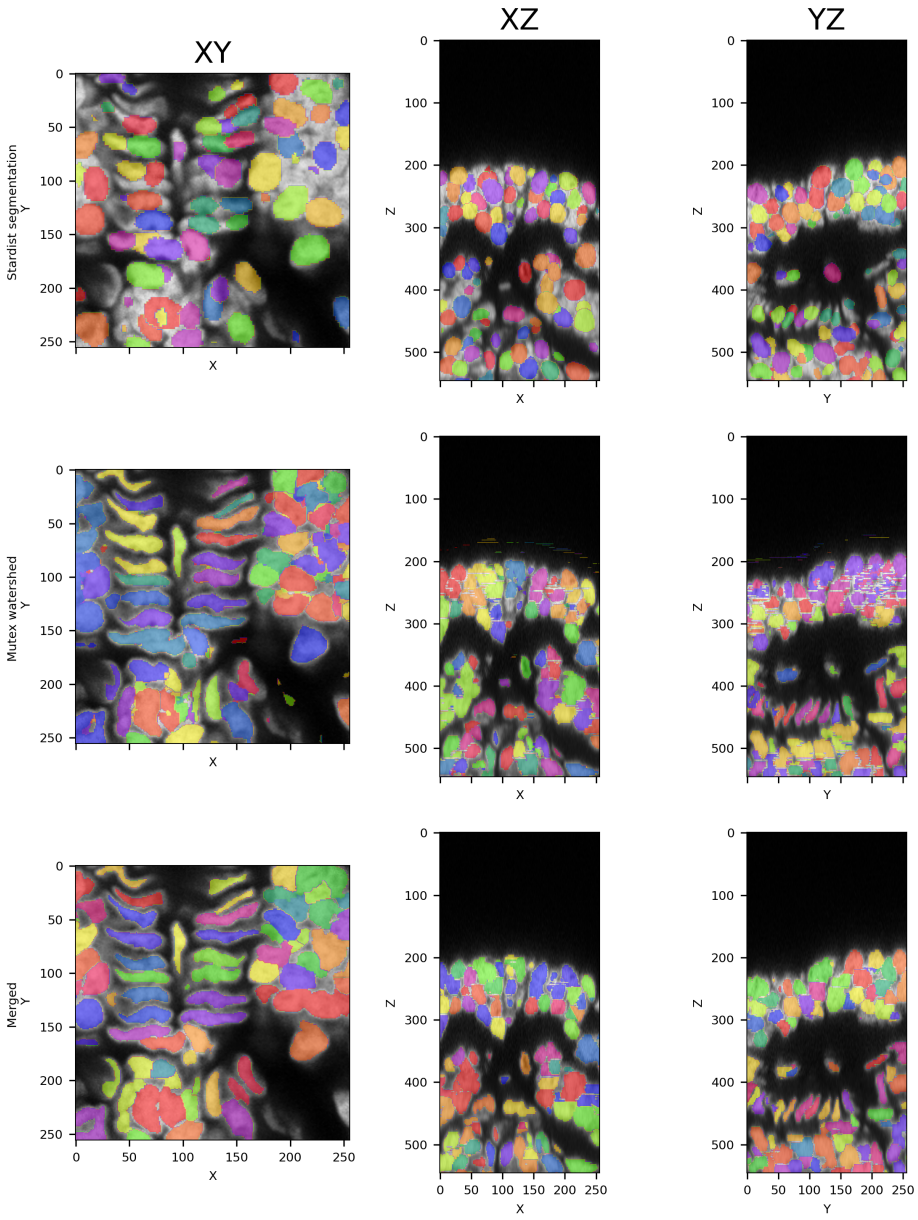


Figure 9: Instance segmentation steps, ROI1.

Unfortunately, no readily available tools were able to deal with this segmentation, therefore it was necessary to collect ground truth and combine existing methods to reach required segmentation accuracy.

3.4.2 *Ground truth collection*

Two blocks of data were chosen in two different datasets, covering the most challenging parts of the volume. ROI 1 (Figure 6) has both examples of flat nuclei with challenging shape and the areas with round, crowded nuclei. ROI 2 (Figure 7) comes from one of the parapodia of the animal, where the nuclei signal is obstructed by the auto-fluorescence of chaetae.

Manual segmentation was performed in Napari by Zülfiye Gülce Serka. Multiple solutions for annotation were tested, such as Cellpose GUI [55], Segment

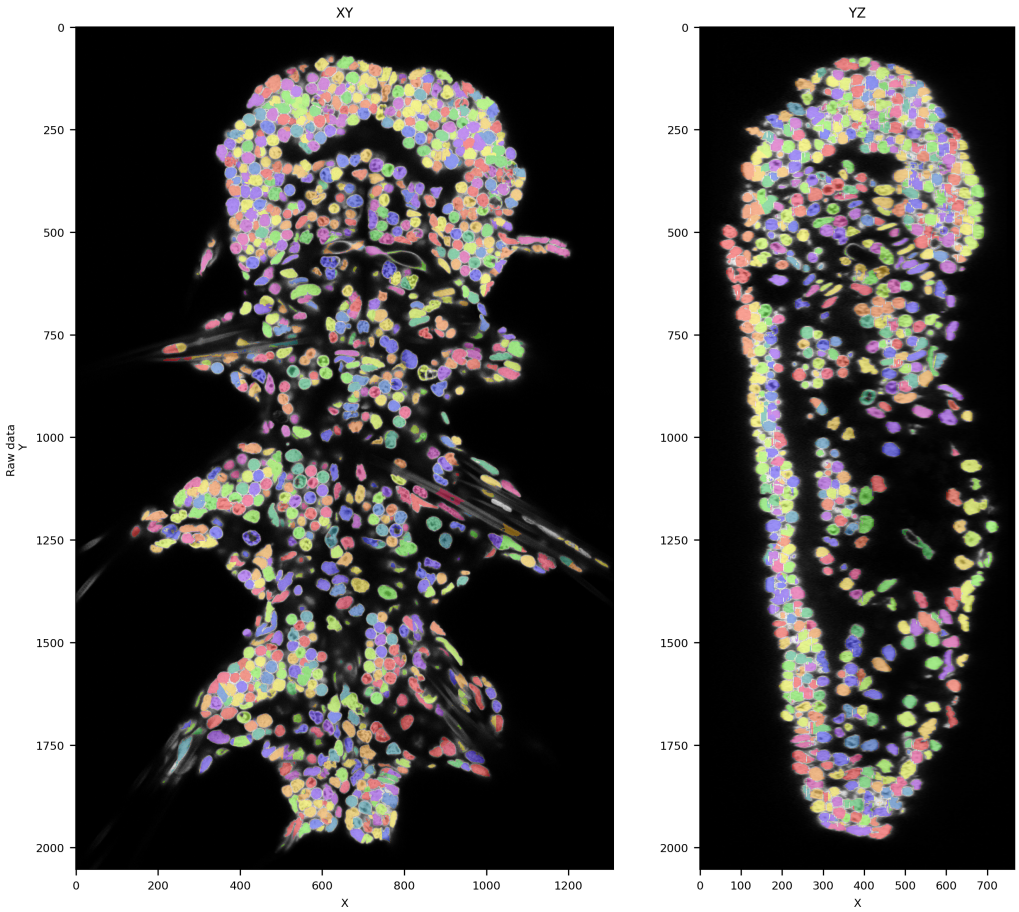


Figure 10: Final merged segmentation of a full volume.

Anything Model [56] and Plantseg [57] splitting/merging based on boundary predictions. However, no pre-trained model was good enough to make initial predictions for correction, therefore annotation was mostly done from scratch using the functionality of the default labels layer. Manual instance segmentation was very difficult and in many cases the boundary between two nuclei can be quite arbitrary with estimated time per slice reaching 40 minutes for slices full of objects. In total 580 nuclei were segmented in ROI 1 and 67 nuclei in ROI 2 in 114 and 118 Z slices accordingly.

3.4.3 *Segmentation quality metrics*

- `Number of nuclei`
- `Semantic intersection over union (IoU)`
- `Precision, recall and accuracy at` $\mathrm{IoU} = 0.1$ `and` $\mathrm{IoU} = 0.5$ To compare instance segmentation with ground truth, intersection over union (IoU) is calculated between each mask in the instance segmentation and the ground truth, then instances which have IoU higher than a threshold are considered matched.

  True positive (TP): number of objects matched between ground truth and instance segmentation

False negative (FN): number of objects in the ground truth that did not match

False positive (FP): number of objects in the instance segmentation that did not match

Precision, recall and accuracy are calculated based on this definition of TP, FN and FP. The higher is the the threshold IoU, the more precisely the objects need to be segmented to count as matching. $IoU = 0.5$ is a value commonly used for evaluation of the instance segmentation. At $IoU = 0.1$ the objects do not need to have very similar shape so the metrics at this threshold can be interpreted as characterizing object detection quality, even if the predicted nuclei shapes are significantly different from the ground truth.

- `Chamfer distance`

  Chamfer distance [58] is metric used to evaluate the similarity between two point sets. Given two point sets A and B, the chamfer distance is defined as the average of the distances from each point in A to its nearest neighbor in B, plus the average of the distances from each point in B to its nearest neighbor in A. This metric characterizes how close on average a point from each of the point clouds is located to the closest point in the other point cloud. This metric is used because centroids of the segmented nuclei are meant to be used for point cloud registration.

3.4.4 *Merging Stardist and Mutex Watershed predictions*

Stardist segmentation was performed using a model trained by Qin Yu using 3D nuclei segmentation dataset annotated using a procedure described in [59]. As shown in 9, Stardist produces segmentation that does not follow the shape of the nuclei - IoU for the foreground prediction is only 0.68 (Table 1) - but roughly matches the correct density of objects as Chamfer distance is pretty low - smaller than the nucleus size. Using this model results in oversegmentation, especially for the nuclei that do not conform to the star-convex inductive bias.

Multicut, the method which was successfully used for segmentation of nuclei in the corresponding EM volume, does not have such strong inductive bias. It requires foreground and boundary semantic segmentation as input. Boundary segmentation is used to split the volume into separate objects. However, as can be seen in Figure 8, boundary evidence can be weak as it is not easy to distinguish between nuclei in Z direction due to lower resolution. At the same time, narrow and curved nuclei get oversegmented because even perfectly correct boundary prediction takes up significant volume inside the nucleus. These two factors create a tread-off: either sensitivity to boundary evidence is high, then it is possible to reduce undersegmentation in the brain of the animal, but narrow nuclei get split, or sensitivity is low, then narrow nuclei are segmented correctly but nuclei in the more crowded regions get merged. Even converting instance segmentation to foreground and boundary probability maps and then applying multicut did not lead to fully correct instance segmentation and this result only got worse when taking the network prediction errors into account.

METRIC	ROI	STARDIST	MUTEX WS	MERGED	GT
N nuclei	ROI1	791	506	**561**	580
	ROI2	140	45	**79**	67
Semantic IoU	ROI1	0.68	0.79	**0.81**	-
	ROI2	0.45	0.67	**0.68**	-
Precision @ 0.1	ROI1	0.85	0.98	**0.99**	-
	ROI2	0.57	**0.85**	**0.85**	-
Recall @ 0.1	ROI1	0.98	0.93	**0.99**	-
	ROI2	0.95	0.88	**0.97**	-
Accuracy @ 0.1	ROI1	0.84	0.92	**0.98**	-
	ROI2	0.55	0.76	**0.82**	-
Precision @ 0.5	ROI1	0.35	0.61	**0.68**	-
	ROI2	0.14	**0.54**	0.39	-
Recall @ 0.5	ROI1	0.48	0.53	**0.66**	-
	ROI2	0.27	0.39	**0.41**	-
Accuracy @ 0.5	ROI1	0.25	0.4	**0.5**	-
	ROI2	0.1	0.29	**0.25**	-
Chamfer distance, μm	ROI1	1.3	1.1	**0.9**	-
	ROI2	2.7	1.7	**1.5**	-

Table 1: Segmentation quality evaluation

Mutex watershed [60] is a method designed for segmentation of neurons in EM data. The neurons have soma which is a large object and axons/dendrites which are narrow and need to be segmented very accurately to enable automated tracing. Mutex watershed is similar to multicut but in addition to foreground and boundary it also takes into account affinity predictions. Affinity with offset $(\Delta x, \Delta y, \Delta z)$ is defined as a probability for a given pixel at the position (x_0, y_0, z_0) that the pixel at the position $(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ belongs to a different instance. This allows to do instance segmentation of both small and large objects in the same dataset, however, with more channels the predictions become more memory and time-consuming.

U-Net model was trained using Plantseg [57] to predict foreground, boundaries and a set of affinities with the following offsets:

$$(1,0,0), (0,1,0), (0,0,1), (3,0,0), (0,3,0), (0,0,3), (9,0,0), (0,9,0), (0,0,9).$$

Figure 8 illustrates the predictions of the model and Figure 9 shows the final instance segmentation after post-processing. Mutex watershed largely segments both the cells with challenging shapes and round nuclei in the crowded regions correctly but there are cases when the nuclei are merged in Z direction, resulting in undersegmentation: mutex watershed detects 506 objects in ROI1 while the correct number is 580 (see Table 1). Visual inspection of the erroneous masks confirms that without knowing the approximate object size it is virtually impossible to split these nuclei correctly without knowing the approximate expected size of the objects.

Both undersegmentation and oversegmentation are detrimental for registration, therefore a heuristic was used to combine Stardist and mutex watershed segmentations. Firstly, both Stardist and mutex watershed were applied to the image as usual. Then in the mutex watershed segmentation instances larger than a certain threshold size were removed. After that Stardist instances were used as seeds to run watershed on the boundary predictions using the removed undersegmented instances as a foreground mask. This procedure allowed to split the large masks into separate nuclei of the correct size while keeping the overall correct foreground shape. As a final step, size filtering was applied to remove masks which were too small to be nuclei.

The final segmentation ("merged" in Figure 9) correctly estimates the number of instances in the ground truth ROIs (Table 1). Precision and recall @ 0.1 as well as low Chamfer distance show that the point cloud derived from the segmentation accurately corresponds to the centroid point cloud of the ground truth segmentation.

3.4.5 *Results and limitations*

Nuclei instance segmentation of the full confocal volume is shown in the Figure 10. Although there is no ground truth annotation for the whole animal, visual inspection shows that the segmentation has consistent quality throughout the volume regardless of intensity variations and difference in resolution between imaging plane and optical slicing direction. The number of nuclei segmented in confocal volumes is around 10600 while EM segmentation has 11477 nuclei. This means that there are some undersegmented instances, although

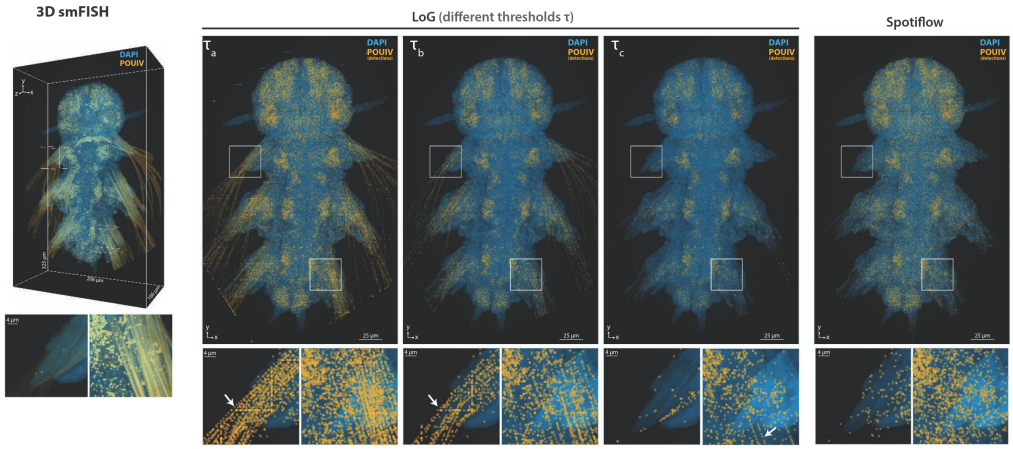


Figure 11: Comparison of LoG spot detection using different threshold with Spotiflow performance on full animal confocal volume of *Platynereis dumerilii*. In the volume render on the left DAPI is shown in blue and raw Brn3 signal in orange. Images on the right show DAPI in blue and spot detection with LoG and Spotiflow in orange. Reproduced from [61].

this discrepancy can be partly explained by the segmentation errors not only in confocal volumes but also in EM segmentation. Despite training the model on a subvolume with autofluorescence, some part of chaetae and the jaw of the animal get segmented, however, these are minor mistakes compared to the total number of nuclei.

The downside of the developed segmentation procedure is the increased processing time and memory consumption. Applying two separate segmentation methods means that neural networks need to be run on the volume twice. The affinity prediction with 11 channels needs to be stored, taking up 100-150 GB for images of size $\sim 800 \times 2000 \times 1500$. Postprocessing, which includes watershed and size filtering, takes about 40 minutes per sample, adding up to the pipeline execution time. Nevertheless, increased demand for computational resources is justified in this case because the segmentation works at the limit of what is possible with this quality of the data.

3.5 SPOT DETECTION

3.5.1 *Problem description*

In smFISH the staining protocol produces images which consist of spots, with each spot corresponding to an mRNA of the gene of interest. Due to intensity variations, autofluorescence and crowding of the spots in cells with high expression levels it is hard to quantify the expression solely by comparing fluorescence intensity. Detection of individual spots allows to compensate for these factors and quantify gene expression based on smFISH images.

In many widely used pipelines ([62], [63]) spot detection relies on Laplacian-of-Gaussian (LoG) - a classical threshold-based method. It works well for the clean data with clearly distinct spots but using it becomes more difficult in the images with varying background and signal intensity, leading both to many

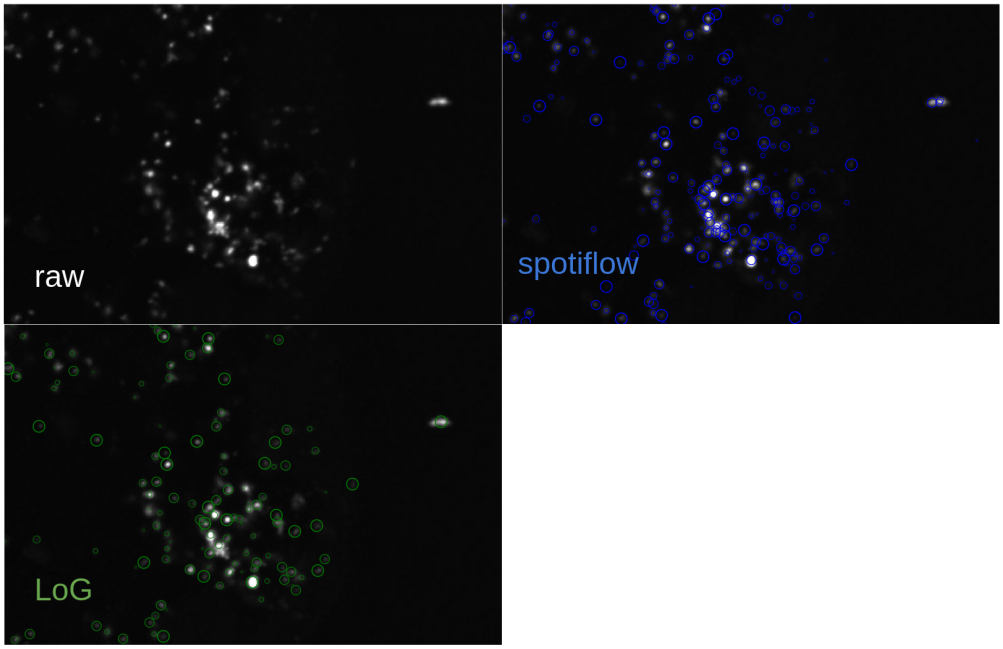


Figure 12: Comparison of spot detection with LoG and Spotiflow for the cells with high gene expression level. Raw signal is shown in white. Spot detections are shown as circles (blue for Spotiflow and green for LoG), smaller circle diameter means that the center of the spot is not in the plane of the image.

false-positive detections in the background and many false-negatives in the crowded areas and areas with lower intensity of the signal.

Testing LoG on the *Platynereis* smFISH volumes has shown that while it is possible to reach decent results by manually adjusting thresholds on a small subvolume of the data, applying LoG with the same setting to the full volume led to variable results throughout the volume. Autofluorescence from chaetae was causing especially large number of false-positives while lowering the sensitivity led to missing detections as shown in Figure 11. In addition, LoG performed poorly in the crowded areas (Figure 12), missing many spots and therefore making it impossible to correctly identify cells with high expression levels.

A lot of deep learning-based methods were developed to solve the spot detection task ([64–67]) but they are hard to use as most of them require an accurate PSF model for creating simulated training data and work only in 2D.

3.5.2 *Spot detection ground truth annotation*

Subvolumes for annotations were chosen from 3 biological replicates, 3 subvolumes per sample. As each sample has 2 gene samples, this resulted in 18 different annotated subvolumes. The subvolumes were chosen to include challenging areas, such as chaetae and cells with high expression levels.

Firstly, the spots in each image were detected using LoG, with the threshold adjusted manually per image. Then the LoG detections were corrected using the functionality of Points layer in Napari. The annotation took approximately 16 hours. It is important to note that these annotations are highly subjective. It

is impossible to determine the exact position of the spot center in a diffraction-limited image. The only way to get the ground truth with precise spot positions is to use simulated data. In spot detection there is a tread-off between the accurate ground truth and the training data which faithfully represents intensity variations, artifacts and noise of the real data.

3.5.3 *Spot detection with Spotiflow*

Annotated data was passed to Albert Dominguez Mantes who used it to develop the 3D version of Spotiflow [61], a deep learning-based method for spot detection. In Spotiflow the position of the spots and the distance to the closest spot is predicted using a CNN, after which post-processing is applied to detect spots and determine their sub-pixel position. The CNN can be trained both using simulated data and using real data, allowing the model to adjust to the data-specific intensity variations and the presence of the fluorescent objects which are not spots. Another advantage of this method is that it uses the 3D CNN, making it easier to distinguish neighboring spots in crowded areas.

Albert trained the 3D model using simulated data and then fine-tuned it using the manual annotations. Resulting model drastically decreased the number of false-positive in chaetae (Figure 11) and performed better than LoG in the cells with high gene expression (Figure 12). More detailed analysis of the results and the method can be found in [61].

Unfortunately, it was not possible to completely eliminate false positives caused by the sample autofluorescence. In addition, inspection of samples stained with different smFISH probes revealed that the cuticle of the animal gets stained in some areas, resulting in the signal that an expert would identify as "false positive". However, the spots look exactly the same as for the correct detections, making it impossible for a local method such as Spotiflow to determine if these detections are correct or incorrect. In such cases it was decided to retain as much signal as possible and perform the filtering during the single-cell gene expression data analysis.

Having a robust model for spot detection allowed to quantify gene expression signal and easily aggregate the signal from different replicates. As discussed above, the raw intensity depends on the rotation of the sample with respect to the imaging plane as well as on the position inside the sample, making it hard to connect intensity values with the expression level. Spot counts are robust to these variations and can be directly compared between replicates imaged with the same smFISH probes in different conditions.

3.6 GENE SIGNAL ASSIGNMENT

After segmentation and spot detection gene signal is assigned to individual nuclei. The volume was split into regions by the distance to the closest nucleus, approximating the cytoplasm of the cell. After that the number of spots detected in each region was calculated and assigned as the gene expression value to each nucleus, as shown in Figure 13. For genes such as StMhc and ROpsin which have so high expression level that it is impossible to detect individual spots, sum and mean of intensity was calculated over each cell region.

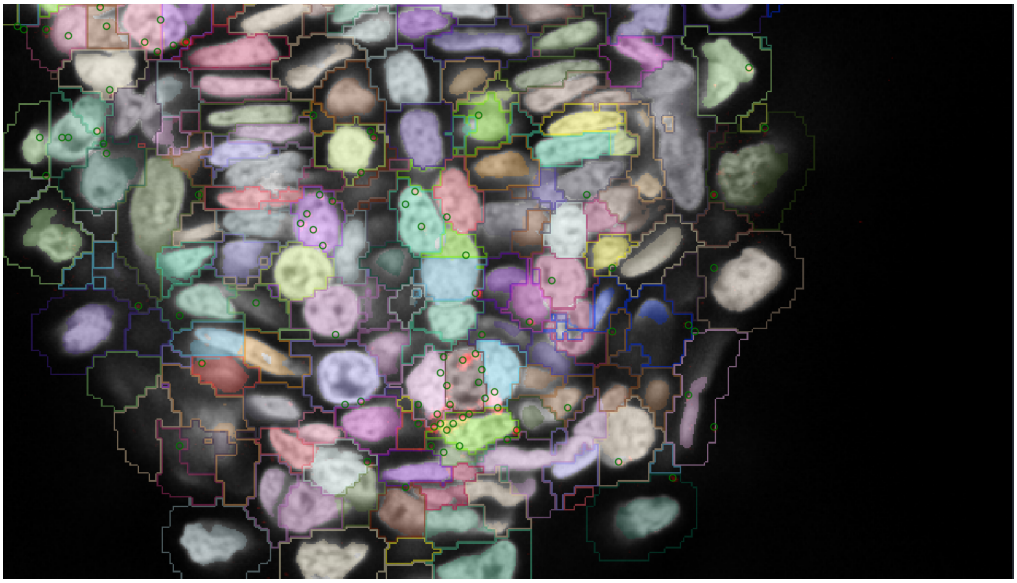


Figure 13: Gene signal assignment to nuclei. DAPI signal (white) with the expression of gene Pkd1 overlaid on top (red). Nuclei are colored according to their segmentation masks and corresponding cell regions are shown contours of the same color. Spots detected by Spotiflow are marked with green circles.

This approach has many disadvantages. It does not take into account the shape of individual cells which can lead to errors. In cases when the cells are located very close to each other, it is impossible to determine if a spot close to the cell boundary should belong to one cell or the other. Potentially a more sophisticated approach could help to take this into account. For example, if one nucleus has many spots assigned to it with high confidence then it is more likely that the spot on the boundary belongs to it rather than to a neighbor with no expression.

3.7 PRELIMINARY SIMILARITY ALIGNMENT

EM segmentation was used as a template for registration. Both EM and confocal volume segmentations were converted from the instance to semantic. Thanks to the SVD pre-alignment step, all the samples were already roughly aligned, which allowed to compare them before registration. All the light samples looked consistently similar but EM volume was approximately 10% smaller than light volumes, based on the nuclei size and the length of the animal body. This difference can be attributed to different fixation protocols. Similar observation was also made in the pilot stage of the project when the samples fixed in different way for confocal microscopy had different size.

To compensate for this difference and make further deformable registration more stable, pre-aligned images were registered using Similarity transform in Elastix [12, 68]. Similarity transform allows translation, rotation and scaling of the moving image. The registration was done using normalized correlation coefficient with 2 levels in the image pyramid schedule (32 and 8) and 20000 spatial samples.

Figure 14: Point cloud alignment with Coherent Point Drift. Each line shows the displacement of one nucleus.

3.8 COHERENT POINT DRIFT

Both EM and light segmentations were converted to point clouds by extracting coordinates of centroids. Coherent Point Drift [69] was used to perform deformable alignment with EM point cloud as a fixed set and light point cloud as moving point set.

In Coherent Point Drift registration problem is posed as follows: maximize probability of fixed point set as samples from the probability distribution of a sum of Gaussians with centers in the points of the moving set:

$$p(\mathbf{x}) = w\frac{1}{N} + (1-w)\sum_{m=1}^{M}\frac{1}{M}p(\mathbf{x}|m), \tag{5}$$

N, M - number of points

$p(\mathbf{x}|m)$ - probability of one point in fixed set x given the position of a moving point m

w - probability of the point being noise (coming from uniform distribution instead of sum of Gaussians)

Non-rigid registration: find displacement field $v(Y)$:

$$T(\mathbf{Y}, v) = \mathbf{Y} + v(\mathbf{Y}), \tag{6}$$

$\mathbf{Y}$ - positions of points in the moving point cloud.

This algorithm has the following parameters:

- w - expected proportion of noise in the dataset
- λ - tradeoff between maximizing probability of points and keeping the spatial structure of the moving point set
- β - parameter defining the width of Gaussians

Registration was performed with parameters $w = 0.00001$, $\lambda = 0.1$, $\beta = 100$ and $max\_iter = 150$ using library Probreg [70]. Registration uses GPU and takes $\sim$ 30min for two point sets with $\sim$ 11000 point each.

Figure 14 demonstrates the displacement field found with CPD for a subset of nuclei. Neighbouring nuclei move consistently as a group, for example, the whole parapodia moves as a whole, indicating that the animal in the confocal volume was imaged in a different pose.

Using point cloud registration allows to reason about the content of the image as a set of objects rather than mere pixel intensities, which matches the human perception of what it means to match two stereotypic animals.

3.9 CELL MATCHING

3.9.1 *Assignment problem formulation*

Coherent Point Drift finds the coordinates of points of the aligned point cloud so additional matching step is necessary to establish correspondences between nuclei from EM and nuclei from the confocal volume. The task of finding the optimal correspondence between the elements of two sets based on the pairwise distance matrix is called assignment problem. It is a fundamental task of combinatorial optimization.

Let M be number of elements in the first set and N - the number of elements in the second set, $M < N$. Let us define the matrix of optimization variables X of size $M \times N$:

$$x_{ij} = \begin{cases} 1, & p_i \text{ matches } m_j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

C - cost matrix of size $M \times N$. In this case C consists of pairwise Euclidean distances between points from different point clouds.

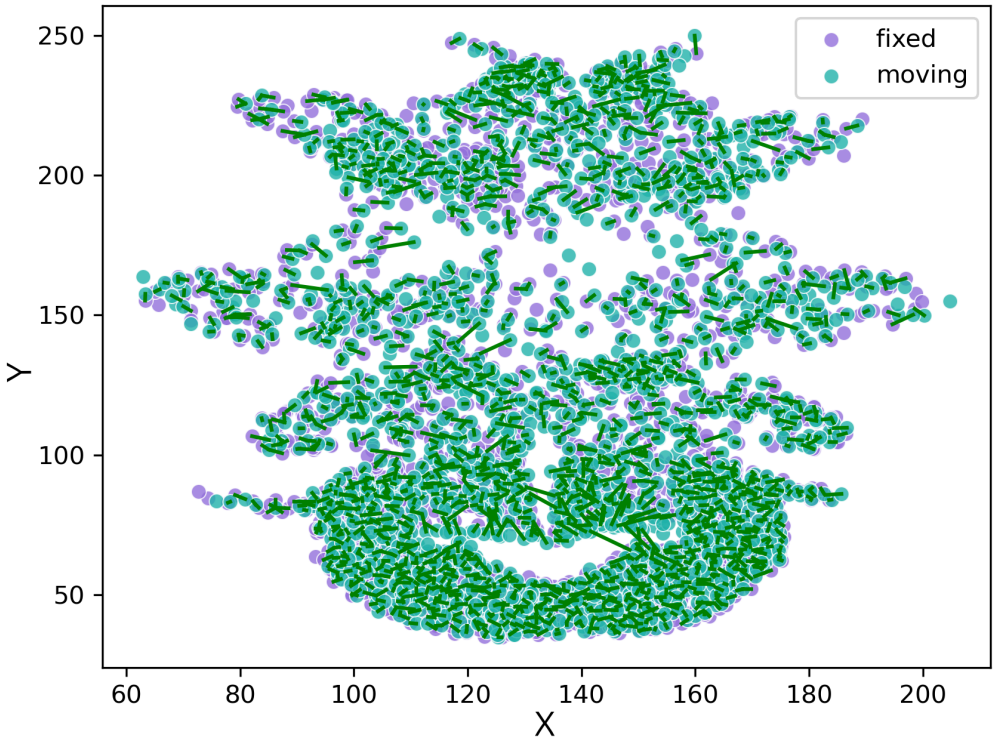


Figure 15: Point matching with ILP. Subset of EM point cloud is shown in purple, subset of light point cloud - in teal. Matched nuclei are connected with green lines.

Objective function:

$$\sum C \cdot X \rightarrow \min \tag{8}$$

Constraints:

$$\sum_{i=1}^{M} X_{i\cdot} >= 1$$

$$\sum_{j=1}^{N} X_{\cdot j} == 1$$

The first constraint means that each point from the smaller point cloud should match to at least one point from the larger point set. The second constraint means that each point from the larger point set must match exactly one point in the smaller point cloud.

Optimization of this objective was implemented using CVXPY [71, 72], an open source package for disciplined convex programming and combinatorial optimization. GLPK (GNU Linear Programming Kit) was used to solve this mixed integer problem.

3.9.2 *Scaling up with a sparse distance matrix*

With approximately 11000 elements in each set, the total number of optimization variables reaches $121 \cdot 10^6$. Solving a problem of this dimensionality is not

feasible therefore the number of considered matches had to be reduced. Since the point clouds are already aligned it is expected that apart from some registration errors the matching nuclei are located close to each other, therefore it is reasonable to limit possible matches to some nearest neighbors of each nucleus.

Based on experiments with synthetic data, sparse distance matrix contained the element if the distance between two nuclei was less than 8 µm. In addition, for each nucleus 30 nearest neighbors were considered as possible matches regardless of the pairwise distance to ensure that the optimization problem always has feasible solutions. With this assumption the number of optimization variables reduced from $\sim 121 \cdot 10^6$ to $\sim 6 \cdot 10^5$. With the reduced number of variables it took approximately 30 minutes to solve this problem using CVXPY and establish the correspondence between nuclei in the EM dataset and confocal volumes (Figure 15).

The formulation of the problem used here ensures that each nucleus in EM and each nucleus in the confocal data will have at least one corresponding object, which is important for transferring annotations such as gene expression and morphology descriptors between the datasets. Using additional optimization procedure for this step helps to compensate for errors in Coherent Point Drift registration.

3.10 DEFORMABLE REGISTRATION WITH MATCHED NUCLEI CENTROIDS AS LANDMARKS

Matching nuclei provide a set of landmarks which can be used to warp the confocal images and overlay them with the EM data.

Thin plate splines are often used for this task as this transform parametrization uses the landmarks as control points [8, 73]. Due to unequal number of segmentation masks, some nuclei essentially match to the same point so instead of enforcing the position of matched points with thin plate splines, B-splines were used to parametrize the deformation and the position of the landmarks was used in the optimization objective instead.

Deformable registration was performed in Elastix with a sum of two metrics as an objective: normalized correlation coefficient between semantic segmentations and Euclidean distance between corresponding points with the weights 1 and 1000. The registration was done with 2 levels in the image pyramid schedule (16 and 8) and 100000 spatial samples. Resulting transformation was applied to the instance segmentation as well as all channels of the raw confocal volumes.

3.11 EXPORT TO MOBIE

MoBIE [74] is a data viewer for exploration of big multi-modal images and associated tabular data based on Big Data Viewer (BDV) [75]. It allows to smoothly browse large images thanks to pyramidal data storage and can be used to map annotations stored in tables to the volumetric segmentation. It was used to interact with the *Platynereis dumerilii* EM volume in the previous

publications [36, 37], therefore the newly registered datasets were added to the already existing MoBIE project.

Transform found in the previous steps was applied to the nuclei segmentation, raw DAPI image and both gene images. After that another rigid transformation was applied to all datasets to orient them correctly as the EM volume in the existing MoBIE project is not oriented along the space axes. Resolution of $0.32 \times 0.32 \times 0.32$ µm was used for all the new datasets to reduce the storage memory.

Adding data to MoBIE browser allowed to overlay EM with registered fluorescence and visualize gene expression values transferred to EM nuclei from multiple confocal replicates.

4 GENE EXPRESSION ANALYSIS

4.1 PIPELINE OVERVIEW

The image analysis pipeline for atlas creation can be split into 3 parts: preprocessing, registration and data visualization. Figures 17, 19, 18 and 16 demonstrate the application of the pipeline to 4 different samples. It is important to point out that these are 4 individual animals which do not necessarily match exactly. Let us briefly summarize all the steps of the registration pipeline.

4.1.1 *Preprocessing*

In the preprocessing phase each confocal dataset is processed independently. The goal of this phase is to enhance and quantify the raw signal and extract objects for registration. This part of the pipeline is highly specific for the data used in this project as it includes the deep learning methods trained using manual annotations of these particular images.

- *Pre-alignment*

  Imaged samples can be oriented randomly in the imaging plane (Figure 16, A). Automatically finding large rotation angles is hard and prone to fail, therefore the pipeline starts with a heuristic-based alignment of body axes of the animal along the axes of the coordinate space: dorsoventral axis along Z, anteroposterior axis along Y and left-right axis along X, with the head of the animal close to the origin. As a result, all the samples are roughly aligned and easy to compare and visualize (Figure 16, B).

- *View merging*

  In fluorescence microscopy, the illumination as well as the detected light is subject to scattering in biological tissues, limiting the penetration depth of the imaging. Decreasing signal intensity makes automated segmentation and spot detection much more challenging and leads to the loss of gene expression in- formation. To compensate for the loss of signal, each animal was imaged from two sides (dorsal and ventral), after which the views were merged. This resulted in a composite volume with uniform intensity, making further processing easier and preventing loss of gene expression signal.

- *Segmentation*

  Each sample was segmented with Stardist and mutex watershed. Thanks to the strong shape prior, Stardist could detect separate nuclei in the areas of the sample where resolution in Z did not allow to reliably predict a boundary between separate nuclei. Mutex watershed was able to accurately segment nuclei of unusual shape and overall segmented the foreground more precisely. The instance segmentations obtained with these

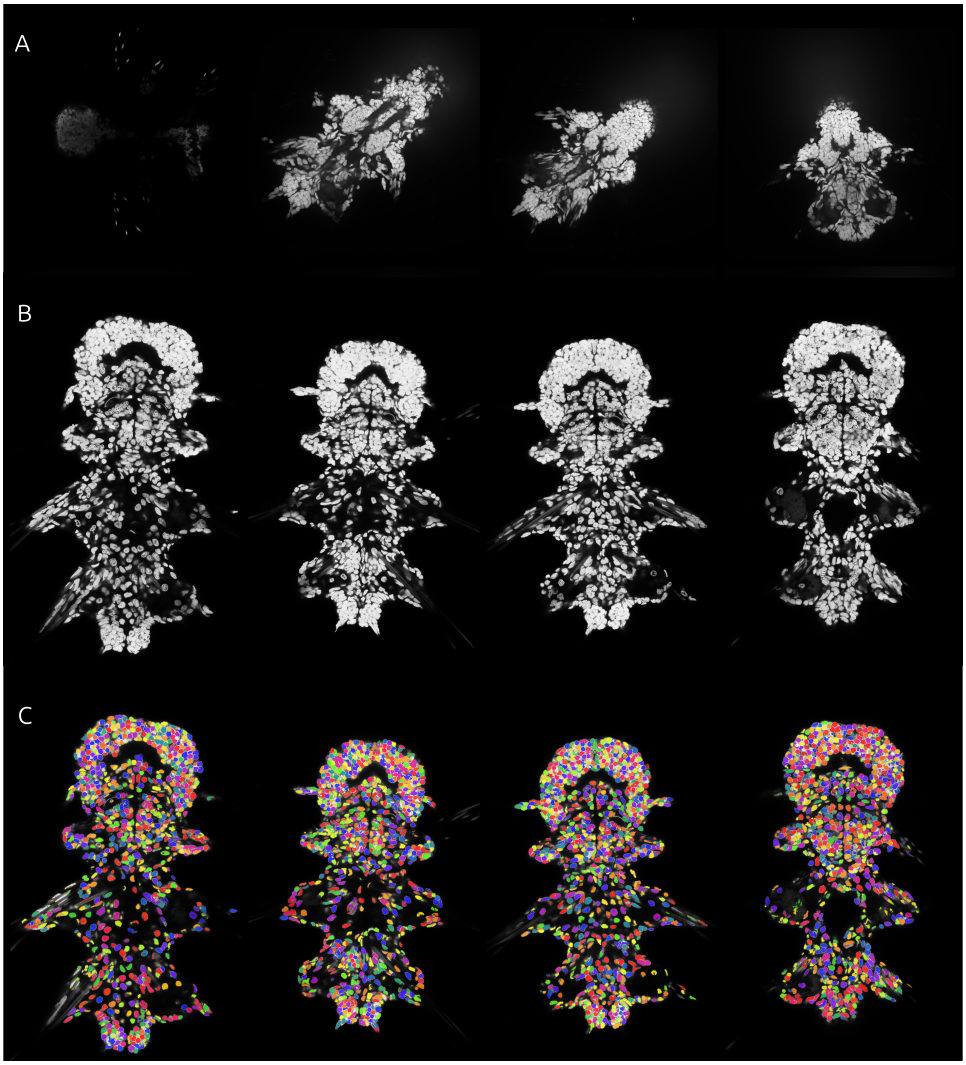


Figure 16: Preprocessing phase of the registration pipeline shown for 4 different animals. A) The same Z slice of each of the input volumes (DAPI staining in white). B) Pre-aligned volumes (DAPI staining in white). C) Instance segmentation (DAPI staining in white, each nucleus mask has different random color).

two methods were combined resulting in ~ 10500 nuclei detected in each sample. Despite the errors, especially caused by the autofluorescence, the segmentation was consistent throughout the volume which is important for using it as a source of registration landmarks (Figure 16, C).

- *Spot detection*

  Spot detection was performed using a custom 3D Spotiflow model. Having a robust model for spot detection allowed to quantify gene expression signal and easily aggregate the signal from different replicates despite intensity variations across and within the samples.

- *Gene signal assignment*

  The volume was split into regions by the distance to the closest nucleus, approximating the cytoplasm of the cell. After that the number of spots

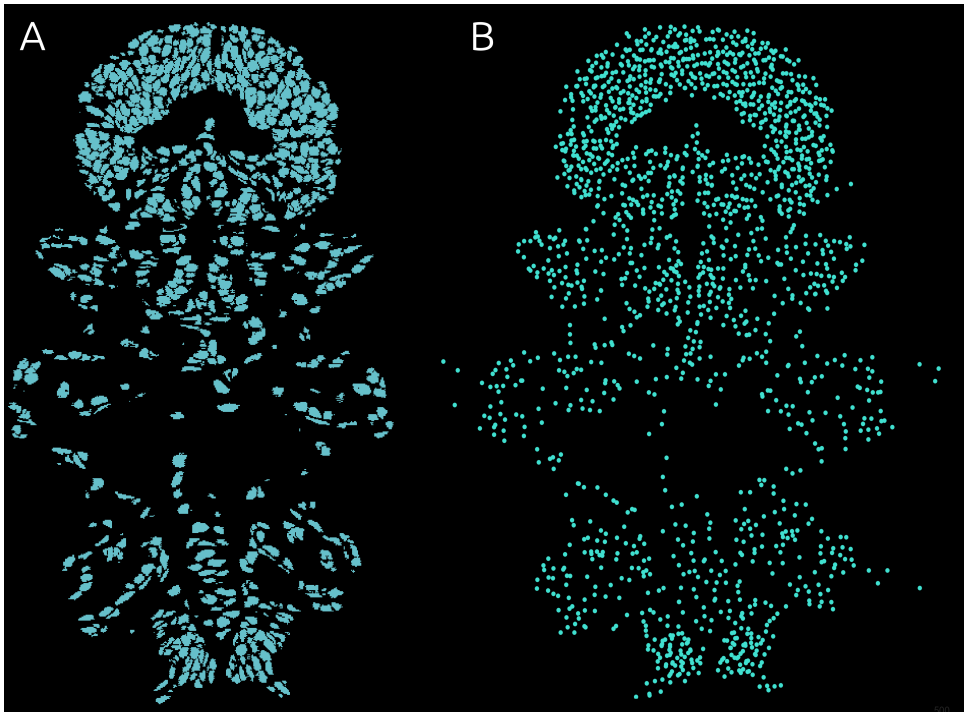


Figure 17: Pipeline EM template

detected in each region was calculated and assigned as the gene expression value to each nucleus.

4.1.2 *Registration*

In this phase all confocal volumes are registered to the EM volume, used as a template. EM segmentation (Figure 17, A) and corresponding centroid point cloud (Figure 17, B) were set as fixed. The main result of this part of the pipeline is the displacement field and nucleus-to-nucleus matching between each nucleus in each of the confocal volumes and the nuclei in EM volume.

This part of the pipeline is fairly generic and can be applied to different biological samples, given that the instance segmentation is already done.

- *Preliminary similarity alignment*

  Before doing deformable alignment, it is necessary to rigidly align the samples to the template as close as possible. In this case similarity alignment (translation + rotation + scaling) was used to compensate for the size difference caused by different fixation and staining protocols. After this step the samples are registered as closely as possible without deformable alignment (Figure 18, A).

- *Coherent Point Drift*

  Both EM and light segmentations were converted to point clouds by extracting coordinates of centroids (Figure 18, B). Coherent Point Drift [26] was used to perform deformable alignment with EM point cloud as a fixed set and light point cloud as moving point set (Figure 18, C) resulting in the aligned point clouds (Figure 18, D).

- *Cell matching*

  Coherent Point Drift finds the coordinates of points of the aligned point cloud so additional matching step was implemented to find a correspondence between nuclei in light and EM. The total distance between matched nuclei was minimized using CVXPY with a constraint that each nucleus in both datasets should match one or more nuclei in the other dataset (Figure 19, A). This matching enabled annotation transfer between the datasets, for example, assignment of gene expression based on spot detection to nuclei in EM.

4.1.3 *Data analysis and visualization*

Final result of registration is the matching between nuclei in confocal volumes and EM template. It is very hard to assess the registration quality based on the point clouds, therefore the last phase was added to the pipeline to overlay raw fluorescence data with the electron microscopy.

- *Deformable registration with matched nuclei centroids as landmarks*

  Matching nuclei provide a set of landmarks which can be used to warp the confocal images and overlay them with the EM data. Deformable transformation was found using Elastix and applied to the segmentation (Figure 19, B) as well as all channels of the raw confocal volumes (Figure 19, C).

- *Visualization in MoBIE*

  Registered instance segmentation as well as all channels of the raw confocal volumes and the final gene expression table were added to the existing MoBIE project. This allowed to interactively visualize multiple datasets together, perform arbitrary direction slicing and visualize the mapping of the quantified gene expression signal to the segmentation masks.

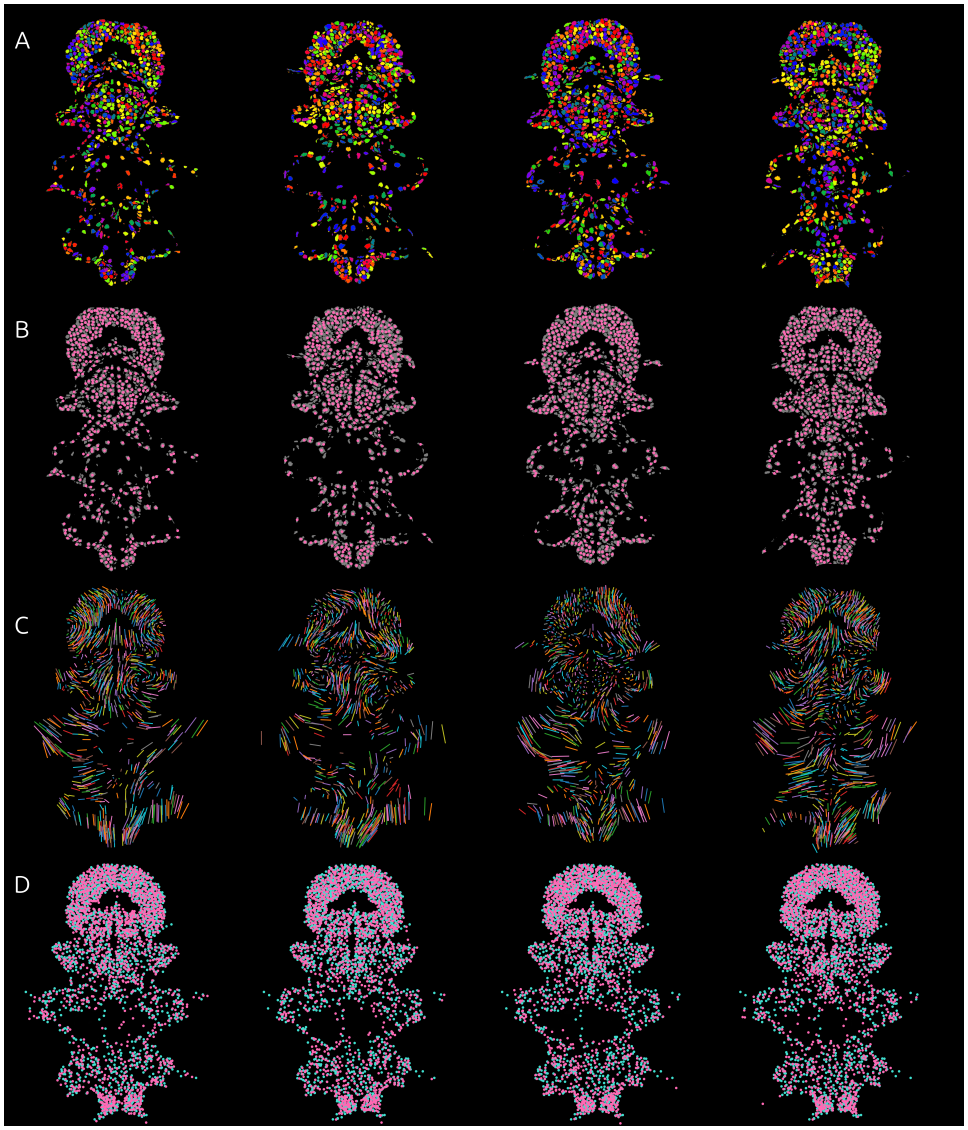


Figure 18: Registration phase of the registration pipeline shown for 4 different animals. A) Volumes registered to EM segmentation using similarity transform (each nucleus mask has different random color). B) Centroid-based point clouds (segmentation shown in gray, point cloud in pink). C) CPD displacement field for a subset of nuclei. Each line is displacement of one nucleus from it position after similarity transform to the final position. D) Light point clouds (pink) registered with EM template point cloud (cyan).

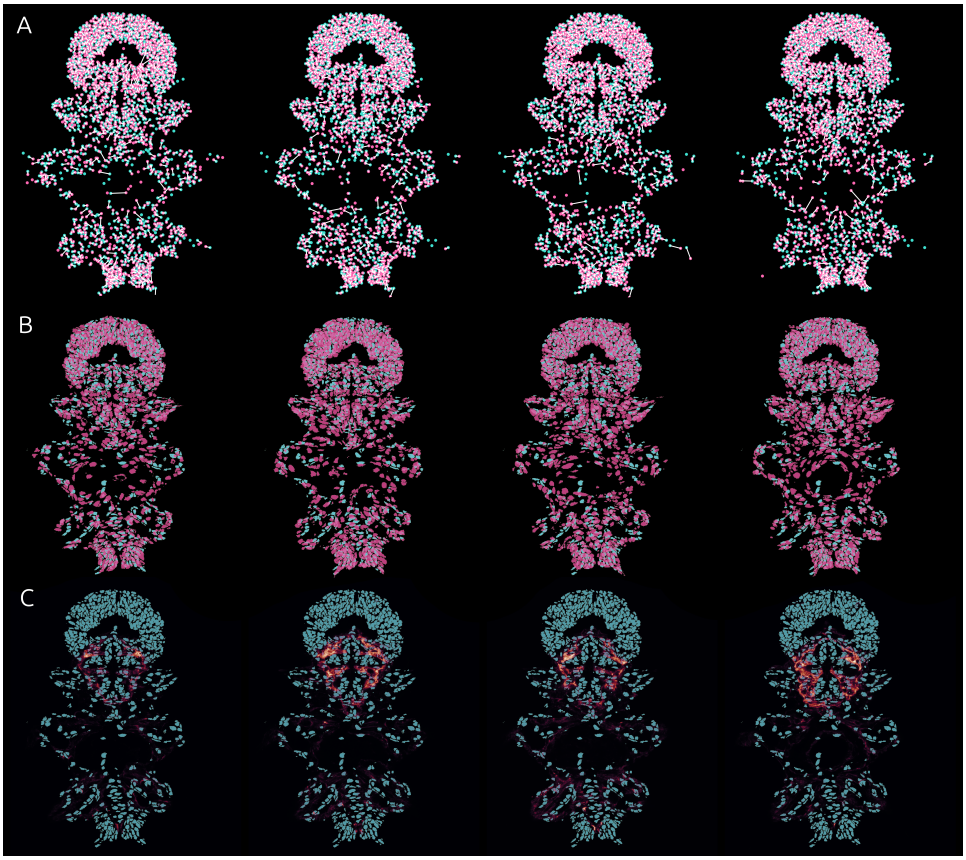


Figure 19: Visualization phase of the registration pipeline shown for 4 different animals. A) Points of the EM point cloud (cyan) matched with the points of the confocal point cloud (pink), matches shown with white lines. B) Deformable transform applied to the semantic segmentation of confocal volumes (pink) to visualize the overlay with EM segmenation template (cyan). C) Deformable transform applied to the muscle staining (one of the gene channels in the confocal volumes) to visualize the overlay with EM segmenation template (cyan).

SAMPLE NAME	GENE 1	GENE 2
ropsin1_pl1	ROpsin	-
ropsin1_pl2	ROpsin	-
ropsin1_pl3	ROpsin	-
ropsin1_pl4	ROpsin	-
StMhc_pl1	StMhc	-
StMhc_pl2	StMhc	-
StMhc_pl3	StMhc	-
StMhc_pl4	StMhc	-

Table 2: List of confocal samples for quality control

4.2 ASSESSMENT OF REGISTRATION QUALITY

Measuring how well registration worked would require manual matching of 11500 nuclei in multiple samples, which is impossible, therefore instead of using matching nuclei to check the registration quality the gene expression signal was used to assess the registration results. Two sets of samples were acquired: 4 replicates with StMhc staining and 4 replicates with ROpsin staining, as listed in Table 2. StMhc is a very highly expressed gene and it is expressed in the muscle cells throughout the body, making it a perfect signal to assess the overall registration quality. ROpsin is expressed in the photosensitive cells, allowing to assess the registration in the head of the animal, which has less StMhc signal.

Visualization of StMhc staining registered with EM (Figure 20) shows that the developed pipeline allows to consistently align confocal volumes with EM. Even with the diffraction-limited confocal signal the replicates overlap and mostly match the muscle segmentation in EM volume. Correlation between the signal assigned to EM nuclei (Figure 21) confirms that the registration pipeline is general enough and works roughly the same for all replicates without adjusting any parameters or needing any manual input.

Visual inspection of the ROpsin stainings registered to EM (Figure 22) further proves that the signal consistently maps to the same cells. In addition to the area around the adult eye of the animal, registration accuracy allows to correctly identify a single photosensitive cell in the different part of the larvae's head (Figure 22, pink box). Correlation between the signal mapped to EM from different replicates for ROpsin is shown in Figure 23. In this case replicate 4 has lower correlation with other samples. Visual analysis did not show any difference in segmentation or registration quality based on the visualization of DAPI staining and nuclei segmentation so lower correlation could be a result of lower quality of the staining. ROpsin is very highly expressed, making it impossible to detect spots, which is why the sum of the signal per cell region was used to quantify expression level per nucleus. This example shows the importance of the spot detection for quantitative gene expression analysis of smFISH samples.

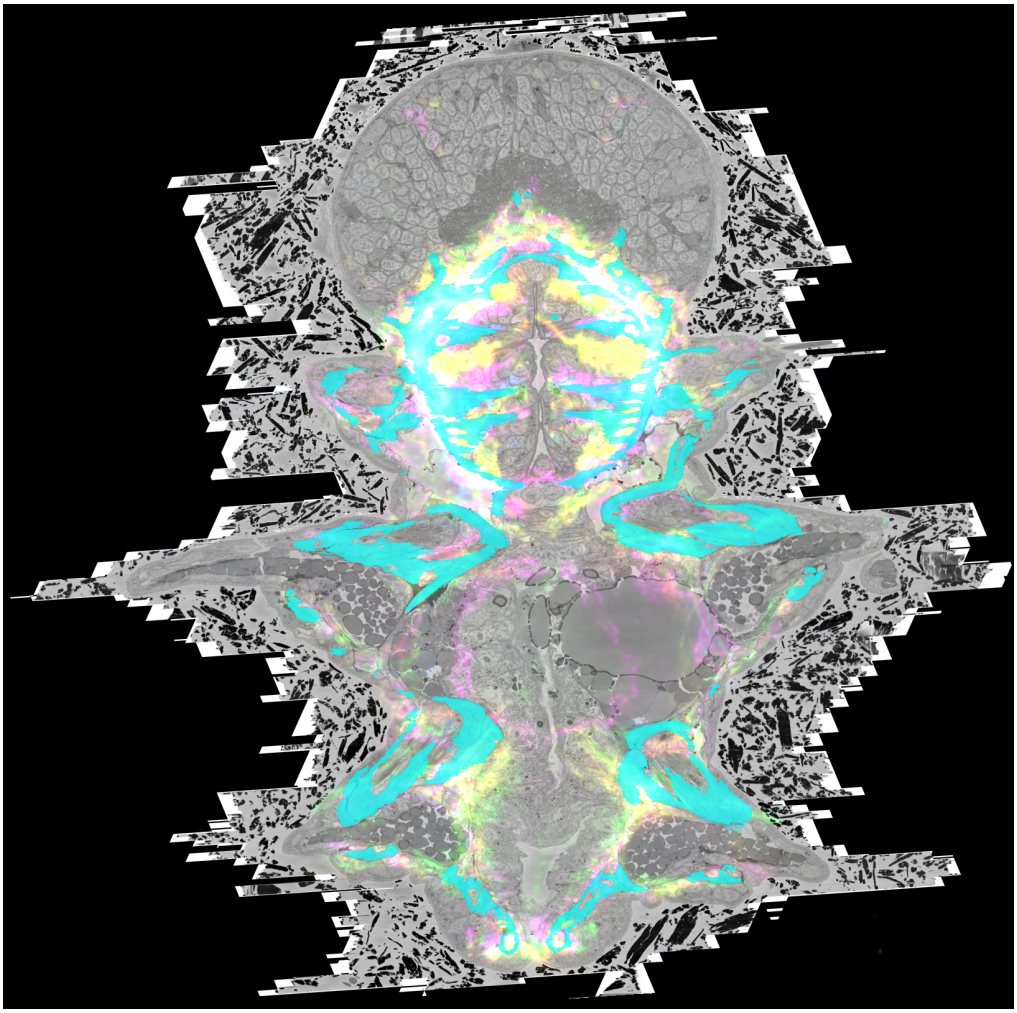


Figure 20: Registration quality control with StMhc staining. EM data shown in grey, solid cyan areas are muscle segmentation in the EM data. 4 replicates of StMhc staining from registered confocal volumes shown in pink, yellow, green and blue.

4.3 GENE EXPRESSION ANALYSIS

Developed pipeline was used to add the expression of 16 genes to the atlas by mapping 35 samples to EM. The samples are listed in Table 3. As a result of registration it is possible to overlay different replicates and view smFISH spots directly on the EM volume, as shown in Figure 24 (top panel) for genes Arx and Pkd1. In addition, the pipeline produces a table which contains an expression vector for each nucleus in EM. The expression level is quantified as a sum of spots mapped to each nucleus in all the replicates. This expression table can be loaded in MoBIE and used for visualization, for example, in Figure 24 (bottom panel) expression level of gene Pkd1 in each EM nucleus.

SAMPLE NAME	GENE 1	GENE 2
prox-brn3_pl1	Prox	Brn3
prox-brn3_pl2	Prox	Brn3
prox-brn3_pl3	Prox	Brn3
prox-brn3_pl4	Prox	Brn3
pkd1_ncald_pl1	Pkd1	NCALD
pkd1_ncald_pl2	Pkd1	NCALD
pkd1_ncald_pl3	Pkd1	NCALD
pkd1_ncald_pl4	Pkd1	NCALD
pkd1_ncald_pl5	Pkd1	NCALD
ptf1-arx_pl1	Ptf1	Arx
ptf1-arx_pl2	Ptf1	Arx
ptf1-arx_pl3	Ptf1	Arx
ptf1-arx_pl4	Ptf1	Arx
copsin1-otx_pl1	COpsin1	OTX
copsin1-otx_pl2	COpsin1	OTX
copsin1-otx_pl3	COpsin1	OTX
copsin1-otx_pl4	COpsin1	OTX
Dbx-Msx_pl1	Dbx_AP080	Msx_AP009
Dbx-Msx_pl2	Dbx_AP080	Msx_AP009
Dbx-Msx_pl3	Dbx_AP080	Msx_AP009
Dbx-Msx_pl4	Dbx_AP080	Msx_AP009
Dbx-Msx_pl5	Dbx_AP080	Msx_AP009
Vglut-Phc2_Pl1	Vglut_AP173	Phc2_AP156
Vglut-Phc2_Pl2	Vglut_AP173	Phc2_AP156
Vglut-Phc2_Pl3	Vglut_AP173	Phc2_AP156
Vglut-Phc2_Pl4	Vglut_AP173	Phc2_AP156
Vglut-Phc2_Pl5	Vglut_AP173	Phc2_AP156
FoxQ2-GoOpsin1_Pl1	FoxQ2_AP249	Go_Opsin1_AP127
FoxQ2-GoOpsin1_Pl2	FoxQ2_AP249	Go_Opsin1_AP127
FoxQ2-GoOpsin1_Pl3	FoxQ2_AP249	Go_Opsin1_AP127
FoxQ2-GoOpsin1_Pl4	FoxQ2_AP249	Go_Opsin1_AP127
Bsx-Hoxa1_Pl1	Bsx_AP203	Hoxa1_AP181
Bsx-Hoxa1_Pl2	Bsx_AP203	Hoxa1_AP181
Bsx-Hoxa1_Pl3	Bsx_AP203	Hoxa1_AP181
Bsx-Hoxa1_Pl4	Bsx_AP203	Hoxa1_AP181

Table 3: List of confocal samples

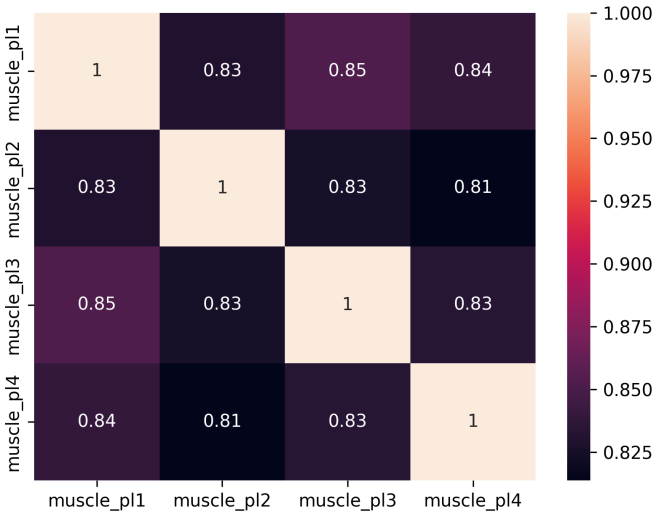


Figure 21: Correlation of the StMhc signal mapped to EM from different replicates.

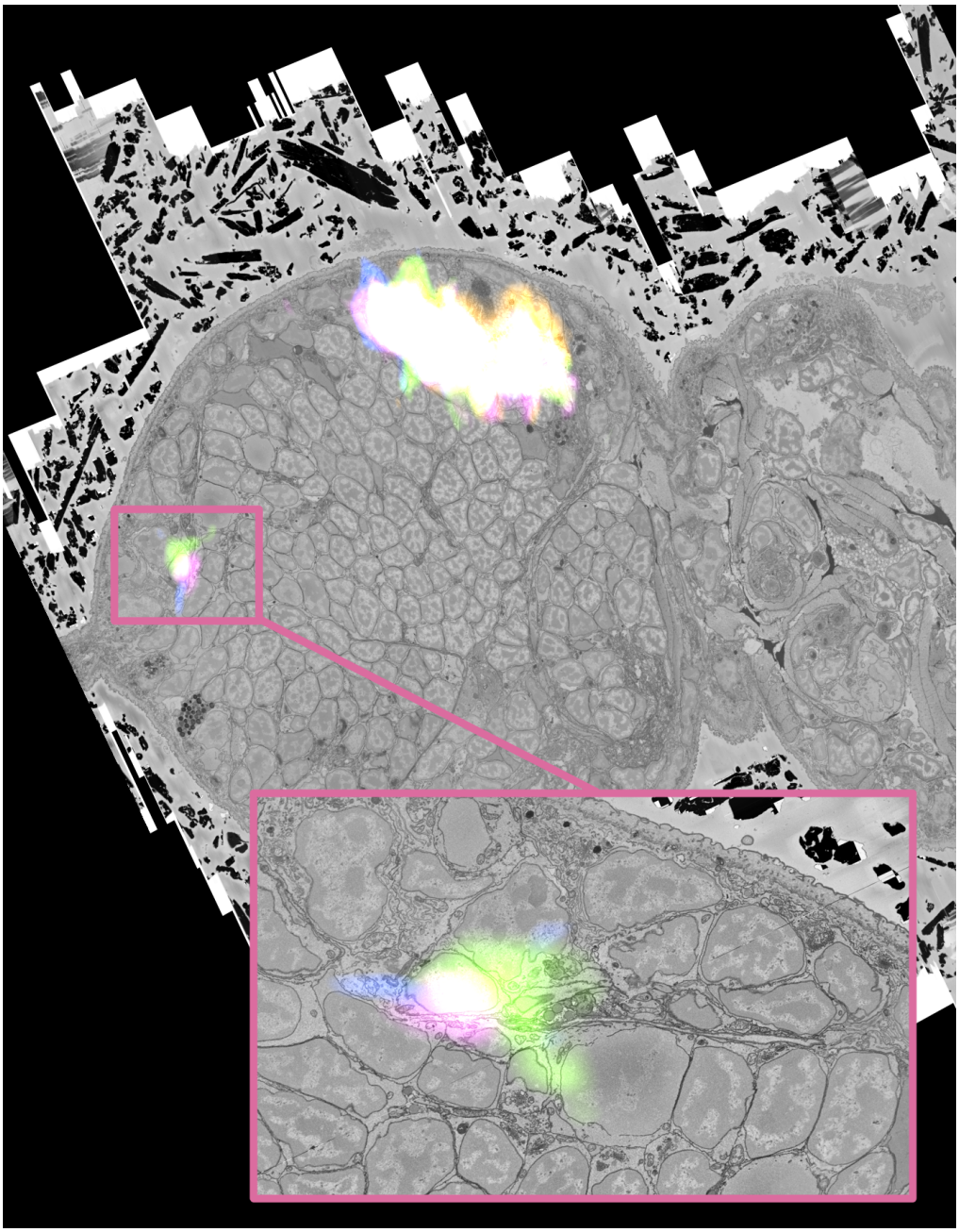


Figure 22: Registration quality control with Ropsin staining. EM data shown in grey. 4 replicates of ROpsin staining from registered confocal volumes shown in pink, yellow, green and blue. The screenshot shows YZ slice of the volume, meaning that the vertical axis is roughly aligned with the direction of optical slicing for confocal microscopy. The pink box highlights the ROpsin expression in a single photosensitive cell.

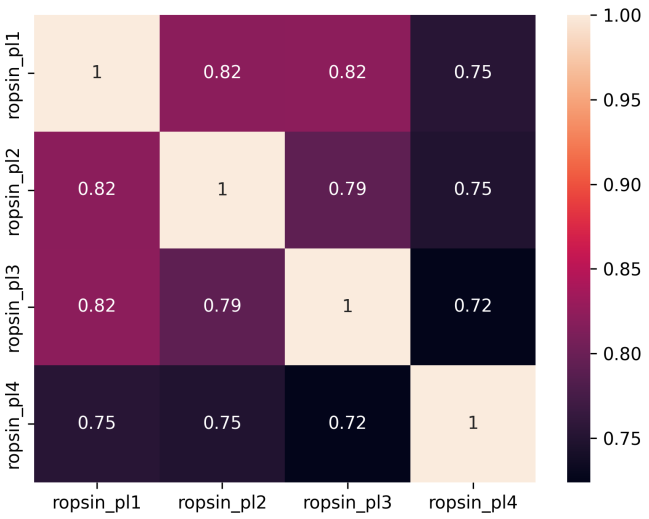


Figure 23: Correlation of the ROpsin signal mapped to EM from different replicates.

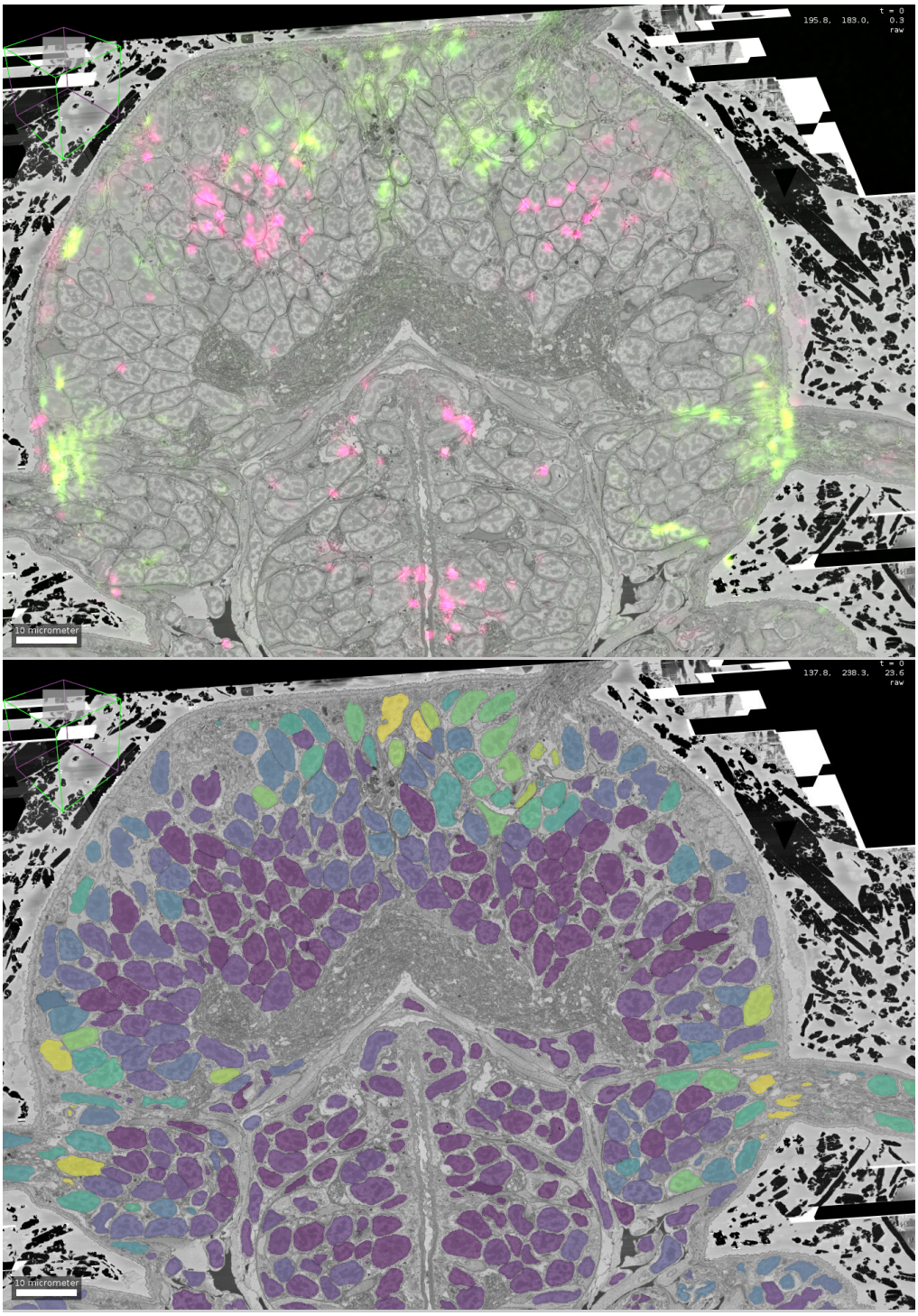


Figure 24: Expression of Pkd1 and Arx mapped to EM. Top: smFISH staining (4 replicates per gene) for Pkd1 (red) and Arx (green) overlaid on top of EM data (gray). Bottom: number of Pkd1 spots mapped to EM nuclei.

4.4 SUMMARY

In this project I developed an automated pipeline which allows to map smFISH confocal microscopy data to electron microscopy. The registration is based on segmenting cell nuclei in both modalities and using them as landmarks. The pipeline can be split into 3 parts.

In the preprocessing part the samples are handled individually: the signal is enhanced via merging confocal views, then the nuclei are automatically segmented and gene expression signal is quantified using a DL-based method.

The second part of the pipeline is the registration to EM template: each sample is registered to the EM using similarity transform, after that the instance segmentations are transformed into point clouds and deformable alignment of point clouds is performed using Coherent Point Drift. Coherent Point Drift places the points of the moving cloud as close to the points of the fixed point cloud while keeping the overall structure of the input point cloud intact. Physical distances between nuclei in EM and aligned nuclei from confocal are used as a cost matrix of the assignment problem, which yields correspondence between nuclei in light an EM.

The third part of the pipeline is the visualization and visual quality control. Corresponding nuclei are used as registered landmarks to warp the raw fluorescent images and overlay them with EM in MoBIE viewer.

Registration requires no manual input, works stably enough to require only 4-5 replicates per gene and matches single cells across modalities. To my knowledge, it is the first data analysis pipeline that allows single-cell level matching between cells of a full organism of this size. Multiple factors contributed to meeting these challenging requirements:

- *High-quality input data*

  I worked together with Luca to iterate on the protocol for microscopy which would produce images which are good enough for the task. To my knowledge, acquiring two views of the same sample is not a standard way of imaging 3D smFISH data. In this case it was worth the overhead of doubling the acquisition time and the amount of raw data to be process as having volumes with relatively uniform intensity distribution allowed to achieve very good segmentation quality. Reproducible and predictable data acquisition is key for any high-throughput data anlysis pipeline.

- *Custom segmentation*

  Quality of registration based on segmentation depends strongly on the segmentation quality. In my experiments, it was possible to register confocal volumes even with the imperfect Stardist segmentation. The segmentation model makes approximately the same errors in every sample and provides reliable registration landmarks. However, EM segmentation has completely different challenges and errors, which is why accurate registration and cell-to-cell matching required ground truth annotation and development of a segmentation method which combined Stardist and mutex watershed.

- *Gradual registration*

Registration progresses from a heuristic SVD-based pre-alignment to solving assignment problem for matching individual nuclei across modalities. Each registration step solves only a part of the problem, making the input more predictable and therefore allowing to set registration algorithm parameters that solve the problem efficiently and generalize to all samples. For example, without the similarity pre-alignment before CPD the deformable alignment would have to take into account not only the pose difference between different samples but also rotation and translation, which would lead to more errors.

- *Using point cloud registration instead of intensity-based registration*

  Intensity-based registration implemented in Elastix is routinely used for biological and medical image analysis. It is based on calculating cross-correlation or mutual information between two images and finding a transformation that would maximize the metric using gradient decent. As all gradient decent methods, this type of registration is prone to getting stuck in local minima, and images of nuclei provide plenty of local minima since correlation becomes very high if some of the objects match. Normally this is circumvented by smoothing the images before registration but for nuclei staining it means removing a lot of details and making precise registration very difficult. Instance segmentation provides meaningful landmarks and point cloud registration operates with nuclei as separate object, not just a set of unrelated pixels.

- *Optimization of registration parameters using gene signal*

  It is very hard to visually estimate the quality of 3D registration, especially if the goal is to reach near-single-cell accuracy. An orthogonal registration quality metric is an invaluable tool for assessing the registration quality without browsing large 3D volumes. Using the StMhc correlation metric introduced earlier I could run a grid search and choose optimal parameters of CPD.

- *CPD is remarkably well-suited for the alignment of nuclei centroid point clouds*

  Popular methods developed for registration of point clouds derived from 3D models of objects or measured in a remote sensing setting, such as PointNet [28] and SuperGlue [27], make use of the normals, which means that the point clouds need to represent the surfaces of objects. This is not true for nuclei as the nuclei are spread throughout the volume of the sample, making the concept of normals meaningless. CPD models point cloud as a sum of Gaussians, effectively utilizing density of points as the main registration feature.

4.5 LIMITATIONS

- Segmentation errors

  Segmentation errors limit the best possible accuracy of single-cell matching but with the resolution of confocal images in Z and cell type diversity of the sample it is very hard to achieve accurate segmentation. Manual

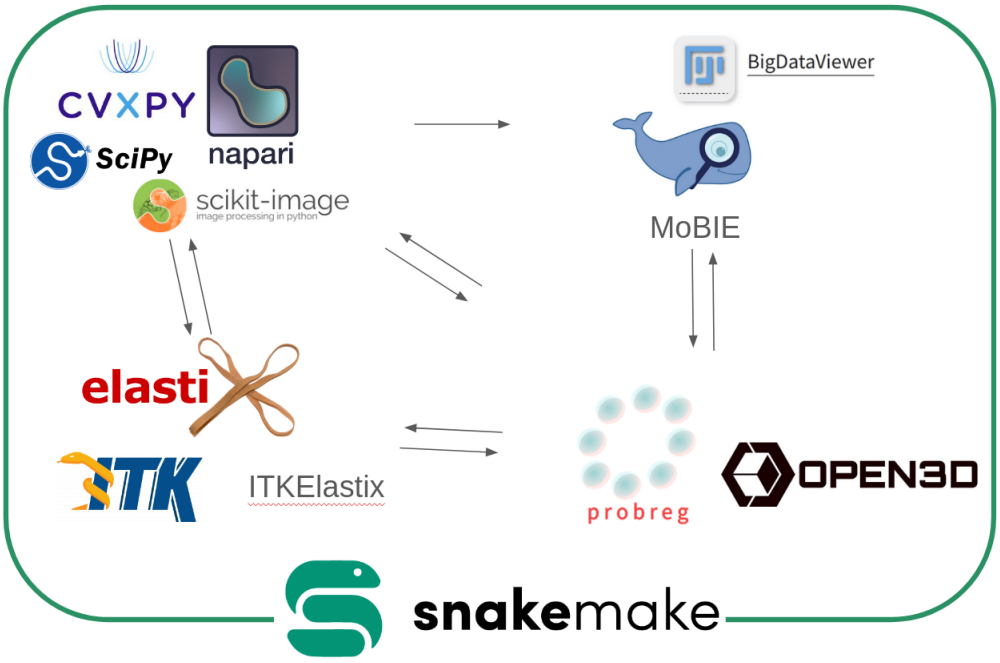


Figure 25: Pipeline tools

correction is also not feasible: adding 2 genes to the atlas means segmenting ~ 11000 nuclei in each of the 5 replicates. Manual correction for just 2 genes would mean curating ~ 55000 nuclei. In addition, EM segmentation used as a template is also not perfect but the errors have a different bias, introducing additional errors.

- How stereotypic is stereotypic?

  As shown for *C. elegans*, even stereotypic animals have some differences. It is not known how different individual larvae is for *Platynereis dumerilii* and without manual matching of all cells it is hard to distinguish between segmentation errors, registration errors and true difference between individuals.

- Low quality in the areas that do not have any landmarks

  Developed pipeline uses segmentation-based registration, which means that only the nuclei landmarks get registered. It means that the areas between nuclei such as neuropil or lipid droplets in the body of the animal do not have any landmarks, sometimes leading to unrealistic deformations. In the same way, it means the although centers of nuclei are matched, it is not possible to correctly map the physical location of the gene expression spots and distinguish, to which cell exactly the spot belongs based on the overlay with EM data.

- Very clunky pipeline (Figure 25)

  The pipeline was built using Snakemake [76]. Each step of the pipeline was a separate Python script which read the data, passed it to the appropriate tool and recorded the results. Due to the size of the volumetric data, reading and writing data took up significant part of the execution

time for several steps of the pipeline. Storing all the intermediate results also significantly increases the storage requirements. A more streamlined version of this pipeline would complete multiple steps simultaneously, without writing and reading intermediate results. In this case I decided to split the pipeline into small independent steps as it facilitated debugging and allowed to interact with the Slurm-based cluster queue more efficiently.

As shown in 25, multiple different ecosystems were used for building registration pipeline. General purpose Python libraries such as `Scipy` and `scikit-image` were used for basic operations with images such as preliminary alignment with SVD. Napari was used for checking and manually correcting segmentation and spot detection results. `CVXPY` was used for the combinatorial optimization. Intensity-based registration was done using C++ library `Elastix` via the Python bindings `ITKElastix`. GPU-based implementation of CPD was taken from `Probreg`, a library for point cloud registration written with `Open3D`, the C++ library for fast storage and operations with 3D point clouds. Finally, the cross-modal visualization was done in ImageJ and specifically MoBIE, the plugin based on BigDataViewer.

Unfortunately, each of these tools use different agreements about the order of axes, default resolutions and ways to write down registration transforms. Figuring out how to pass data between the libraries was a major challenge during pipeline development.

- Application to new datasets is limited by the preprocessing step

  Despite being the most studied problem in biological image analysis, testing methods for segmentation of confocal volumes and collecting ground truth took several months of work, despite numerous claims in the literature that nuclei segmentation is a solved problem. Since this pipeline is underpinned by the high quality instance nuclei segmentation, application of the same registration methods to new datasets is limited by the preprocessing step.

4.6 DISCUSSION AND OUTLOOK

Fully automated cross-modal registration of not exactly identical 3D samples with near-single-cell accuracy is a challenging task. It was possible to solve it by making use of the recent advances in the area of 3D nuclei segmentation to detect relevant landmarks and then implementing gradual registration procedure. Instead of trying to adjust registration parameters to each dataset individually, I took an approach in which subsequent steps compensate for the errors of the previous steps:

- Multiview registration
  - Loss of signal due to imaging depth
- Segmentation
  - Intensity variations

  - Lack of obvious matching landmarks between modalities
  - Resolution mismatch
- Coherent Point Drift registration
  - Differences between individual organisms
  - Segmentation errors
- Point matching with integer linear programming
  - Registration and segmentation errors
  - Different number of nuclei in individual samples

CPD-based registration can be reused for other biological data, including 3D and 2D samples (for example, consecutive tissue slices). It is modality-agnostic as long as instance segmentation is available for all images. It is not limited to segmentation of nuclei as long as it is possible to detect the landmark objects in all modalities. My experiments with Stardist segmentation have shown that as long as segmentation makes consistent errors (for example, always oversegments the same elongated nuclei in all samples), the developed pipeline can be applied and achieves good results for creating overlays of the images.

Practically application of the CPD-based registration to new datasets has two main challenges. Firstly, since the single-cell matching and overall registration quality depend on the segmentation accuracy, developments in the area of instance segmentation would improve the usability of segmentation-based registration. Secondly, more software development and standardization efforts are needed to integrate all the various tools (Figure 25) and make registration more user-friendly.

Datasets collected for registration quality control (Table 2) in combination with the EM volume of *Platynereis dumerilii* represent a unique benchmark, which will become a valuable resource for large-scale cross-modal biological image registration method development.

It was possible to register about 50 individual smFISH volumes, creating an atlas of 16 genes mapped to EM volume. Having nuclei-to-nuclei correspondence allows to easily transfer annotations, such as gene expression levels and morphological features, between modalities.

Part III

SPATIAL SINGLE-CELL ISOTOPE TRACING

Contributions

- I acquired MALDI and microscopy data for all the preliminary experiments (not presented here) and part of the datasets presented in this chapter, performed data processing using SpaceM and developed data analysis pipeline
- Måns Ekelöf acquired MALDI data for part of the datasets
- Mohammed Shahraz acquired microscopy data for part of the datasets
- Marteinn T. Snaebjornsson and Lisa Schlicker performed isotope tracing experiments and provided samples for analysis as well as control bulk mass spectrometry data
- Alberto Bailoni performed data processing using SpaceM for part of the datasets

5 SINGLE-CELL ISOTOPE TRACING OF HETEROGENEOUS CELL CULTURES

5.1 ISOTOPE TRACING

The state of a biological system cannot be fully described without taking into account the whole range of lipids and small metabolites which are the substrates and products of enzyme reactions. Mass spectrometry, a chemical analysis technique used to provide absolute or relative intracellular or extracellular metabolite levels, is a broad and sensitive method to detect differences in metabolic states between conditions. However, measuring changes in concentrations of individual metabolites does not always allow detailed conclusions about biological activity, as most metabolites are part of several important and interlinked metabolic pathways. Isotope tracing is a well established method in biochemistry used to estimate relative pathway activities, qualitative changes in pathway contributions via alternative metabolic routes and nutrient contribution to the production of different metabolites .

In isotope tracing experiments in cell culture, one of the nutrients in the growth medium is replaced with a nutrient labeled with non-radioactive heavy atoms. As mass spectrometry separates molecules based on mass, the incorporation of heavier atoms in a molecule provides an unambiguous label while keeping molecular properties close to identical to the naturally occurring analogues. The labeled nutrient is taken up by the cell and metabolized, producing various labeled metabolites. If the cells are in metabolic steady state, meaning that all intracellular and extracellular metabolic fluxes stay constant in the course of the experiment, the concentration of labeled metabolites changes over time and levels out at the isotopic steady state. The time required to reach this state depends on turnover rate for the pathways of interest. labeling experiments can be performed both in dynamic and steady state. Dynamic measurements, during which samples are analyzed in several time points, can be used to infer metabolic fluxes, while in steady state analysis of isotopologue patterns allows to draw conclusions on nutrient usage and relative pathway contributions [79].

In response to changing conditions cells rewire their metabolism, balancing between energetic demands and biosynthetic requirements. It is critical to understand metabolism on the single-cell level since, similar to gene expression, metabolic response of cells to environment, disease, or therapy can be highly heterogeneous. **In this project my goal was to adapt bulk isotope tracing to study nutrient usage and relative pathway contributions at single-cell level.**

5.2 SPATIAL SINGLE-CELL MASS SPECTROMETRY

A mass spectrometry measurement requires a sample of the material to analyze, which must be broken up into free molecular ions. One common tech-

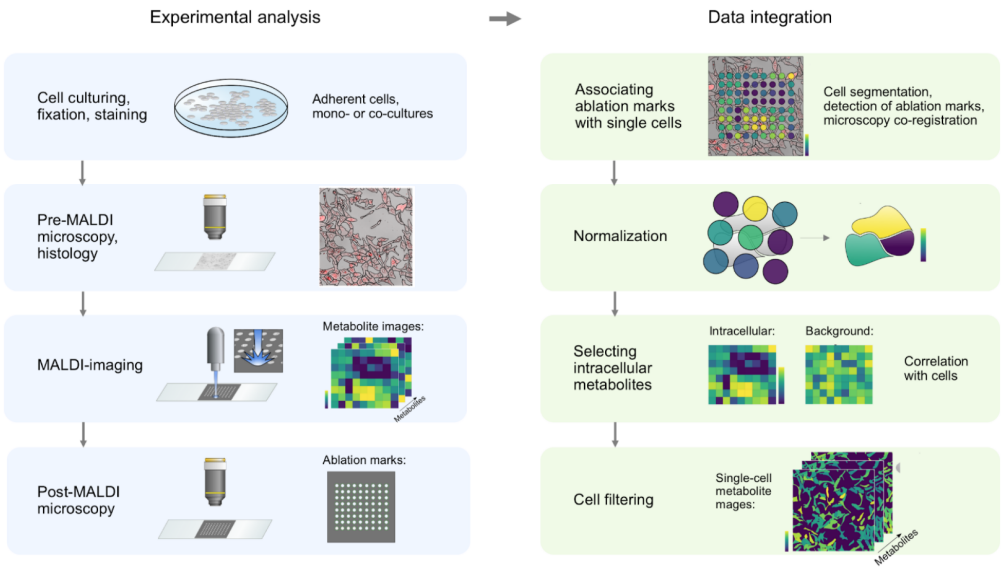


Figure 26: Protocol for single-cell metabolomics (reproduced from [77]). Cells are grown on a glass slide, fixed and stained, then three rounds of measurements are performed. Firstly, bright field and fluorescent pre-MALDI microscopy is done to collect data about location of cells. Then MALDI imaging is done on a part of the well with ablation marks forming a grid and uniformly sampling region of interest. After that post MALDI bright field microscopy is conducted to get the location of ablation marks. Ablation marks and cells are segmented from microscopy images and overlaid, therefore mass spectra are matched with the corresponding cells.

nique is to use a pulsed ultraviolet laser beam to eject material directly from a sample surface (Matrix-Assisted Laser Desorption Ionization, or MALDI) [80]. The resulting cloud of partially ionized material is injected into a mass spectrometer which provides a measure of mass and amount of the constituent molecules. By sampling a target surface in a suitable grid pattern and connecting each chemical measurement to a spatial location, MALDI mass spectrometry provides a multi-channel molecular image that can be correlated with other imaging modalities [81].

One application of mass spectrometry imaging is to identify differences between single cells in tissue sections or from culture. For the latter, a convenient preparation is to grow adherent cells directly on a microscope slide. To accurately assign the resulting metabolite images to cells, it is necessary to use high resolution microscopy for registration. The SpaceM [77] method for image registration illustrated in Figure 26 utilizes a multi-step pipeline, where optical microscopy is performed before and after the MALDI experiment, to compensate for the low resolution of MALDI imaging (>5 µm / pixel). The initial optical image is used to determine the location and optional fluorescent labelling of each cell. The post-MALDI image is used to determine the exact location of each laser shot through segmenting the visible marks in the image. In this way, the laser shots addressing each cell can be identified and grouped for further enrichment analysis.

The advantage of this method is that it allows to correlate mass spectrometry signal with microscopy, providing additional information about the cell

phenotype. However, due to small size of ablation marks, each ablation contains very small amount of material leading to low sensitivity of the method compared to bulk mass spectrometry, so only very abundant metabolites get detected. Moreover, the biological material is ablated directly from the slide and the ions are transferred to the mass analyzer, meaning that there is no chromatography step. This significantly limits the power of the method to resolve isomers - molecules with the same mass but different structure.

5.3 ANALYSIS OF ISOTOPOLOGUE DISTRIBUTIONS OF FATTY ACIDS INCORPORATED IN LIPIDS

When the cells use labeled carbon source, the labeled atoms get incorporated into various metabolites, therefore detection and interpretation of the isotopologue distribution of different molecules can provide valuable information about the cell's metabolism [82, 83]. Based on the limitations of the single-cell metabolomics and preliminary experiments, it was decided to use the isotopologue distributions of lipid-incorporated fatty acids as a readout for the glucose utilization.

The Figure 27, A shows the fatty acid synthesis pathway. Fatty acid synthase produces fatty acids by joining units of two-carbon cytosolic AcCoA to first form 16C palmitate, which can then be desaturated or further elongated to produce a variety of fatty acids used to build lipids. The primary source contributing to the cytoplasmic AcCoA is glucose, metabolized in the mitochondria to AcCoA and then transported to the cytoplasm. Other contributors include glutamine, acetate, some amino acids and fatty acids. When the conditions change, the cells can switch between different carbon sources, for example, cancer cells do it to adapt to the metabolic constraints of the tumour microenvironment, in particular tumour hypoxia and nutrient deprivation.

If one of the nutrients is labeled, then some proportion of the AcCoA pool is also labeled. Since fatty acids are made from stochastic condensation of labeled and unlabeled acetyl-CoA, acetyl group labeling can be inferred by binomial fitting of fatty acid labeling [84]. The free fatty acids are further incorporated in the lipids.

Although it is possible to detect free fatty acids, the absolute size of the pool is not as large as the total pool of the cell's lipids. Moreover, the samples are prone to contamination with fatty acids coming from the laboratory equipment, which can significantly change the measured proportion of labeled fatty acids [84]. At the same time, spectra of the lipids contain many overlapping isotopologue peaks, making it very hard to analyze the spectra.

To increase the signal intensity, avoid the influence of contamination and simplify spectra interpretation, the approach illustrated in Figure 27, B was used. First, ions in the negative mode in mass range 600-1000 m/z were isolated. This mass range includes most of the cell's lipids but not the free fatty acids. Then fragmentation was performed using higher-energy collisional dissociation (HCD) at a normalized collision energy of 25. Fragmentation at this energy causes lipids to break into the fatty acids and the heads, therefore fatty acids from all lipids can be analyzed together, increasing the signal. Finally, the scan range was set to 100-400 m/z at 140,000 nominal resolving power

with 500 ms ion injection time. This mass range contains the fatty acids. It is important to note that thanks to the isolation range free fatty acids are not detected with this method, therefore the measured isotopologue distributions faithfully represent the cell's carbon utilization.

5.4 FATTY ACID SPECTRA INTERPRETATION

In addition to the normalized isotopologue peak intensities, a binomial model for the fatty acid synthesis described in [84] was used. Fatty acids are synthesized by randomly taking 2-carbon monomers from the cytosolic acetyl-CoA pool. The resulting single-cell isotopologue distribution for every fatty acid was normalized by its sum again. Scanpy v.1.8.1 [60] was used for all single-cell data analysis.

In addition to the normalized isotopologue peak intensities, a binomial model for the fatty acid synthesis described in [84] was used. Fatty acids are synthesized by randomly taking 2-carbon monomers from the cytosolic AcCoA pool. Therefore, if the fraction of labeled monomers equals p, the probability that a given fatty acid has $2i$ labeled carbons, can be described with the following model:

$$P_{binom}(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, ...n$$

$$I_0 = uptake + (1 - uptake) \times P_{binom}(0)$$

$$I_i = (1 - uptake) \times P_{binom}(i), i = 1, 2, ..., n$$

where uptake is the fraction of the unlabeled FA directly taken from the media, p is the labeling degree of cytosolic AcCoA pool, and n is the number of AcCoA molecules used for the synthesis (number of carbons in the fatty acid / 2). Longer fatty acids can be synthesized both using labeled palmitate made by the cell de novo and unlabeled palmitate, therefore we used a modified model for C18:

$$I_0 = uptake + (1 - uptake) \times P_{binom}(0)$$

$$I_1 = (1 - uptake) \times (uptake_{C16} + (1 - uptake_{C16}) \times P_{binom}(1))$$

$$I_i = (1 - uptake) \times (1 - uptake_{C16}) \times P_{binom}(i), i = 2, 3, ..., n$$

where $uptake_{C16}$ is the fraction of palmitate taken from the medium.

Fitting a binomial model to the single-cell isotopologue distributions allows summarizing them as two numbers: uptake, which characterizes the fraction of the fatty acid which was not synthesized de novo but directly taken from the medium, and acetyl-CoA pool labeling degree p, which describes relative contribution of the labeled substrate to the fatty acid synthesis compared to other carbon substrates consumed by the cell.

To produce a control experiment for glucose labeling, batches of cells were grown in labeled medium under normal or hypoxic (0.5% O_2) conditions. In the hypoxic state, cell respiration is inhibited, preventing lipid synthesis from glucose in favor of other carbon sources. After culture, cells from both conditions were prepared separately as well as in mixed sample where the hypoxic cells were modified to express GFP for identification (28, A). The cells grown under hypoxia form a robust negative control sample, where almost no label incorporation is observed, consistent with the bulk measurements (28, E). Labeling degree of fatty acids allowed to distinguish between hypoxic and normoxic cells with near-perfect accuracy (28, C).

5.5 ACLY KNOCKDOWN EXPERIMENTS

Another set of liver cells were modified with two shRNA oligonucleotide sequences to knock down the enzyme ACLY, which is responsible for the transfer of glucose-derived carbon from the TCA cycle through citric acid. The selective inhibition of this pathway was expected to produce a reduction in fatty acid labelling similar to, but lower in magnitude than the hypoxia control. In agreement with previous studies using bulk analysis of palmitate isotopologue distribution in ACLY wildtype vs knockout mouse embryo fibroblasts, silencing ACLY caused a reduction in the fractional labelling of acetyl-CoA from glucose but only a minor increase in the M+0 fraction. This suggests that other substrates, such as acetate, are used to synthesize acetyl-CoA under these conditions. The effect was quantifiable for both knockdowns treatments (29, A, C), and in one case isotope tracing revealed two distinct sub-populations of cells, with a total distribution in agreement with bulk data (29, D, E).

5.6 APPLICATION TO TISSUE SAMPLES

The isotope tracing strategy is flexible, and can be scaled well up to the whole organism level, limited mainly by the high cost of the labelled tracer. MALDI experiments with whole mice fed with a diet enriched with heavy labelled glucose revealed differences in glucose transportation using the label incorporation in downstream lipids as the proxy measure. Applied to a mouse glioblastoma model, it was possible to visualize spatial differences in lipid synthesis, elongation, and desaturation pathways [78].

5.7 DISCUSSION

13C-SpaceM is a method to combine stable-isotope tracing with spatial single-cell metabolomics to determine de novo fatty acid biosynthesis in cancer cells at single-cell resolution. 13C-SpaceM builds upon SpaceM [77] with the major changes including the mass spectrometry method constructed to mimic chemical saponification and fatty acid detection as in bulk stable isotope tracing as well as the data analysis. In particular, it was shown how using the binomial model applied to the isotopologue distribution helps estimate such labeling parameters as uptake and pool labeling degree for individual cells. 13C-SpaceM

as employed in this study uses all-ion fragmentation to release fatty acids from cellular lipids at m/z range of 600-1000 and ionized in the negative mode.

The ability of 13C-SpaceM to detect changes in glucose-dependent fatty acid synthesis, and its single-cell resolution, was validated by co-plating cells that had previously been cultured under either normoxia or hypoxia. It was furher shown that 13C-SpaceM can monitor the effect of perturbations of the metabolic network in cancer cells. Both bulk and single-cell analysis revealed a marked shift in palmitate isotopologue distribution following silencing of ACLY, correctly reflecting a reduced contribution of 13C-glucose to the lipogenic acetyl-CoA pool. Using a binomial model to quantify fractional labelling from the isotopologue distribution, we estimated the labelling degree of acetyl-CoA in control and ACLY silenced cells at single-cell level. In addition to correctly reproducing bulk results on average, 13C-SpaceM was able to reveal a substantial degree of heterogeneity in fractional labelling of lipogenic acetyl-CoA within the cell population.

MALDI-based single-cell metabolomics provides integration of mass spectrometry data with microscopy, at a cost of lower sensitivity and inability to distinguish between isomers or use standard for quantitative analysis. 13C-SpaceM bypasses these limitations by using all-ion fragmentation: pooling together fatty acids from all cell lipids increases the signal-to-noise ratio and makes the mass spectra clean and easy to interpret. As the biological meaning of the isotopologue distribution is determined by the ratios between isotopologues of the same molecule instead of absolute detected amounts of chemically different molecules, 13CSpaceM enables robust comparison of carbon source utilization between single cells.

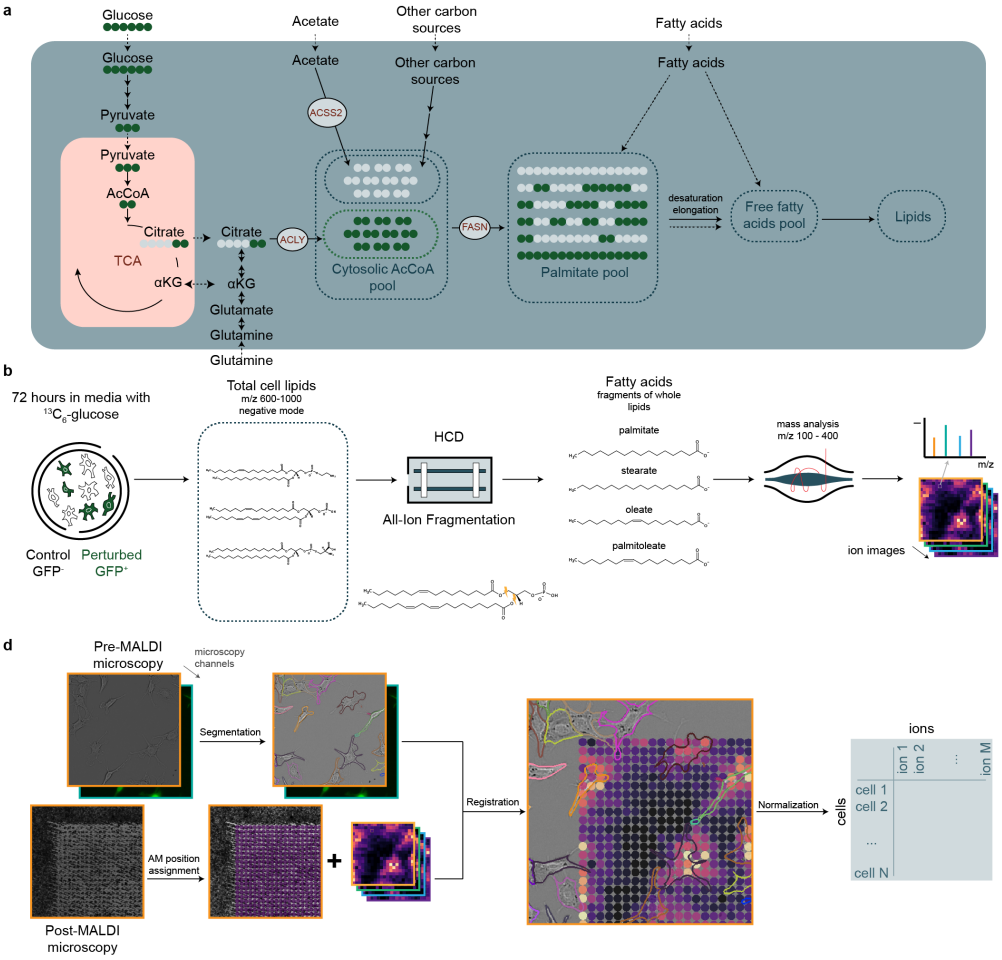


Figure 27: 13C-SpaceM workflow as applied to interrogation of de novo fatty acid synthesis (reproduced from [78]). a) Generation of cytoplasmic acetyl-CoA and subsequent fatty acid and lipid synthesis from stable-isotope labelled 13C6-glucose. Glucose-derived pyruvate is metabolized in the mitochondria and exported to the cytoplasm in the form of citrate. ATP-citrate lyase (ACLY) converts citrate to acetyl-CoA, thereby contributing to the cytoplasmic (lipogenic) acetyl-CoA pool. De novo fatty acid synthesis by fatty acid synthase (FASN) results in the formation of palmitate (C16:0), which is either further modified or directly incorporated into cellular lipids. Exogenously taken up fatty acids are also incorporated into cellular lipids but are not labelled. b) Using all-ion fragmentation imaging mass spectrometry for 13C-SpaceM. Cells are grown in medium supplemented with uniformly labelled 13C6-glucose for 72 hours. The combination of wide-range isolation of parent lipid ions followed by HCD fragmentation and selective isolation of fatty acid fragments allows high-sensitivity measurements of fatty acids incorporated into lipids, conceptually similar to bulk extractions followed by saponification but with the benefit of retained spatial information and the exclusion of free fatty acids. c) Integration of microscopy and imaging mass spectrometry in 13C-SpaceM to obtain single-cell profiles. Pre-MALDI microscopy and post-MALDI microscopy images containing information about the cell outlines and areas ablated by MALDI-imaging are registered. Ion intensities from MALDI-imaging are assigned to single cells through a normalization procedure.

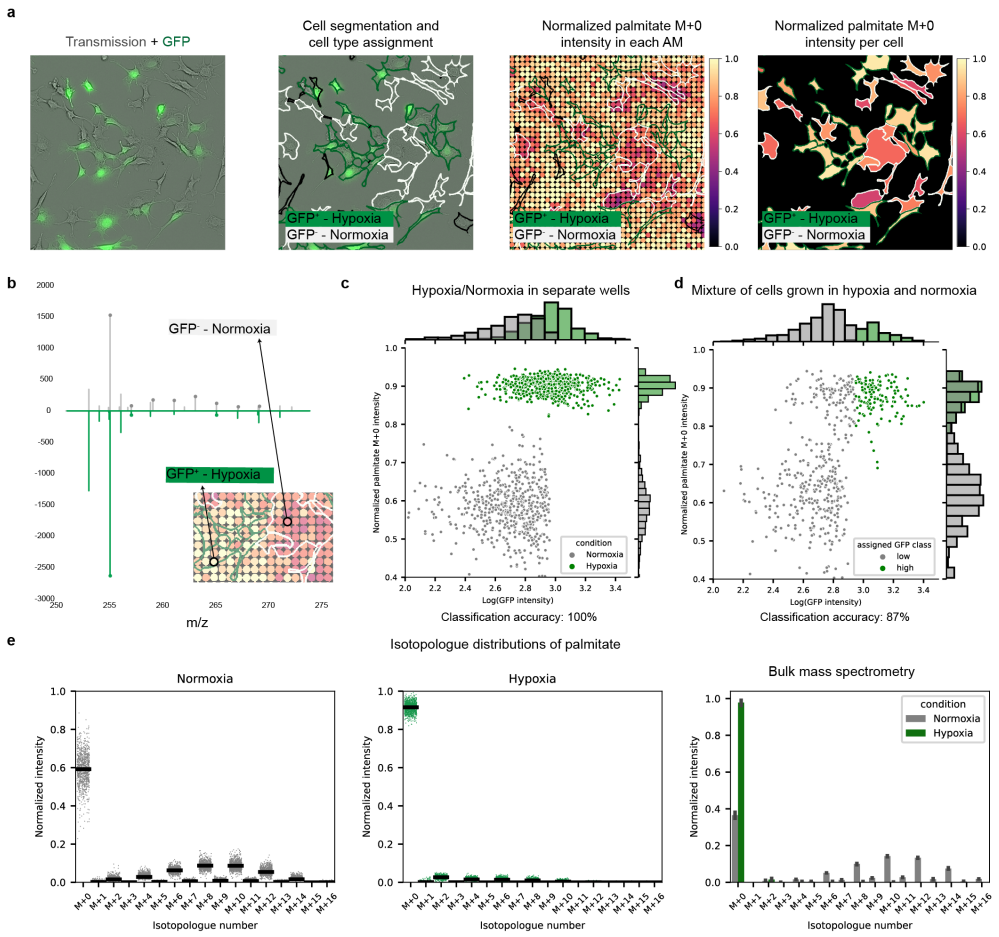


Figure 28: Validating 13C-SpaceM by interrogating de novo fatty acid synthesis in spatially-heterogeneous normoxia-hypoxia model (reproduced from [78]). a) Illustration of microscopy and imaging mass spectrometry data from the model of co-plated primary murine liver cancer cells originally cultured under normoxia (GFPneg) and hypoxia (GFPpos). The GFP signal was used for discerning the culturing conditions (white and green cell outlines shown normoxic and hypoxic cells, respectively). The normalized intensities of the M+0 isotope of palmitate (representing the fraction of unlabelled palmitate) are shown for MALDI-imaging pixels and as assigned to the single cells. b) Mass spectra for individual pixels mapped to cells from a normoxic cell (white outline) and a hypoxic cell (green outline). The peaks corresponding to palmitate isotopologues are marked by grey or green dots. c) Discerning cells cultured under normoxia vs. hypoxia using 13C-SpaceM for the cells mono-plated for each culturing condition. Scatterplot and histograms show the values of the GFP-reporter (ground truth for telling the condition) and the normalized intensity of the M+0 isotope of palmitate representing the fraction of unlabeled palmitate for single cells. Different colors are used to show cells with different true state label, known from the growth conditions for the particular well, and different cell state assigned using GFP signal. Prediction accuracy refers to the prediction of the true state based on the isotopologue distributions. d) Same analysis as in (c) for spatially-heterogeneous co-plated cells from both conditions. In this case accuracy refers to prediction of the cell state assigned using GFP signal. e) Confusion matrices for the prediction of culture condition based on 13C-SpaceM or GFP in cells grown separately and comparison of the two predictions for co-plated cells. f) Comparison of single-cell vs. bulk intensities for the M+0 isotope of palmitate. Single-cell intensities are from the spatially-heterogeneous co-plated model. Bulk intensities are from cells of each mono-cultured condition subjected to total fatty acid analysis by saponification followed by LC-MS. For single-cell data, black lines show average values. For bulk data, data are displayed as mean ± standard deviation across 3 replicates.

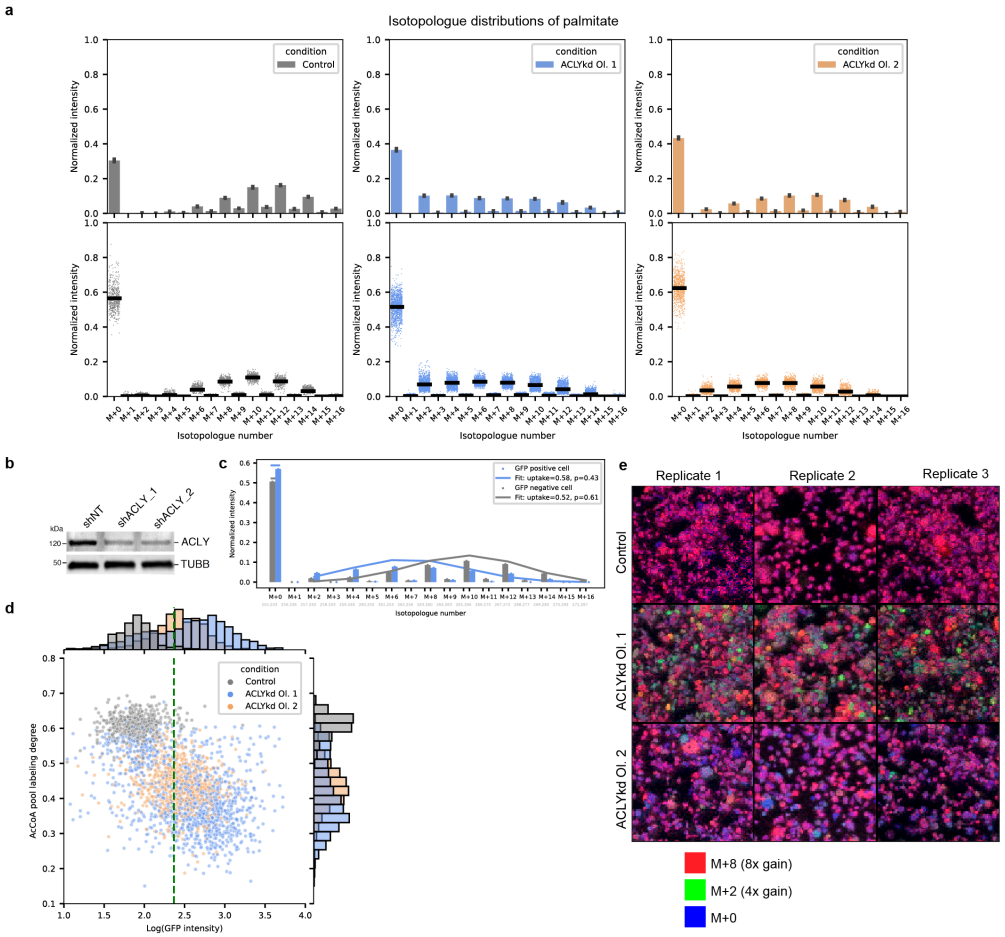


Figure 29: Single-cell quantitative analysis of lipogenic acetyl-CoA production and heterogeneity (reproduced from [78] a) Comparison of bulk and single-cell analysis of isotopologue distribution for palmitate after 72 hours of ACLY gene silencing and culture in the presence of 13C6-glucose. Bulk data are displayed as mean ± standard deviation across 3 replicates. For single-cell data, black lines show average values. b) Normalized single-cell isotopologue distributions for two individual cells, one from the control and the other from ACLYkd oligo 1 (shown as bar plots). Lines show fit of the fatty acid labelling binomial model: horizontal lines for M+0 showing an estimated uptake, and connected lines for M+2i showing (1 - uptake) * binomial(i). Legend shows parameters of the binomial model fit. c) Single-cell analysis of the estimated acetyl-CoA pool labelling degree (p) as calculated using the fatty acid labelling binominal model for the three conditions: control (grey), ACLYkd oligo 1 (blue), and ACLYkd oligo 2 (blue). Green dashed line shows 95% quantile of the GFP intensity distribution in the control condition, which was used to classify cells as GFPpos vs. GFPneg. d) Spatial metabolic imaging of de novo fatty synthesis for the control, ACLYkd oligo 1, and ACLYkd oligo 2 conditions. Abundance of different palmitate isotopologue peaks are displayed in different channels: M+0 (blue), M+2 (green) and M+8 (red). Each channel is normalized to the total ion count (TIC).

Part IV

CHALLENGES OF FEATURE NORMALIZATION IN SEGMENTATION OF LARGE BIOLOGICAL IMAGES

Contributions

- I developed the train/eval disparity and tile mismatch metrics, designed and implemented all presented experiments
- Edoardo D'Imprima acquired organoids datasets and provided ground truth for semantic boundary segmentation

6

CHALLENGES OF FEATURE NORMALIZATION IN BIOIMAGE SEGMENTATION

6.1 INTRODUCTION

Segmentation is an essential first step in automated quantification for various types of microscopy images. As shown in the previous chapters, in particular, segmentation is an important step in automated registration and multimodal integration pipelines as it helps to find meaningful landmarks and identify separate objects. It is vital for finding matching features in different modalities based on semantics and not visual similarity.

Recent advances in natural image analysis accelerated the development of biological image segmentation. However, there are differences between biological and natural images that preclude direct transfer of the methods and tools between these domains. Collecting a large dataset of annotated images and then training a generalist model is a popular approach in computer vision which led to the fast development of deep learning-based methods, for example, ImageNet [86], CityScapes [87] and Segment Anything dataset [56] significantly aided the advances in computer vision. Homogeneity in the image formation model and range of the depicted objects in a vast amount of imagery aquired with regular cameras and smartphones allows to reach the necessary level of generalization.

In contrast to that, biological images are much more diverse. Investigation of biological diversity, discovery of unexpected phenotypes and development of novel imaging protocols leads to high variability of datasets. This explains the popularity of human-in-the-loop approaches to segmentation. In this paradigm, a small amount of data is either annotated from scratch or from the output of pre-trained models (Cellpose [52], Stardist [53], uSAM [88]), then new model is either trained from scratch or fine-tuned. After that the next rounds of correction and training happen, until the desired quality is reached. In this setup instead of aiming to train the most general large model using a predetermined dataset, the goal is to quickly fit a smaller model to the specific dataset so it is important to be able to quickly train a smaller model for an unseen image type.

In addition to the higher diversity, the size of biological images is often much larger. For example, a single electron microscopy volume for connectomics [89] or light sheet [90] volume can reach multiple terabyte scale. Researchers routinely acquire images of approximately $1000 \times 1000 \times 1000$ pixels which take up 10-100 GB, exceeding the GPU memory. Nevertheless, these volumetric images represent one sample and need to be analyzed as a whole. Contrary to that, natural image datasets (ImageNet [86], CityScapes [87] and Segment Anything dataset [56]) usually contain relatively small 2D images which fully fit into GPU memory or can be resized without the loss of the signal of interest, allowing to use relatively large batch size.

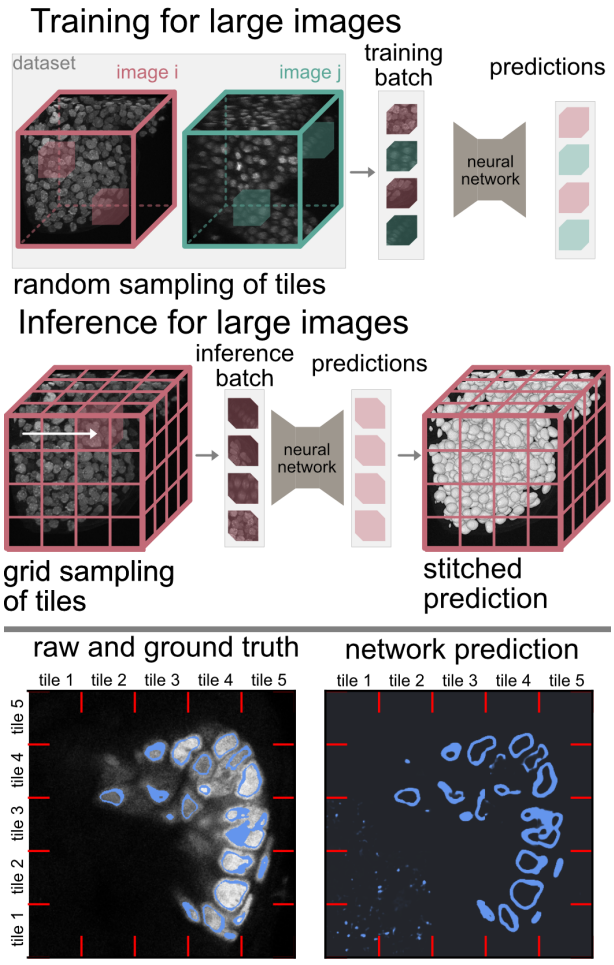


Figure 30: Illustration of the pipeline for processing images larger than GPU memory (reproduced from [85]): random sampling of tiles during training and sliding window inference. Example of artifacts caused by tiling: hallucinations in the background tiles and discontinuous predictions due to mismatch between neighboring tiles.

Larger than GPU memory image size prompted usage of the sliding window inference shown in Figure 30. The main idea of this approach is splitting the large volume into parts that fit into memory. Firstly, during training random patches are sampled from the annotated images and combined into a batch. Then, during evaluation the image is split into a grid of tiles which are processed with a neural network independently. After that the tile-wise predictions are stitched back to form the full prediction of arbitrary size.

For simplicity, both 2D and 3D patches are referred to as tiles. The size of the tiles is chosen to be in between the the receptive field size of the network [91] and the maximum size determined by GPU memory. Typically used tile size is around 96x96x96 - 256x256x256 pixels. Given the memory of the widely used GPUs, the batch size during training can go down to as little as 2-3 tiles per batch. During inference it is not necessary to store gradients, therefore either tile size or batch size can be larger.

Tile-wise prediction and stitching might are not just a technical step, as the patch sampling and stitching can affect prediction quality and create stitching

artifacts (Figure 30). This was shown not only in biological image analysis [92] but also in medical [93, 94] and aerial [95] image analysis domains which also operate with images larger than GPU memory.

According to the literature, it is possible to achieve seamless stitching with CNNs for prediction if edge effects are taken into account [96].

Let us consider the common case in which the size of the input image and output prediction coincides, such as for semantic segmentation. For CNNs only a limited area of the input image called receptive field (RF) affects the prediction value in a given pixel [97, 98]. For the output pixels which are closer to the edge of the tile than $\frac{1}{2}$RF part of the input values within the RF is replaced by zeros, changing the predicted values and causing mismatch between tiles. A common solution to this issue is sampling tiles with overlap and using only the central part of each tile where all pixels have the full context for stitching. In this setup the cropped area is called halo, and its minimal sufficient size is $\frac{1}{2}$RF.

According to my results, even with large halo sliding window inference can cause tiling artifacts. There are two types of the artifacts: abrupt discrepancies between predictions in neighboring tiles and background hallucinations, as shown in panel C of 30 and third column of 31.

In practice many pipelines using sliding window inference do not aim to completely eliminate the tiling artifacts but rather minimize them by using the largest possible tile size in inference and averaging. For example, nnU-Net [99] (a popular method for medical segmentation) and Cellpose [52] (a popular tool for cell segmentation in microscopy images) compute weighted average of the tiles, with the weight decreasing from 1 in the middle of the tile to 0 at the border, smoothing the border between tiles and making artifacts less visible. Additional thresholding based on intensity and filtering by the object size can be used to remove artifacts in the background like in the Figure 30, panel C. Sometimes the artifacts become less visible when the semantic predictions are converted into instance segmentation.

Although heuristic postprocessing helps to minimize the tiling artifacts, additional steps make the whole pipeline harder to use and apply to new datasets. In my experience tiling artifacts become more pronounced in the transfer setting, further reducing usability of the networks. According to my observation, even after taking into account the edge effects, predictions inside the valid central area do not necessarily match between tiles. It has been shown that for the otherwise fixed model and training procedure the prediction accuracy can depend significantly on the tile size [100, 101]. According to my experiments, the main cause of these effects in the feature normalization within the network.

In this project I aimed to find the source of the tiling artifacts in the sliding window inference pipelines and explore the connection between tile size, batch size, normalization strategy and prediction quality in large biological image segmentation.

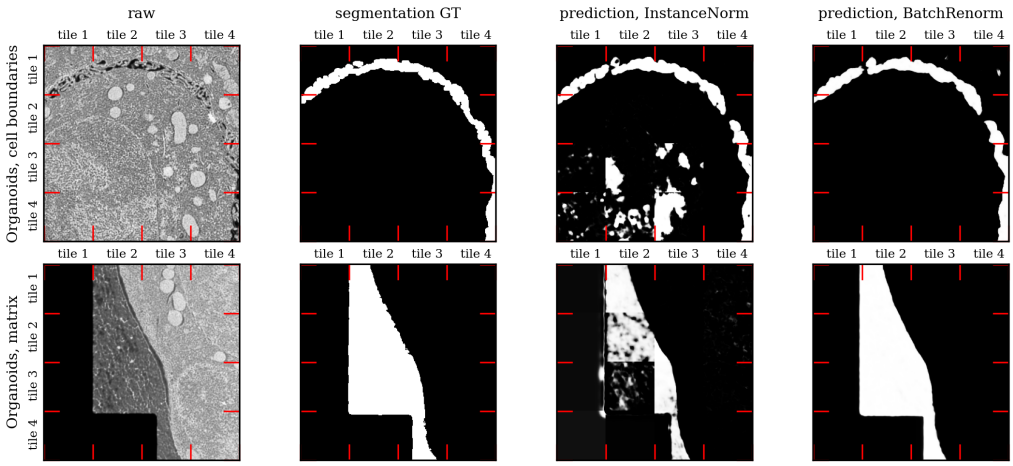


Figure 31: Examples of predictions with and without tiling artifacts (reproduced from [85]).

6.2 METHODS

6.2.1 *Sliding window inference with halo*

Input data can be a 2D image or a 3D volume with one or multiple channels. Here both images and volumes are referred to as images. Input image I is a matrix of size $C_{in} \times Z \times Y \times X$, where C_{in} is a number of channels, Z, Y, X are image sizes in pixels. For inference the image is split into a grid of N tiles of size $C_{in} \times Z_t \times Y_t \times X_t$. Tiles which do not fully fit in the image are padded with zeros. Corresponding stitched prediction P is a matrix of size $C_{seg} \times Z \times Y \times X$ formed by tile-wise predictions of size $C_{seg} \times Z_t \times Y_t \times X_t$.

To ensure that observed tiling artifacts were not caused by edge effects, theoretical (TRF) and effective receptive field (ERF) of the networks was calculated following [98]. TRF is defined as all pixels which belong to the computation graph for the target pixel in the output space. ERF was estimated using the following procedure:

1. Tiles were sampled from validation dataset and processed with a network
2. For each tile the gradients of the prediction at the center of the output tile with respect to each pixel of the input are calculated using Pytorch's automatic differentiation
3. The absolute values of the gradients for each tile are normalized to the range between 0 and 1 and averaged over all tiles.
4. Resulting distribution was modeled as a Gaussian distribution and the width of the peak (3σ) was reported as a size of effective receptive field.

Overlap equal to the size of halo was used to minimize the computational overhead. The halo part of the tile was completely removed before stitching.

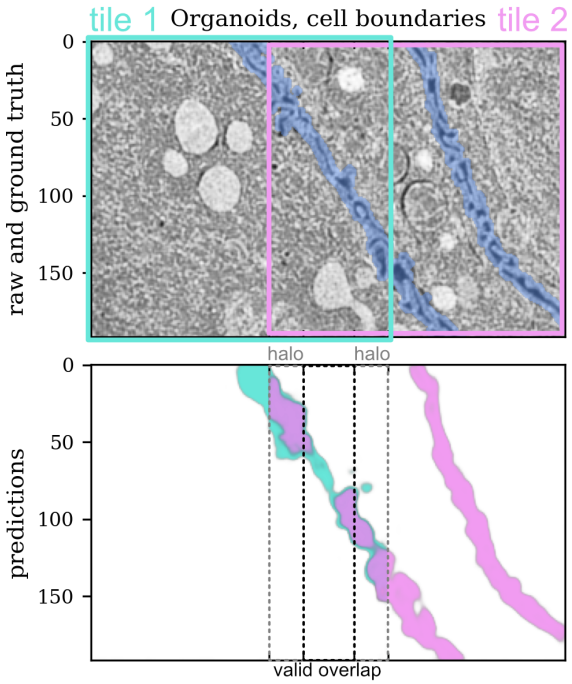


Figure 32: Explanation of tile mismatch metric (reproduced from [85]).

6.2.2 *Evaluation metrics*

Overall prediction quality Sliding window inference was used to produce stitched predictions for the whole volumes. In the cases when tile-wise input normalization was tested, each tile was normalized separately before being passed through the network. Dice score is a metric which is often used to characterize the quality of semantic segmentation. Dice score for prediction P and ground truth labels L of shape $C \times Z \times Y \times X$:

$$\text{Dice}(P, L) = \frac{2 \cdot \sum_{z,y,x} PL}{\sum P^2 + \sum L^2},$$

where sum was taken over the volume separately for each channel. Different classes can be more difficult to predict, therefore median Dice score was reported for each predicted class.

Train/eval disparity According to my experiments, difference between normalization in train and eval mode can cause unexpected change in the network prediction. To investigate this difference in addition to "dice, eval mode" in 5 I reported 6 "dice, train mode". This metric is calculated similarly to "dice, eval mode" except the model was set to the train mode (`model.train()`). For a more direct comparison and quantification I introduce train/eval disparity metric:

$$\text{train/eval disparity} = 1 - \text{Dice}(P_{train}, P_{eval}),$$

where P_{train} is the prediction done with `model.train()` and P_{eval} - with `model.eval()` for the boundary class.

In this case 0 means that the predictions perfectly match and higher value means larger difference between predictions. The inference in both modes was done on the same images from the validation set, so the difference comes only

from the behavior of the model in different modes and not overfitting to the training data.

Tile mismatch

Quantification of tiling artifacts is tricky because the magnitude of the artifacts depends on the similarity of the content of neighboring tiles in different parts of the image. The tiles of size $192 \times 192 \times 288$ were sampled in a grid with a stride of 64 pixels, allowing to test the artifacts in the whole volume. Sampled tiles were split into two overlapping tiles of size $192 \times 192 \times 192$. After that both tiles were processed with the network to compare the predictions in the overlapping region. Only the valid part of the overlap was taken for comparison, as shown in 32. The following metrics were calculated:

$$max\_dist = \max_{i=1}^{M}(|O_{i1} - O_{i2}|)$$

$$tile\ mismatch = median(1 - Dice(O_{i1}, O_{i2})),$$

where O_{i1} and O_{i2} are predictions in the valid overlap region in tile i and M is number of sampled tiles. In $max\ dist$ max is taken over all channels and all tiles, making it an indicator of wether given setup produces tiling artifacts. In cases when $max\ dist = 0$ I explicitly report "no" instead of $tile\ mismatch$ to emphasize that the predictions perfectly match. $tile\ mismatch$ was calculated per channel and characterizes the magnitude of the artifacts. Values for boundary channel were reported in 5, 6 and 7.

6.2.3 *Normalization strategies*

With no postprocessing of the predictions, there are two parts of the pipeline when normalization occurs: preprocessing of the input data and feature map normalization inside the network.

6.2.3.1 *Input normalization*

Microscopy data often has oversaturated regions or single pixels so quantile normalization is a good general method to normalize images without letting a few outlier pixel values dominate the normalization. Firstly, quantile normalization was applied to each volume as a whole to convert the data type and value range from the microscope output to $float32$ and $[0..1]$ range.

$$quantile\_norm(I, q_{min}, q_{max}) = \frac{I - q_{min}}{q_{max}},$$

where q_{min} is the lower quantile and q_{max} is higher quantile. $q_{min} = 0.01$ and $q_{max} = 0.98$ were used for all datasets. Then the values are clipped to $[0..1]$ range to avoid having negative values.

A common input normalization strategy is applying quantile normalization to each tile separately. Supposedly, it makes tiles more comparable and ensures that the input of the network always has the same range.

The following normalization strategies were considered:

- *Global*: apply quantile normalization only to the whole volume.

- *Tile-wise*: first apply quantile normalization to the whole volume, then to each sample tile, both during training and during inference.

6.2.3.2 *Feature normalization*

General formula for the normalization operation with input x, output y is:

$$y = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}},$$

where μ and σ are normalization parameters and ϵ is a small constant used for numerical stability. Parameters μ and σ can be estimated directly from the input $\mu = E[x]$ and $\sigma^2 = \mathrm{Var}[x]$, where average can be taken either over each sample independently or over the whole batch. Alternatively, global normalization parameters independent of the current input can be used. A common strategy is to estimate the parameters as a running average over multiple samples: $p_{new} = (1 - momentum) \times p_{old} + momentum \times p_t$. The update speed is determined by the $momentum$.

- `BatchNorm`

  Training: use statistics of the current batch and collect running average. Inference: use saved running average.

- `InstanceNorm`

  Both training and inference: use statistics of each input sample.

  In Pytorch this behavior can be changed by setting `track_running_stats=True`. Then during inference the statistics collected during training will be utilized, same as in `BatchNorm`.

- `BatchRenorm` [102]

  Both training and inference: use running average statistics.

  It was observed that using running statistics during training leads to the weights exploding. To overcome this issue but still use the same statistics both during training and during inference it was suggested to consider μ and σ as external parameters which do not require gradients.

$$\frac{x - \mu}{\sigma} = \frac{x - E[x]}{\sqrt{\mathrm{Var}[x]}} \cdot r + d,$$

  where $r = \frac{\mathrm{Var}[x]}{\sigma}$ and $d = \frac{E[x] - \mu}{\sigma}$. For a given batch r and d are treated as external parameters and are not used during backpropagation. This way `BatchRenorm` ensures that that all layers are trained on internal representations that will be actually used during inference.

  Training with `BatchRenorm` consists of 3 phases:

  1. $r = 1$, $d = 0$: the same training as `BatchNorm`. This part needs to be long enough for the model to almost converge.
  2. r and d are slowly increased to their final values.
  3. r and d are calculated based on the running average statistics and current input statistics.

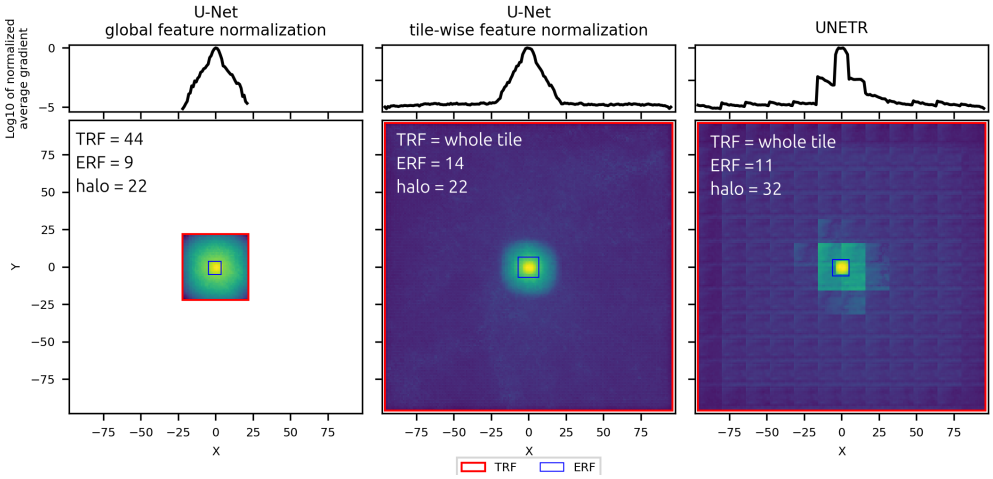


Figure 33: Log10 of mean gradient of the central pixel output with respect to input (reproduced from [85])

- `Identity`

  Skip feature normalization.

  Normalization layers were introduced to reduce sensitivity to initialization and allow for higher learning rates. However, it is possible to train a model without normalization layers. In this setup normalization layers were replaced with `Identity` layers.

6.3 EXPERIMENTS

6.3.1 *Datasets*

Datasets prepared for competitions are often cropped to contain only the object-centered parts of the images and do not always reflect the variability observed within one image in practice.

The following publicly available datasets were taken from biological publications to represent the challenges of biological image analysis in terms of the data variability within one dataset.

organoids [103]: instance segmentation of cells in patient-derived colorectal cancer organoids, electron microscopy (FIB/SEM), $40 \times 30 \times 30$ nm^3 voxel size. The volume of size $1350 \times 1506 \times 1647$ pixels was split 70/30 into train and validation dataset with the first 945 slices along Z axis being in train and the other 405 slices - in validation. Ground truth instance masks were converted into cell boundaries, foreground - pixels inside the cells - and extracellular matrix - space outside of the cells but inside the measured area. The task was to predict 4 channels: boundaries, foreground, extracellular matrix and background.

plants [59]: instance segmentation of nuclei in Arabidopsis ovules, confocal fluorescent microscopy $0.28 \times 0.13 \times 0.13$ µm^3 per pixel resampled to isotropic resolution of $0.13 \times 0.13 \times 0.13$ µm^3 per pixel. The dataset consists of 5 volumes, 4 of which were used for training and 1 for testing. The size of each volume was around $500 \times 1000 \times 1000$ pixels.

embryo [104]: instance segmentation of nuclei in mouse embryos, confocal fluorescent microscopy, $1 \times 0.21 \times 0.21$ μm^3 per pixel resampled to isotropic resolution of $0.2 \times 0.2 \times 0.2$ μm^3. The dataset consists of 22 train volumes and 13 validation volumes. The size of each volume is around $600 \times 1000 \times 700$ pixels.

For **plants** and **embryo** datasets ground truth instance masks were converted into cell boundaries, foreground - pixels inside the nuclei and background. The task was to predict 3 channels: boundaries, foreground and background.

transfer setup: networks trained for **embryo** dataset were used to segment **plants** dataset and vice versa. The difference between two confocal datasets is that the images in embryo datasets are more noisy and have larger areas where there are no nuclei in the sample.

6.3.2 *Architectures*

3D U-Net

3D U-Net with 2 downsampling steps and 2 convolutional blocks per level was used for all datasets. Each convolutional block consists of convolutional layer, normalization layer and a non-linearity. The normalization layer in all the blocks were replaced with one of the normalization layers listed above. All models were trained with tile size $(192, 192, 192)$. Tile size $(128, 128, 128)$ was used during sliding window inference.

3D UNETR

UNETR consists of a ViT encoder and convolutional decoder. In the encoder the volume is split into $16 \times 16 \times 16$ patches and embeddings in the ViT are normalized using `LayerNorm` after each round of attention. With `LayerNorm` all features are normalized using average and variance of the features in each patch independently instead of using statistics of the whole image. There exists previous work on using other normalization strategies in ViT [105] but here `LayerNorm` was kept in all experiments as a standard approach. Convolutional decoder consists of 4 upsampling steps with 2 convolutional blocks at each level of the U-Net-like architecture. Each convolutional block consists of convolutional layer, normalization layer and a non-linearity. The normalization layer in all the blocks were replaced with one of the normalization layers listed above.

6.3.3 *Implementation*

All experiments were implemented using MONAI[1] framework. The models were trained on a slurm cluster with NVIDIA GeForce RTX 3090 GPUs with the batch size of 1 and gradient accumulation for 8 steps, using the Adam optimizer with initial learning rate of 0.001 for 25,000 iterations. The only augmentation was flip with probability 0.5. During training tiles of size $192 \times 192 \times 192$ were sampled from the volumes randomly. Dice loss averaged over channels was used for training. Dice is not defined when the label has no non-zero pixels, therefore for tiles with empty channels the empty channels were not used

1 https://monai.io

Architecture	ERF	TRF	halo
U-Net, global feature norm	9	44	22
U-Net, tile-wise feature norm	14	whole tile	22
UNETR	11	whole tile	32

Table 4: Receptive fields and halo for different architectures (reproduced from [85])

in averaging. U-Net with two downsampling steps and 32, 64 and 128 feature maps in the first, second and bottleneck levels was used. For all normalization layers that use running average to collect statistics `momentum` was set to 0.01. UNETR encoder had $16 \times 16 \times 16$ patch size, 12 attention heads and hidden embedding size 768. Convolutional patch projection and trainable positional embedding were used, following the recommendations in the UNETR paper.

6.4 RESULTS

6.4.1 *Receptive field*

I calculated theoretical and effective receptive field for the architectures with different normalization layers.

For U-Net with the global feature normalization (`BatchNorm`, `BatchRenorm`, `Identity`) only a limited region around the central pixel constitutes TRF. In agreement with the well-known results, only the small portion of the TRF makes a large difference for the prediction so the ERF is quite small as shown in the left panel of 33.

For the U-Net with normalization based on the current input (`InstanceNorm`) ERF does not change, as it is determined by the number of layers and stride in convolutional layers, which remain fixed. However, TRF takes up the whole tile as values of all pixels in the tile make a contribution to the prediction for the central pixel via normalization. Seamless stitching is based on the assumption that only a limited part of the tile constitutes TRF, so for the U-Net with `InstanceNorm` stitching can not be completely artifact-free.

Based on these results, I used halo calculated using TRF of the U-Net with global feature normalization (4) for all experiments involving U-Net.

For UNETR TRF also takes the whole input tile, both due to normalization in the convolutional decoder and attention mechanism which uses all patches. The averaged gradients image 33 has clear square pattern coming from the fact that in the encoder the image is downsampled by a factor 16 and then upsampled back to the original resolution.

Surprisingly, ERF of the UNETR turned out to be the same as for the U-Net even though self-attention is supposed to make use of the parts of the image which are not directly adjecent. Since only 3 central patches have large effect on the central pixel prediction, $\texttt{halo} = 32$ was used for all UNETR experiments (4).

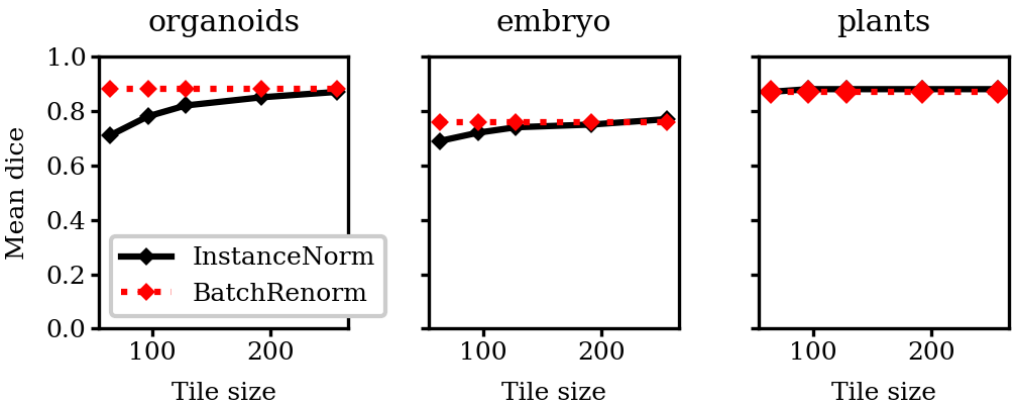


Figure 34: Dependency of performance on tile size (reproduced from [85]).

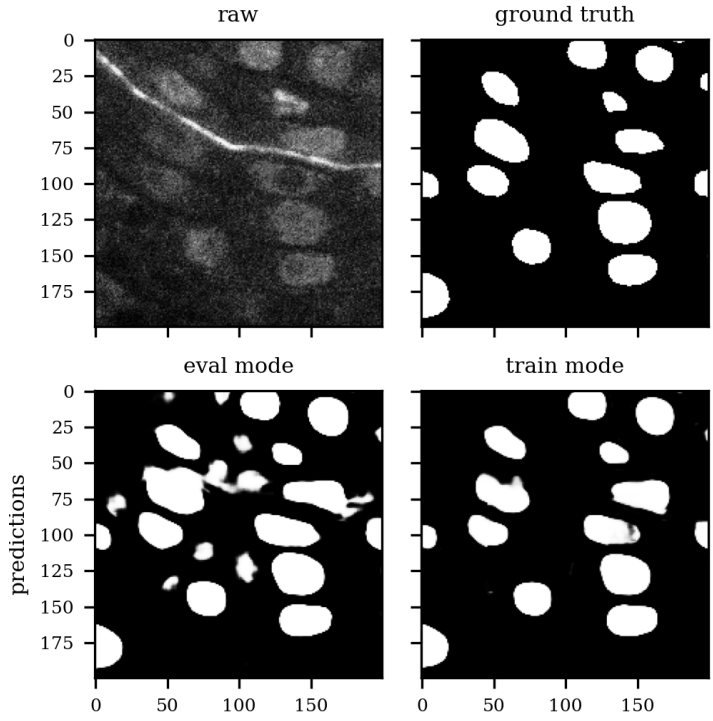


Figure 35: Example of train/eval disparity (reproduced from [85]).

6.4.2 *Tile-wise feature normalization with `InstanceNorm` causes tiling artifacts*

Tiling artifacts can manifest in two ways: as abrupt discrepancy between predictions or as background hallucinations. If the predictions are mostly correct, but do not match between the neighboring tiles, it makes the segmented objects appear sliced, such as in the boundary predictions for embryo dataset in the third column of 31. If some tiles have drastically different content from the average content of the dataset, for example, do not have all the classes, such as in the background patches, local normalization can lead to the wrong predictions. All examples in column three of 31 have this issue. The square edges of hallucinations coinciding with the tile borders suggest that it is a tile-wise normalization issue and not the problem of network not being able to distinguish the objects.

All experiments with tile-wise normalization exhibit some degree of tile mismatch (Table 5 and Table 6), although input normalization (`BatchNorm`, tile-wise input norm) causes less artifacts than the tile-wise feature normalization alone (`InstanceNorm`, global input norm).

For UNETR the severity of the artifacts significantly decreases if the global normalization is used in the convolutional part of the network, even if self-attention makes seamless stitching impossible. Taking into account ERF size, the artifacts become almost unnoticeable.

Tile-wise normalization makes it hard to make correct predictions for classes which cover areas larger than tile size, such as foreground and matrix in organoids dataset. However,`InstanceNorm` performs better than `BatchNorm` for boundaries, which might motivate usage of tile-wise feature normalization even at the cost of introducing the artifacts.

For `InstanceNorm` the prediction quality shows strong dependence on the tile size: the larger the tile and the closer the feature distribution of each tile to the statistics of the full dataset, the better the predictions. This makes it harder to reuse the model in the settings when the user of the model has stricter hardware limitations.

In the transfer experiments I observed that the artifacts caused by tile-wise feature normalization become more severe when the model is applied to a new dataset (Table 7). This makes the models with tiling artifacts less generalizable even in the cases when tile-wise prediction quality is high.

6.4.3 *Mismatch in batch statistics causes `train` and `eval` mode performance disparity for `BatchNorm`*

As shown in 5, `BatchNorm` produces significantly better prediction quality for foreground and matrix classes of the organoids dataset than `InstanceNorm`. These classes take up continuous areas larger than the tile size, so the absence of tiling artifacts improves the prediction score a lot. However, the quality of the boundary prediction is much worse and the predictions for other datasets are worse than with `InstanceNorm`. The reason for this is the disparity between the performance of the network in train and evaluation mode. `BatchNorm` uses the statistics of the current batch during training and collected running average statistics during evaluation. This approach relies on the batch statistics being stable enough and being representative of the whole dataset. If batches are too small then the running average can be very different from the statistics used during the training, which leads to deteriorating performance. Doing inference in the train mode supports this hypothesis: the performance of the networks with `BatchNorm` recovers, at the cost of tiling artifacts. In our experiments train and evaluation mode predictions can be quite different, with train/eval disparity reaching 0.84 for the organoids dataset. Qualitatively, I observe that the changes in the prediction occur in the areas of the least prediction certainty. For example, in 35 the train mode prediction is quite close to the ground truth and eval mode predictions are wrong in the areas where the background has a bright fluorescent object or brighter parts of the cytoplasm. This type of mistakes significantly hinders usage of the predictions for further postprocessing.

Table 5: U-Net results table

Normalization layer			BatchNorm	BatchNorm	InstanceNorm	InstanceNorm	BatchRenorm	Identity
Input norm			global	tile-wise	global	tile-wise	global	global
Feature norm			global	global	tile-wise	tile-wise	global	global
dice, eval mode	organoids	foreground	**0.95**	0.88	0.94	0.93	**0.95**	**0.95**
		boundaries	0.49	0.15	**0.75**	**0.75**	0.73	0.72
		matrix	0.82	0.61	0.61	0.61	**0.85**	0.83
	plants	foreground	0.85	0.88	**0.89**	**0.89**	**0.89**	0.52
		boundaries	0.69	0.73	**0.75**	0.74	**0.75**	0.26
	embryo	foreground	0.66	0.68	0.67	0.72	**0.74**	0.59
		boundaries	0.55	0.55	0.55	0.53	**0.58**	0.44
dice, train mode	organoids	foreground	0.94	0.93	0.94	0.93	**0.95**	0.95
		boundaries	**0.75**	**0.75**	**0.75**	**0.75**	0.73	0.72
		matrix	0.61	0.61	0.61	0.61	**0.85**	0.83
	plants	foreground	**0.89**	**0.89**	**0.89**	**0.89**	**0.89**	0.52
		boundaries	**0.75**	0.74	**0.75**	0.74	**0.75**	0.26
	embryo	foreground	0.67	0.72	0.67	0.72	**0.74**	0.59
		boundaries	0.55	0.52	0.55	0.53	**0.58**	0.44
train/eval disparity	organoids		0.48	0.84	**0**	**0**	**0**	**0**
	plants		0.17	0.11	**0**	**0**	**0**	**0**
	embryo		0.37	0.29	**0**	**0**	**0**	**0**
tile mismatch		organoids	**no**	0.03	0.11	0.11	**no**	**no**
		plants	**no**	0.01	0.05	0.05	**no**	**no**
		embryo	**no**	0.09	0.16	0.20	**no**	**no**

Table 6: UNETR results table

Decoder norm layer			BatchNorm	InstanceNorm	BatchRenorm	Identity
Input norm			global	global	global	global
Feature norm			global	tile-wise	global	global
dice, eval mode	organoids	foreground	0.93	0.89	**0.95**	0.91
		boundaries	0.39	0.60	**0.72**	0.60
		matrix	0.03	0.10	**0.87**	0.28
	plants	foreground	**0.84**	**0.85**	**0.85**	0.52
		boundaries	0.64	**0.69**	0.66	0.21
	embryo	foreground	**0.67**	0.56	**0.67**	0.00
		boundaries	0.49	0.49	**0.53**	0.00
dice, train mode	organoids	foreground	0.89	0.89	**0.95**	0.91
		boundaries	0.60	0.60	**0.72**	0.60
		matrix	0.10	0.10	**0.87**	0.28
	plants	foreground	**0.85**	**0.85**	**0.85**	0.52
		boundaries	**0.69**	**0.69**	0.66	0.21
	embryo	foreground	0.56	0.56	**0.67**	0.00
		boundaries	0.49	0.49	**0.53**	0.00
train/eval disparity	organoids		0.36	**0**	**0**	**0**
	plants		0.20	**0**	**0**	**0**
	embryo		0.33	**0**	**0**	**0**
tile mismatch	organoids		**0.00**	0.10	**0.00**	**0.00**
	plants		**0.00**	0.04	**0.00**	**0.00**
	embryo		**0.00**	0.06	**0.00**	**0.00**

The described effect is different from the usual difference between train and evaluation performance. Normally it is expected that performance of the model is worse on the validation set than on the training set. The explanation is overfitting: validation set contains cases which were not present in the training set and the model did not "generalize" enough, resulting in wrong predictions in unfamiliar cases. The difference in performance stems from the differences in the data and the typical solution is to apply more random augmentations to make the training data distribution closer to the validation data. However this comparison is based on the assumption that the prediction quality is approximately the same in train and eval mode. In this case I run evaluation and calculate train/eval disparity metric using the same data from the validation dataset. This issue is not remedied by using more training data because no matter how much data is used for training, the mismatch in statistics will stay the same.

6.4.4 *BatchRenorm corrects for train/eval disparity and provides seamless stitching*

`InstanceNorm` avoids the statistics mismatch by always using the statistics of the input both during training and during inference at the cost of causing tiling artifacts. `BatchNorm` does not introduce the tiling artifacts by using the same normalization parameters in each tile but can lose performance due to small batch size.

Since normalization layers cause these issues, I tested what happens if the normalization layers are removed from the network. Unfortunately, this led to significantly worse performance for plants dataset (`Identity` in Table 5 and Table 6). It could be possible to improve the performance of the networks with no normalization by finding more suitable training parameters as the learning rate was set to quite high value (`learning rate=0.001` for all experiments). Relative independence of learning parameters is an important advantage of using normalization and necessity to adjust training parameters for each dataset makes the pipeline harder to use.

Another solution is to use the same global normalization statistics both in training and in inference, however, directly using running average statistics during training causes network's weights to explode [102]. Batch Renormalization method uses a reparametrization trick to avoid the divergence and stabilize training while using the same statistics in both modes. I found that networks with `BatchRenorm` produce same or better predictions as `InstanceNorm` (Table 5, Table 6) including the transfer setup (Table 7) while allowing for seamless stitching due to global feature normalization, as shown in the last column of Figure 31 and Figure 36.

6.5 DISCUSSION

The severity of tiling artifacts depends a lot on the particular data, tile size, input normalization, sampling strategy, postprocessing and other parameters of the image analysis pipeline. In some cases only careful comparison of predictions done with different tile size can reveal that some segmentation errors stem from tile-wise normalization issues. Often suboptimal performance is

Table 7: U-Net transfer results

Normalization layer			BatchNorm		BatchNorm		InstanceNorm		InstanceNorm		BatchRenorm		Identity	
Input norm			global		tile-wise		global		tile-wise		global		global	
Feature norm			global		global		tile-wise		tile-wise		global		global	
Target \Source			plant	embryo	plant	embryo	plant	embryo	plant	embryo	plant	embryo	plant	embryo
dice, eval mode	plants	foreground	0.85	0.36	0.88	0.42	0.89	0.50	0.89	**0.53**	0.89	0.50	0.52	0.21
		boundaries	0.69	0.11	0.73	0.14	0.75	0.27	0.74	0.22	0.74	**0.29**	0.26	0.25
	embryo	foreground	0.51	0.66	0.51	0.68	0.54	0.67	0.54	0.72	**0.59**	0.74	0.45	0.59
		boundaries	0.18	0.55	0.14	0.55	0.28	0.55	0.28	0.53	**0.29**	0.58	0.12	0.44
tile mismatch	plants		no	**no**	0.01	0.00	0.05	0.10	0.05	0.07	no	**no**	no	**no**
	embryo		**no**	no	0.17	0.09	0.18	0.16	0.25	0.20	**no**	no	**no**	no

simply attributed to underfitting or too small amount of ground truth annotations. As deep learning pipelines include so many parts, it can be challenging to determine the cause of poor performance. To make it easier to find out the cause of segmenation errors, I introduced two simple metrics: *tile mismatch* and *train/eval disparity*. They can help to determine if a particular method has issues with normalization without retraining the model. Both metrics are applicable for any sliding window inference pipeline even for dense prediction tasks other than segmentation.

In practice, pipelines both with tile mismatch and high train/eval disparity have been successfully applied to segment biological data. However, in my experience, training a deep learning model remains a time-consuming task with unpredictable results. Let us consider how described effects contribute to this.

Even with tile-wise feature normalization the mismatch between tiles can be negligible if the images are very homogeneous in content and tile size is big enough that each tile mostly reproduces the global statistics of the image, such as for plants dataset. Still, even in that case the inference performance depends on the tile size and can change unpredictably in the transfer setting, putting higher requirements on the hardware for inference and making the pipeline harder to use, especially for non-expert users.

Tracing low validation accuracy to the instability in batch statistics is not straightforward too, as it is normally expected that the performance on validation dataset should be worse than on training data. If the train loss is low but the accuracy on validation data does not improve, it is often attributed to the insufficient amount of ground truth, however, train/eval disparity prevents from utilizing the ground truth efficiently.

Both tile mismatch and train/eval disparity are inherent to working with highly diverse large images: it is necessary to estimate the normalization parameters consistent for the whole image due to the seamless stitching constraint but it is not possible to run inference for the whole image during training due to time constraints. `BatchRenorm`, originally introduced for natural image classification, solves this by splitting training in two stages. First, the network is trained with `BatchNorm`, making use of fast convergence and independence of the optimization parameters such as learning rate. When the training largely converged the normalization layer switches to using the accumulated running average statistics. In my experience it is very important to wait until convergence before switching and continue training with relatively low learning rate as the statistics change from tile-wise to global gradually to achieve the best final accuracy. Although this method has additional hyperparameters, in my experiments keeping the same values as in the original publications provided good enough quality.

6.6 CONCLUSION

In this work I have demonstrated the tread-off in segmentation of large biological images: using tile-wise statistics for feature normalization in the network during inference (`InstanceNorm`) leads to tiling artifacts even if edge effects were taken into consideration while using tile-wise statistics in training and

global accumulated statistics during inference (`BatchNorm`) leads to unexpected difference between network performance during training and inference due to small batch size and high variability of data.

`InstanceNorm` [106] is widely used in segmentation networks. It uses statistics of the current input for normalization both during training and evaluation, making predictions in every pixel dependent on the content of the whole tile and causing tiling artifacts.

In contrast with `InstanceNorm`, `BatchNorm` [107] uses the statistics of the current input during training but switches to saved running average statistics in the evaluation mode, allowing to achieve the seamless stitching. However, the performance of the network can deteriorate in evaluation mode compared to training mode if the statistics of individual batches are not representative of the full dataset. It often happens in biological image analysis due to small batch size.

I suggest `BatchRenorm` [102] which uses the same global normalization statistics both during training and inference as a solution which produces artifact-free predictions without compromising on the prediction quality.

Biological images are a very diverse domain with a variety of image modalities and segmentation tasks. In some cases it might be worth having some tiling artifacts in order to win a few more percent in the prediction quality. In other cases, for example, in registration, tile edges can create false features which make downstream analysis impossible, so it might be more important to achieve artifact-free prediction even if it means sacrificing tile-wise prediction quality. I hope that the theoretical framework for reasoning about normalization and the *tile mismatch* and *train/eval disparity* metrics will aid researchers in reaching optimal performance in their pipeline given the limited amount of ground truth and make some aspects of deep learning a little less mysterious.

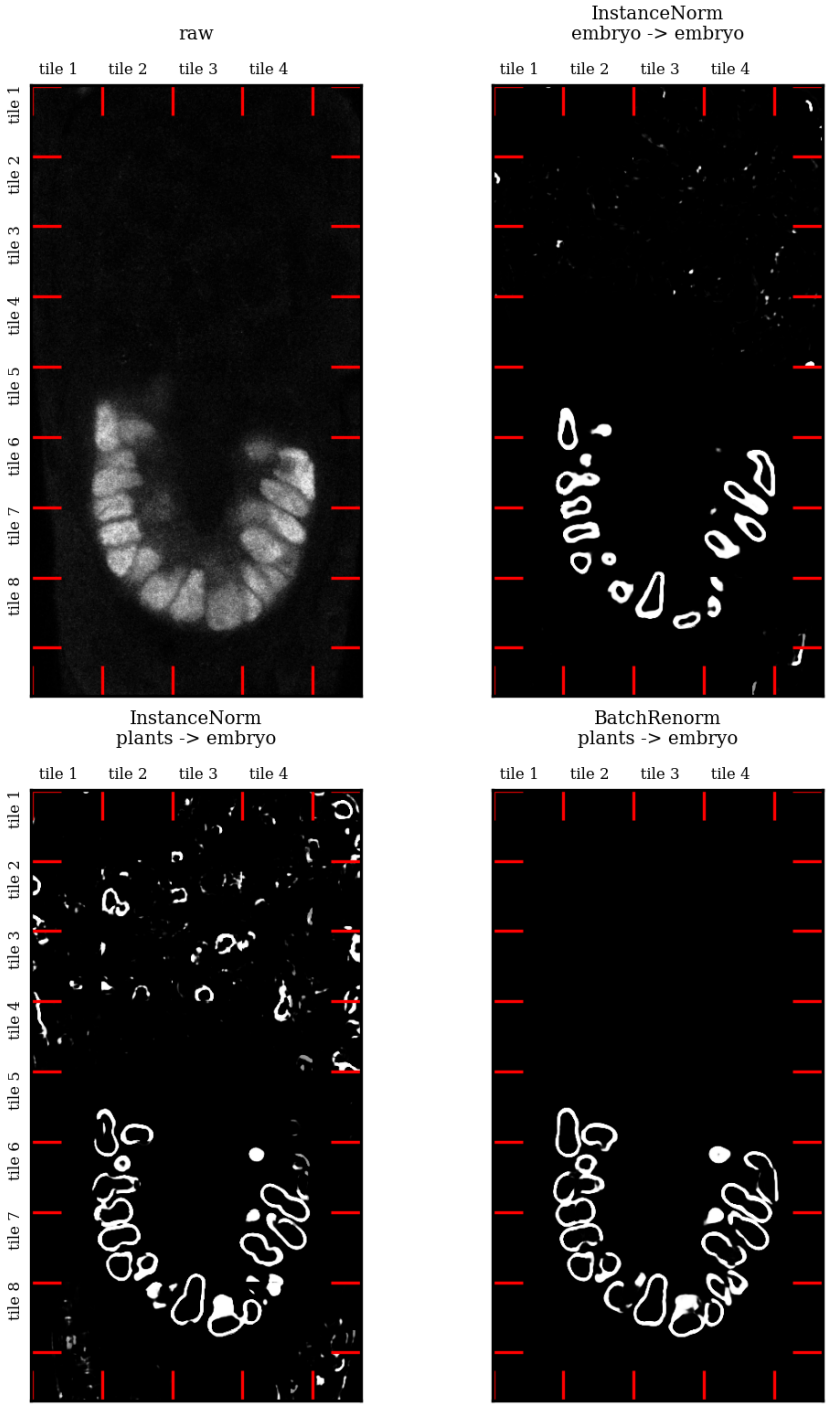


Figure 36: Tiling gets worse in transfer (reproduced from [85])

Part V

CONCLUSIONS AND OUTLOOK

7 CONCLUSIONS AND OUTLOOK

7.1 SUMMARY

During my PhD I worked on correlative workflows including different data modalities, such as imaging mass spectrometry, wide field and confocal microscopy, electron microscopy and scRNAseq.

In the first project described in this thesis I demonstrated automated registration of multiple 3D smFISH datasets with the EM image stack. I used deep learning-based segmentation to detect nuclei in both modalities and then established correspondence between nuclei in confocal volumes and in EM template using a multi-step registration procedure which consists of intensity-based registration, point cloud registration and integer linear programming. I quantified gene expression in smFISH volumes with a custom deep learning-based spot detection method and transferred them to nuclei in EM, therefore assigning gene expression vectors directly to the cells in EM.

In the second project I used already existing single-cell metabolomics method in which correlation of two imaging modalities was based on artificial fiducials and semi-manual registration. I modified the mass spectrometry protocol and have shown that isotope tracing can be applied at single-cell level to find out how much each cell relies on the labeled carbon source.

My experience shows that correlative workflows can be a useful method which helps to interpret both modalities and gives more than just an overlay of two images - however, at a cost of additional difficulties with registration and additional constraints imposed on both correlated analytical methods. Despite following a similar general principle, multimodal workflows are highly specific for the modalities, samples and research questions.

Automated registration remains a bottleneck in correlative workflows. Both in my work and in literature segmentation has been shown to provide robust landmarks for cross-modal registration. This is why in the last part of the thesis I looked into the factors that make segmentation models for biological images harder to train and transfer to new, unseen data.

Characteristic features of biological image analysis domain: high diversity of datasets and large size of the images. Practical scenario for the usage of deep learning models in biological research is human-in-the-loop type of pipelines where a small model is trained for a specific task. Especially for 3D data, the size of one image by far surpasses GPU memory, necessitating sliding window inference. Splitting image into tiles and stitching can result in artifacts, which stem either from edge effects, normalization in the preprocessing pipeline or normalization within the network. It turned out that some commonly used normalization strategies lead to tiling artifacts and result in a tread-off between artifact-free prediction and keeping high-quality predictions under the small batch size conditions and high variability within sample. I tested Batch Renormalization strategy earlier proposed to tackle a similar problem in natural

images and found that it allows to avoid tiling artifacts and improve transferability of the segmentation models.

7.2 CROSS-MODAL REGISTRATION: OUTLOOK

Precise single-cell level registration is a key step in the joint analysis of different modalities. The main challenges in this task are the difference in object appearance and resolution gap between imaging methods. Segmentation can be used to alleviate the appearance difference or find key points for registration. Intensity-based registration and point cloud-based registration have been used for registration of segmented biological images, however there are some problems and use cases that remain unaddressed:

- *Only local registration is feasible*

  Available algorithms provide local registration which means that the result is highly dependent on the initialization. They require good preliminary alignment which is hard to achieve without using manual registration or heuristics such as principal components alignment. This results in registration pipelines that either consist of many registration steps aiming to gradually improve alignment accuracy or require manual adjustment to handle samples in different starting orientation. Robust global alignment algorithm suitable for large-scale biological data would allow to significantly simplify registration pipelines and make registration more accessible to non-expert users.

- *It is hard to incorporate visual features and domain knowledge*

  Available methods match either individual pixel intensity values or the point cloud densities but not the single objects, which is the final goal of this analysis. Consequently, there is no obvious way to make use of other information such as shape descriptors, visual features or high-level information such as prior knowledge about anatomical regions. An algorithm that is able to take this kind of annotations as input would make registration results more predictable and allow to guide registration process in challenging cases while remaining scalable. End-user would be able to steer registration without having to provide extensive manual landmarks for each dataset.

- *There are no readily available methods for registration of partial data*

  Another registration task that remains largely overlooked due to its complexity is the registration when parts of the sample are missing. For example, in the *P. dumerilii* project it would save a lot of imaging time if it were possible to image only a small region of interest, such as a specific brain region, and automatically map it to the full-body template. However, due to lack of the computational methods, it was necessary to image the whole sample.

I believe that the dataset created for the registration quality control of *P. dumerilii* will serve as a valuable benchmark and help to advance the field of

large-scale cross-modal biological image registration and address the problems listed above.

In my opinion, combination of state of the art learning-based methods, such as rotation-invariant graph neural networks (GNNs), together with the specialized combinatorial optimization algorithms is a promising approach for automated biological image registration.

In this type of registration instance segmentation is first converted into kNN graph and each node is assigned two types of features: rotation-invariant geometric features, similar to PointNet [28], and non-geometric features, such as shape descriptors, learned visual features or any other prior information. Next, the GNN is applied to obtain features suitable for registration. The learning process automatically blends geometric and non-geometric features and finds their combination which makes the final representation most distinct and suitable for object matching, similar to how human eye looks for anything that stands out to establish correspondence between two images.

Since the features do not depend on the rotation of the image, the nodes of two graphs can be matched in the feature space using a combinatorial optimization algorithm, providing global registration which does not depend on the initial alignment. Similar approaches were successfully used for registration of the point clouds derived from LiDAR scans. Despite the apparent similarity, point clouds derived from biological data have unique properties so, akin to the segmentation methods adapted from natural images, GNN-based registration will need to be reinvented to achieve global, cross-modal, semantic-aware object matching for biological samples.

Rapid development of imaging technology now allows to bring together modalities with a very large resolution gap and different mechanisms of contrast creation. As workflows such as CLXEM are becoming more routine, and single cell spatial 'omics keeps a high rate of improvement both in experimental and in commercial settings, the need for integration approaches is posed to rise to a new level. I hope that this thesis will contribute both ideas and tools for this exciting field.

BIBLIOGRAPHY

[1] Donald E. Knuth. "Computer Programming as an Art." In: *Communications of the ACM* 17.12 (1974), pp. 667–673.

[2] Joshua A. Pierson, Jie E. Yang, and Elizabeth R. Wright. "Recent advances in correlative cryo-light and electron microscopy." In: *Current Opinion in Structural Biology* 89 (Dec. 2024), p. 102934. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2024.102934. URL: http://dx.doi.org/10.1016/j.sbi.2024.102934.

[3] Notash Shafiei et al. "Correlative light and electron microscopy for human brain and other biological models." In: *Nature Protocols* (Mar. 2025). ISSN: 1750-2799. DOI: 10.1038/s41596-025-01153-9. URL: http://dx.doi.org/10.1038/s41596-025-01153-9.

[4] Hugh Tanner, Olivia Sherwin, and Paul Verkade. "Labelling strategies for correlative light electron microscopy." In: *Microscopy Research and Technique* 86.8 (Feb. 2023), 901–910. ISSN: 1097-0029. DOI: 10.1002/jemt.24304. URL: http://dx.doi.org/10.1002/jemt.24304.

[5] Paolo Ronchi et al. "High-precision targeting workflow for volume electron microscopy." In: *Journal of Cell Biology* 220.9 (June 2021). ISSN: 1540-8140. DOI: 10.1083/jcb.202104069. URL: http://dx.doi.org/10.1083/jcb.202104069.

[6] Perrine Paul-Gilloteaux, Xavier Heiligenstein, Martin Belle, Marie-Charlotte Domart, Banafshe Larijani, Lucy Collinson, Graça Raposo, and Jean Salamero. "eC-CLEM: flexible multidimensional registration software for correlative microscopies." In: *Nature Methods* 14.2 (Jan. 2017), 102–103. ISSN: 1548-7105. DOI: 10.1038/nmeth.4170. URL: http://dx.doi.org/10.1038/nmeth.4170.

[7] David P. Hoffman et al. "Correlative three-dimensional super-resolution and block-face electron microscopy of whole vitreously frozen cells." In: *Science* 367.6475 (Jan. 2020). ISSN: 1095-9203. DOI: 10.1126/science.aaz5357. URL: http://dx.doi.org/10.1126/science.aaz5357.

[8] John A. Bogovic, Philipp Hanslovsky, Allan Wong, and Stephan Saalfeld. "Robust registration of calcium images by learned contrast synthesis." In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, Apr. 2016, 1123–1126. DOI: 10.1109/isbi.2016.7493463. URL: http://dx.doi.org/10.1109/isbi.2016.7493463.

[9] Sergey V. Loginov, Job Fermie, Jantina Fokkema, Alexandra V. Agronskaia, Cilia De Heus, Gerhard A. Blab, Judith Klumperman, Hans C. Gerritsen, and Nalan Liv. "Correlative Organelle Microscopy: Fluorescence Guided Volume Electron Microscopy of Intracellular Processes." In: *Frontiers in Cell and Developmental Biology* 10 (Apr. 2022). ISSN: 2296-634X. DOI: 10.3389/fcell.2022.829545. URL: http://dx.doi.org/10.3389/fcell.2022.829545.

[10] Nadav Scher, Katya Rechav, Perrine Paul-Gilloteaux, and Ori Avinoam. "In situ fiducial markers for 3D correlative cryo-fluorescence and FIB-SEM imaging." In: *iScience* 24.7 (July 2021), p. 102714. ISSN: 2589-0042. DOI: 10.1016/j.isci.2021.102714. URL: http://dx.doi.org/10.1016/j.isci.2021.102714.

[11] Larissa Heinrich et al. "Whole-cell organelle segmentation in volume electron microscopy." In: *Nature* 599.7883 (Oct. 2021), 141–146. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03977-3. URL: http://dx.doi.org/10.1038/s41586-021-03977-3.

[12] S. Klein, M. Staring, K. Murphy, M.A. Viergever, and J. Pluim. "elastix: A Toolbox for Intensity-Based Medical Image Registration." In: *IEEE Transactions on Medical Imaging* 29.1 (Jan. 2010), 196–205. ISSN: 1558-254X. DOI: 10.1109/tmi.2009.2035616. URL: http://dx.doi.org/10.1109/TMI.2009.2035616.

[13] Daniel Krentzel, Matouš Elphick, Marie-Charlotte Domart, Christopher J. Peddie, Romain F. Laine, Cameron Shand, Ricardo Henriques, Lucy M. Collinson, and Martin L. Jones. "CLEM-Reg: An automated point cloud based registration algorithm for correlative light and volume electron microscopy." In: (May 2023). DOI: 10.1101/2023.05.11.540445. URL: http://dx.doi.org/10.1101/2023.05.11.540445.

[14] Lukas Heumos et al. "Best practices for single-cell analysis across modalities." In: *Nature Reviews Genetics* 24.8 (Mar. 2023), 550–572. ISSN: 1471-0064. DOI: 10.1038/s41576-023-00586-w. URL: http://dx.doi.org/10.1038/s41576-023-00586-w.

[15] Joachim De Jonghe et al. "scTrends: A living review of commercial single-cell and spatial 'omic technologies." In: *Cell Genomics* 4.12 (Dec. 2024), p. 100723. ISSN: 2666-979X. DOI: 10.1016/j.xgen.2024.100723. URL: http://dx.doi.org/10.1016/j.xgen.2024.100723.

[16] Chenglong Xia, Hazen P. Babcock, Jeffrey R. Moffitt, and Xiaowei Zhuang. "Multiplexed detection of RNA using MERFISH and branched DNA amplification." In: *Scientific Reports* 9.1 (May 2019). ISSN: 2045-2322. DOI: 10.1038/s41598-019-43943-8. URL: http://dx.doi.org/10.1038/s41598-019-43943-8.

[17] Nils Eling, Michael D. Morgan, and John C. Marioni. "Challenges in measuring and understanding biological noise." In: *Nature Reviews Genetics* 20.9 (May 2019), 536–548. ISSN: 1471-0064. DOI: 10.1038/s41576-019-0130-6. URL: http://dx.doi.org/10.1038/s41576-019-0130-6.

[18] Indhupriya Subramanian, Srikant Verma, Shiva Kumar, Abhay Jere, and Krishanpal Anamika. "Multi-omics Data Integration, Interpretation, and Its Application." In: *Bioinformatics and Biology Insights* 14 (Jan. 2020), p. 117793221989905. ISSN: 1177-9322. DOI: 10.1177/1177932219899051. URL: http://dx.doi.org/10.1177/1177932219899051.

[19] Sophia K. Longo, Margaret G. Guo, Andrew L. Ji, and Paul A. Khavari. "Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics." In: *Nature Reviews Genetics* 22.10 (June 2021), 627–644. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00370-8. URL: http://dx.doi.org/10.1038/s41576-021-00370-8.

[20] Szeliski and Coughlan. "Hierarchical spline-based image registration." In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*. IEEE Comput. Soc. Press, 1994, 194–201. DOI: 10.1109/cvpr.1994.323829. URL: http://dx.doi.org/10.1109/cvpr.1994.323829.

[21] B.S. Reddy and B.N. Chatterji. "An FFT-based technique for translation, rotation, and scale-invariant image registration." In: *IEEE Transactions on Image Processing* 5.8 (Aug. 1996), 1266–1271. ISSN: 1941-0042. DOI: 10.1109/83.506761. URL: http://dx.doi.org/10.1109/83.506761.

[22] Xiaoxin Guo, Zhiwen Xu, Yinan Lu, and Yunjie Pang. "An Application of Fourier-Mellin Transform in Image Registration." In: *The Fifth International Conference on Computer and Information Technology (CIT'05)*. IEEE, 2005, 619–623. DOI: 10.1109/cit.2005.62. URL: http://dx.doi.org/10.1109/CIT.2005.62.

[23] Fatemehzahra Darzi and Thomas Bocklitz. "A Review of Medical Image Registration for Different Modalities." In: *Bioengineering* 11.8 (Aug. 2024), p. 786. ISSN: 2306-5354. DOI: 10.3390/bioengineering11080786. URL: http://dx.doi.org/10.3390/bioengineering11080786.

[24] Vincent Leroy, Yohann Cabon, and Jérôme Revaud. *Grounding Image Matching in 3D with MASt3R*. 2024. DOI: 10.48550/ARXIV.2406.09756. URL: https://arxiv.org/abs/2406.09756.

[25] David G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints." In: *International Journal of Computer Vision* 60.2 (Nov. 2004), 91–110. ISSN: 0920-5691. DOI: 10.1023/b:visi.0000029664.99615.94. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

[26] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. "Deep learning in medical image registration: a review." In: *Physics in Medicine amp; Biology* 65.20 (Oct. 2020), 20TR01. ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab843e. URL: http://dx.doi.org/10.1088/1361-6560/ab843e.

[27] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. "SuperGlue: Learning Feature Matching With Graph Neural Networks." In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020, 4937–4946. DOI: 10.1109/cvpr42600.2020.00499. URL: http://dx.doi.org/10.1109/cvpr42600.2020.00499.

[28] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017, 77–85. DOI: 10.1109/cvpr.2017.16. URL: http://dx.doi.org/10.1109/CVPR.2017.16.

[29] Aleksandar Shtedritski, Christian Rupprecht, and Andrea Vedaldi. *SHIC: Shape-Image Correspondences with no Keypoint Supervision*. 2024. DOI: 10.48550/ARXIV.2407.18907. URL: https://arxiv.org/abs/2407.18907.

[30] Haowen Deng, Tolga Birdal, and Slobodan Ilic. *PPFNet: Global Context Aware Local Features for Robust 3D Point Matching*. 2018. DOI: 10.48550/ARXIV.1802.02669. URL: https://arxiv.org/abs/1802.02669.

[31] B. Duygu Özpolat et al. "The Nereid on the rise: Platynereis as a model system." In: *EvoDevo* 12.1 (Sept. 2021). ISSN: 2041-9139. DOI: 10.1186/s13227-021-00180-3. URL: http://dx.doi.org/10.1186/s13227-021-00180-3.

[32] Antje HL Fischer, Thorsten Henrich, and Detlev Arendt. "The normal development of Platynereis dumerilii (Nereididae, Annelida)." In: *Frontiers in Zoology* 7.1 (2010), p. 31. ISSN: 1742-9994. DOI: 10.1186/1742-9994-7-31. URL: http://dx.doi.org/10.1186/1742-9994-7-31.

[33] Raju Tomer, Alexandru S. Denes, Kristin Tessmar-Raible, and Detlev Arendt. "Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium." In: *Cell* 142.5 (Sept. 2010), 800–809. ISSN: 0092-8674. DOI: 10.1016/j.cell.2010.07.043. URL: http://dx.doi.org/10.1016/j.cell.2010.07.043.

[34] Albina Asadulina, Aurora Panzera, Csaba Verasztó, Christian Liebig, and Gáspár Jékely. "Whole-body gene expression pattern registration in Platynereis larvae." In: *EvoDevo* 3.1 (Dec. 2012). ISSN: 2041-9139. DOI: 10.1186/2041-9139-3-27. URL: http://dx.doi.org/10.1186/2041-9139-3-27.

[35] Hernando Martínez Vergara, Paola Yanina Bertucci, Peter Hantz, Maria Antonietta Tosches, Kaia Achim, Pavel Vopalensky, and Detlev Arendt. "Whole-organism cellular gene-expression atlas reveals conserved cell types in the ventral nerve cord of Platynereis dumerilii." In: *Proceedings of the National Academy of Sciences* 114.23 (June 2017), 5878–5885. ISSN: 1091-6490. DOI: 10.1073/pnas.1610602114. URL: http://dx.doi.org/10.1073/pnas.1610602114.

[36] Hernando M. Vergara et al. "Whole-body integration of gene expression and single-cell morphology." In: *Cell* 184.18 (Sept. 2021), 4819–4837.e22. ISSN: 0092-8674. DOI: 10.1016/j.cell.2021.07.017. URL: http://dx.doi.org/10.1016/j.cell.2021.07.017.

[37] Valentyna Zinchenko, Johannes Hugger, Virginie Uhlmann, Detlev Arendt, and Anna Kreshuk. "MorphoFeatures for unsupervised exploration of cell types, tissues, and organs in volume electron microscopy." In: *eLife* 12 (Feb. 2023). ISSN: 2050-084X. DOI: 10.7554/elife.80918. URL: http://dx.doi.org/10.7554/eLife.80918.

[38] David A. Salamanca-Díaz, Stephan M. Schulreich, Alison G. Cole, and Andreas Wanninger. "Single-Cell RNA Sequencing Atlas From a Bivalve Larva Enhances Classical Cell Lineage Studies." In: *Frontiers in Ecology and Evolution* 9 (Jan. 2022). ISSN: 2296-701X. DOI: 10.3389/fevo.2021.783984. URL: http://dx.doi.org/10.3389/fevo.2021.783984.

[39] Patricia Álvarez Campos et al. "Annelid adult cell type diversity and their pluripotent cellular origins." In: *Nature Communications* 15.1 (Apr. 2024). ISSN: 2041-1723. DOI: 10.1038/s41467-024-47401-6. URL: http://dx.doi.org/10.1038/s41467-024-47401-6.

[40] Kaia Achim et al. "Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body." In: *Molecular Biology and Evolution* 35.5 (Jan. 2018). Ed. by Gunter Wagner, 1047–1062. ISSN: 1537-1719. DOI: 10.1093/molbev/msx336. URL: http://dx.doi.org/10.1093/molbev/msx336.

[41] Laura Piovani, Daniel J. Leite, Luis Alfonso Yañez Guerra, Fraser Simpson, Jacob M. Musser, Irepan Salvador-Martínez, Ferdinand Marlétaz, Gáspár Jékely, and Maximilian J. Telford. "Single-cell atlases of two lophotrochozoan larvae highlight their complex evolutionary histories." In: *Science Advances* 9.31 (Aug. 2023). ISSN: 2375-2548. DOI: 10.1126/sciadv.adg6034. URL: http://dx.doi.org/10.1126/sciadv.adg6034.

[42] Harry M. T. Choi, Maayan Schwarzkopf, Mark E. Fornace, Aneesh Acharya, Georgios Artavanis, Johannes Stegmaier, Alexandre Cunha, and Niles A. Pierce. "Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust." In: *Development* 145.12 (June 2018). ISSN: 0950-1991. DOI: 10.1242/dev.165753. URL: http://dx.doi.org/10.1242/dev.165753.

[43] Yu Toyoshima et al. "Neuron ID dataset facilitates neuronal annotation for whole-brain activity imaging of C. elegans." In: *BMC Biology* 18.1 (Mar. 2020). ISSN: 1741-7007. DOI: 10.1186/s12915-020-0745-2. URL: http://dx.doi.org/10.1186/s12915-020-0745-2.

[44] Eviatar Yemini, Albert Lin, Amin Nejatbakhsh, Erdem Varol, Ruoxi Sun, Gonzalo E. Mena, Aravinthan D.T. Samuel, Liam Paninski, Vivek Venkatachalam, and Oliver Hobert. "NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans." In: *Cell* 184.1 (Jan. 2021), 272–288.e11. ISSN: 0092-8674. DOI: 10.1016/j.cell.2020.12.012. URL: http://dx.doi.org/10.1016/j.cell.2020.12.012.

[45] Dagmar Kainmueller, Florian Jug, Carsten Rother, and Gene Myers. "Active Graph Matching for Automatic Joint Segmentation and Annotation of C. elegans." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Springer International Publishing, 2014, 81–88. ISBN: 9783319104041. DOI: 10.1007/978-3-319-10404-1_11. URL: http://dx.doi.org/10.1007/978-3-319-10404-1_11.

[46] Christoph Karg, Sebastian Stricker, Lisa Hutschenreiter, Bogdan Savchynskyy, and Dagmar Kainmueller. *Fully Unsupervised Annotation of C. Elegans*. 2025. DOI: 10.48550/ARXIV.2503.07348. URL: https://arxiv.org/abs/2503.07348.

[47] P.J. Besl and Neil D. McKay. "A method for registration of 3-D shapes." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (Feb. 1992), 239–256. ISSN: 2160-9292. DOI: 10.1109/34.121791. URL: http://dx.doi.org/10.1109/34.121791.

[48] Victoria E. Centonze and John G. White. "Multiphoton Excitation Provides Optical Sections from Deeper within Scattering Specimens than Confocal Imaging." In: *Biophysical Journal* 75.4 (Oct. 1998), 2015–2024. ISSN: 0006-3495. DOI: 10.1016/s0006-3495(98)77643-x. URL: http://dx.doi.org/10.1016/S0006-3495(98)77643-X.

[49] Aurore Masson, Paul Escande, Céline Frongia, Grégory Clouvel, Bernard Ducommun, and Corinne Lorenzo. "High-resolution in-depth imaging of optically cleared thick samples using an adaptive SPIM." In: *Scientific Reports* 5.1 (Nov. 2015). ISSN: 2045-2322. DOI: 10.1038/srep16898. URL: http://dx.doi.org/10.1038/srep16898.

[50] Gustavo de Medeiros, Nils Norlin, Stefan Gunther, Marvin Albert, Laura Panavaite, Ulla-Maj Fiuza, Francesca Peri, Takashi Hiiragi, Uros Krzic, and Lars Hufnagel. "Confocal multiview light-sheet microscopy." In: *Nature Communications* 6.1 (Nov. 2015). ISSN: 2041-1723. DOI: 10.1038/ncomms9881. URL: http://dx.doi.org/10.1038/ncomms9881.

[51] Thorsten Beier et al. "Multicut brings automated neurite segmentation closer to human performance." In: *Nature Methods* 14.2 (Feb. 2017), 101–102. ISSN: 1548-7105. DOI: 10.1038/nmeth.4151. URL: http://dx.doi.org/10.1038/nmeth.4151.

[52] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. "Cellpose: a generalist algorithm for cellular segmentation." In: *Nature Methods* 18.1 (Dec. 2020), 100–106. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01018-x. URL: http://dx.doi.org/10.1038/s41592-020-01018-x.

[53] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. "Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy." In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Mar. 2020, 3655–3662. DOI: 10.1109/wacv45572.2020.9093435. URL: http://dx.doi.org/10.1109/wacv45572.2020.9093435.

[54] Dmitry Ershov et al. "TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines." In: *Nature Methods* 19.7 (June 2022), 829–832. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01507-1. URL: http://dx.doi.org/10.1038/s41592-022-01507-1.

[55] Marius Pachitariu and Carsen Stringer. "Cellpose 2.0: how to train your own model." In: *Nature Methods* 19.12 (Nov. 2022), 1634–1641. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01663-4. URL: http://dx.doi.org/10.1038/s41592-022-01663-4.

[56] Alexander Kirillov et al. *Segment Anything*. 2023. DOI: 10.48550/ARXIV.2304.02643. URL: https://arxiv.org/abs/2304.02643.

[57] Adrian Wolny et al. "Accurate and versatile 3D segmentation of plant tissues at cellular resolution." In: *eLife* 9 (July 2020). ISSN: 2050-084X. DOI: 10.7554/elife.57613. URL: http://dx.doi.org/10.7554/eLife.57613.

[58] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf. "Parametric correspondence and chamfer matching: two new techniques for image matching." In: *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'77. Cambridge, USA: Morgan Kaufmann Publishers Inc., 1977, 659–663.

[59] Athul Vijayan et al. "A deep learning-based toolkit for 3D nuclei segmentation and quantitative analysis in cellular and tissue context." In: *Development* 151.14 (July 2024). ISSN: 1477-9129. DOI: 10.1242/dev.202800. URL: http://dx.doi.org/10.1242/dev.202800.

[60] Steffen Wolf, Alberto Bailoni, Constantin Pape, Nasim Rahaman, Anna Kreshuk, Ullrich Köthe, and Fred A. Hamprecht. "The Mutex Watershed and its Objective: Efficient, Parameter-Free Graph Partitioning." In: (2019). DOI: 10.48550/ARXIV.1904.12654. URL: https://arxiv.org/abs/1904.12654.

[61] Albert Dominguez Mantes et al. "Spotiflow: accurate and efficient spot detection for fluorescence microscopy with deep stereographic flow regression." In: (Feb. 2024). DOI: 10.1101/2024.02.01.578426. URL: http://dx.doi.org/10.1101/2024.02.01.578426.

[62] Arthur Imbert, Wei Ouyang, Adham Safieddine, Emeline Coleno, Christophe Zimmer, Edouard Bertrand, Thomas Walter, and Florian Mueller. "FISH-quant v2: a scalable and modular tool for smFISH image analysis." In: *RNA* 28.6 (Mar. 2022), 786–795. ISSN: 1469-9001. DOI: 10.1261/rna.079073.121. URL: http://dx.doi.org/10.1261/rna.079073.121.

[63] Shannon Axelrod, Matthew Cai, Ambrose Carr, Jeremy Freeman, Deep Ganguli, Justin Kiggins, Brian Long, Tony Tung, and Kevin Yamauchi. "starfish: scalable pipelines for image-based transcriptomics." In: *Journal of Open Source Software* 6.61 (May 2021), p. 2440. ISSN: 2475-9066. DOI: 10.21105/joss.02440. URL: http://dx.doi.org/10.21105/joss.02440.

[64] T. Wollmann, C. Ritter, J. N. Dohrke, J.-Y. Lee, R. Bartenschlager, and K. Rohr. "Detnet: Deep Neural Network For Particle Detection In Fluorescence Microscopy Images." In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, Apr. 2019, 517–520. DOI: 10.1109/isbi.2019.8759234. URL: http://dx.doi.org/10.1109/ISBI.2019.8759234.

[65] Kristen R Maynard, Madhavi Tippani, Yoichiro Takahashi, BaDoi N Phan, Thomas M Hyde, Andrew E Jaffe, and Keri Martinowich. "dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images in complex tissues." In: *Nucleic Acids Research* 48.11 (May 2020), e66–e66. ISSN: 1362-4962. DOI: 10.1093/nar/gkaa312. URL: http://dx.doi.org/10.1093/nar/gkaa312.

[66] Prabhakar R. Gudla, Koh Nakayama, Gianluca Pegoraro, and Tom Misteli. "SpotLearn: Convolutional Neural Network for Detection of Fluorescence In Situ Hybridization (FISH) Signals in High-Throughput Imaging Approaches." In: *Cold Spring Harbor Symposia on Quantitative*

Biology 82 (2017), 57–70. ISSN: 1943-4456. DOI: 10.1101/sqb.2017.82.033761. URL: http://dx.doi.org/10.1101/sqb.2017.82.033761.

[67] Bastian Th Eichenberger, YinXiu Zhan, Markus Rempfler, Luca Giorgetti, and Jeffrey A Chao. "deepBlink: threshold-independent detection and localization of diffraction-limited spots." In: *Nucleic Acids Research* 49.13 (July 2021), 7292–7297. ISSN: 1362-4962. DOI: 10.1093/nar/gkab546. URL: http://dx.doi.org/10.1093/nar/gkab546.

[68] Konstantinos Ntatsis, Niels Dekker, Viktor van der Valk, Tom Birdsong, Dženan Zukić, Stefan Klein, Marius Staring, and Matthew McCormick. "itk-elastix: Medical image registration in Python." In: *Proceedings of the 22nd Python in Science Conference*. SciPy. SciPy, 2023, 101–105. DOI: 10.25080/gerudo-f2bc6f59-00d. URL: http://dx.doi.org/10.25080/gerudo-f2bc6f59-00d.

[69] Andriy Myronenko and Xubo Song. "Point-Set Registration: Coherent Point Drift." In: (2009). DOI: 10.48550/ARXIV.0905.2635. URL: https://arxiv.org/abs/0905.2635.

[70] Kenta-Tanaka et al. *probreg*. Version 0.1.6. Sept. 29, 2019. URL: https://probreg.readthedocs.io/en/latest/.

[71] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. "A rewriting system for convex optimization problems." In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[72] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization." In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[73] B AVANTS, C EPSTEIN, M GROSSMAN, and J GEE. "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain." In: *Medical Image Analysis* 12.1 (Feb. 2008), 26–41. ISSN: 1361-8415. DOI: 10.1016/j.media.2007.06.004. URL: http://dx.doi.org/10.1016/j.media.2007.06.004.

[74] Constantin Pape et al. "MoBIE: a Fiji plugin for sharing and exploration of multi-modal cloud-hosted big image data." In: *Nature Methods* 20.4 (Feb. 2023), 475–476. ISSN: 1548-7105. DOI: 10.1038/s41592-023-01776-4. URL: http://dx.doi.org/10.1038/s41592-023-01776-4.

[75] Tobias Pietzsch, Stephan Saalfeld, Stephan Preibisch, and Pavel Tomancak. "BigDataViewer: visualization and processing for large image data sets." In: *Nature Methods* 12.6 (May 2015), 481–483. ISSN: 1548-7105. DOI: 10.1038/nmeth.3392. URL: http://dx.doi.org/10.1038/nmeth.3392.

[76] Felix Mölder et al. "Sustainable data analysis with Snakemake." In: *F1000Research* 10 (Apr. 2021), p. 33. ISSN: 2046-1402. DOI: 10.12688/f1000research.29032.2. URL: http://dx.doi.org/10.12688/f1000research.29032.2.

[77] Luca Rappez, Mira Stadler, Sergio Triana, Rose Muthoni Gathungu, Katja Ovchinnikova, Prasad Phapale, Mathias Heikenwalder, and Theodore Alexandrov. "SpaceM reveals metabolic states of single cells." In: *Nature Methods* 18.7 (July 2021), 799–805. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01198-0. URL: http://dx.doi.org/10.1038/s41592-021-01198-0.

[78] Elena Buglakova et al. "Spatial single-cell isotope tracing reveals heterogeneity of de novo fatty acid synthesis in cancer." In: *Nature Metabolism* 6.9 (Sept. 2024), 1695–1711. ISSN: 2522-5812. DOI: 10.1038/s42255-024-01118-4. URL: http://dx.doi.org/10.1038/s42255-024-01118-4.

[79] Joerg M Buescher et al. "A roadmap for interpreting 13 C metabolite labeling patterns from cells." In: *Current Opinion in Biotechnology* 34 (Aug. 2015), 189–201. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2015.02.003. URL: http://dx.doi.org/10.1016/j.copbio.2015.02.003.

[80] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp. "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." In: *International Journal of Mass Spectrometry and Ion Processes* 78 (Sept. 1987), 53–68. ISSN: 0168-1176. DOI: 10.1016/0168-1176(87)87041-6. URL: http://dx.doi.org/10.1016/0168-1176(87)87041-6.

[81] Richard M. Caprioli, Terry B. Farmer, and Jocelyn Gile. "Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS." In: *Analytical Chemistry* 69.23 (Dec. 1997), 4751–4760. ISSN: 1520-6882. DOI: 10.1021/ac970888i. URL: http://dx.doi.org/10.1021/ac970888i.

[82] Cholsoon Jang, Li Chen, and Joshua D. Rabinowitz. "Metabolomics and Isotope Tracing." In: *Cell* 173.4 (May 2018), 822–837. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.055. URL: http://dx.doi.org/10.1016/j.cell.2018.03.055.

[83] Maciek R. Antoniewicz. "A guide to 13C metabolic flux analysis for the cancer biologist." In: *Experimental & Molecular Medicine* 50.4 (Apr. 2018), 1–13. ISSN: 2092-6413. DOI: 10.1038/s12276-018-0060-y. URL: http://dx.doi.org/10.1038/s12276-018-0060-y.

[84] Sergey Tumanov, Vinay Bulusu, and Jurre J. Kamphorst. "Analysis of Fatty Acid Metabolism Using Stable Isotope Tracers and Mass Spectrometry." In: *Metabolic Analysis Using Stable Isotopes*. Elsevier, 2015, 197–217. DOI: 10.1016/bs.mie.2015.05.017. URL: http://dx.doi.org/10.1016/bs.mie.2015.05.017.

[85] Elena Buglakova, Anwai Archit, Edoardo D'Imprima, Julia Mahamid, Constantin Pape, and Anna Kreshuk. *Tiling artifacts and trade-offs of feature normalization in the segmentation of large biological images*. 2025. DOI: 10.48550/ARXIV.2503.19545. URL: https://arxiv.org/abs/2503.19545.

[86] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "ImageNet: A large-scale hierarchical image database." In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[87] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. "The Cityscapes Dataset for Semantic Urban Scene Understanding." In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[88] Anwai Archit, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Sagnik Gupta, Andreas Dengel, Sheraz Ahmed, and Constantin Pape. "Segment Anything for Microscopy." In: (Aug. 2023). DOI: 10.1101/2023.08.21.554208. URL: http://dx.doi.org/10.1101/2023.08.21.554208.

[89] Sven Dorkenwald et al. "Neuronal wiring diagram of an adult brain." en. In: *Nature* 634.8032 (Oct. 2024), pp. 124–138.

[90] Xiongtao Ruan et al. "Image processing tools for petabyte-scale light sheet microscopy data." en. In: *Nat. Methods* 21.12 (Dec. 2024), pp. 2342–2352.

[91] Josef Lorenz Rumberger, Xiaoyan Yu, Peter Hirsch, Melanie Dohmen, Vanessa Emanuela Guarino, Ashkan Mokarian, Lisa Mais, Jan Funke, and Dagmar Kainmueller. "How Shift Equivariance Impacts Metric Learning for Instance Segmentation." In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2021, 7108–7116. DOI: 10.1109/iccv48922.2021.00704. URL: http://dx.doi.org/10.1109/iccv48922.2021.00704.

[92] T. Falk et al. "U-Net – Deep Learning for Cell Counting, Detection, and Morphometry." In: *Nature Methods* 16 (2019), pp. 67–70. URL: http://lmb.informatik.uni-freiburg.de/Publications/2019/FMBCAMBBR19.

[93] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, 234–241. ISBN: 9783319245744. DOI: 10.1007/978-3-319-24574-4_28. URL: http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[94] Özgün undefinediçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Springer International Publishing, 2016, 424–432. ISBN: 9783319467238. DOI: 10.1007/978-3-319-46723-8_49. URL: http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[95] Bohao Huang, Daniel Reichman, Leslie M. Collins, Kyle Bradbury, and Jordan M. Malof. "Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations." In: *arXiv: Computer Vision and Pattern Recognition* (2018). URL: https://api.semanticscholar.org/CorpusID:86748431.

[96] Michael Majurski and Peter Bajcsy. "Exact Tile-Based Segmentation Inference for Images Larger than GPU Memory." In: *Journal of Research of the National Institute of Standards and Technology* 126 (June 2021). ISSN: 2165-7254. DOI: 10.6028/jres.126.009. URL: http://dx.doi.org/10.6028/jres.126.009.

[97] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks." In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29. Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/c8067ad1937f728f51288b3eb986afaa-Paper.pdf.

[98] Vincent Loos, Rohit Pardasani, and Navchetan Awasthi. "Demystifying the effect of receptive field size in U-Net models for medical image segmentation." In: *Journal of Medical Imaging* 11.05 (Oct. 2024). ISSN: 2329-4302. DOI: 10.1117/1.jmi.11.5.054004. URL: http://dx.doi.org/10.1117/1.JMI.11.5.054004.

[99] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." In: *Nature Methods* 18.2 (Dec. 2020), 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: http://dx.doi.org/10.1038/s41592-020-01008-z.

[100] G. Anthony Reina, Ravi Panchumarthy, Siddhesh Pravin Thakur, Alexei Bastidas, and Spyridon Bakas. "Systematic Evaluation of Image Tiling Adverse Effects on Deep Learning Semantic Segmentation." In: *Frontiers in Neuroscience* 14 (Feb. 2020). ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00065. URL: http://dx.doi.org/10.3389/fnins.2020.00065.

[101] Calimanut-Ionut Cira, Miguel-Ángel Manso-Callejo, Naoto Yokoya, Tudor Sălăgean, and Ana-Cornelia Badea. "Impact of Tile Size and Tile Overlap on the Prediction Performance of Convolutional Neural Networks Trained for Road Classification." In: *Remote Sensing* 16.15 (July 2024), p. 2818. ISSN: 2072-4292. DOI: 10.3390/rs16152818. URL: http://dx.doi.org/10.3390/rs16152818.

[102] Sergey Ioffe. "Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models." In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/c54e7837e0cd0ced286cb5995327d1ab-Paper.pdf.

[103] Edoardo D'Imprima, Marta Garcia Montero, Sylwia Gawrzak, Paolo Ronchi, Ievgeniia Zagoriy, Yannick Schwab, Martin Jechlinger, and Julia Mahamid. "Light and electron microscopy continuum-resolution imaging of 3D cell cultures." In: *Developmental Cell* 58.7 (Apr. 2023), 616–632.e6. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2023.03.001. URL: http://dx.doi.org/10.1016/j.devcel.2023.03.001.

[104] Vladyslav Bondarenko et al. "Embryo-uterine interaction coordinates mouse embryogenesis during implantation." In: *The EMBO Journal* 42.17 (July 2023). ISSN: 1460-2075. DOI: 10.15252/embj.2022113280. URL: http://dx.doi.org/10.15252/embj.2022113280.

[105] Zhuliang Yao, Yue Cao, Yutong Lin, Ze Liu, Zheng Zhang, and Han Hu. "Leveraging Batch Normalization for Vision Transformers." In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, Oct. 2021. DOI: 10.1109/iccvw54120.2021.00050. URL: http://dx.doi.org/10.1109/iccvw54120.2021.00050.

[106] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. *Instance Normalization: The Missing Ingredient for Fast Stylization*. 2016. DOI: 10.48550/ARXIV.1607.08022. URL: https://arxiv.org/abs/1607.08022.

[107] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, 2015, pp. 448–456. URL: https://proceedings.mlr.press/v37/ioffe15.html.