Assisting in the Reuse of Artificial Intelligence Approaches in Biological Image Processing

Zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik des
Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von
Hamideh Hajiabadi

Tag der mündlichen Prüfung: 27. Januar 2025
Erster Gutachter: Prof. Dr. Anne Koziolek
Zweiter Gutachter: Prof. Dr. Stefan Wagner
Dritter Gutachter: TT-Prof. Dr. Lennart Hilbert

Abstract

Modern scientific software, particularly in fields like biological image analysis, faces significant challenges due to the complexity and variability of the data. While Deep Learning (DL) and Machine Learning (ML) provide powerful tools to address these challenges, integrating them into domain-specific workflows with quality assurance often requires tailored solutions. Many biologists, as non-experts in computer science, encounter difficulties in selecting and effectively reusing these advanced tools, particularly when ensuring the trustworthiness and reliability of the results is critical. Generalized approaches often fail to address the unique conditions of biological imaging. This thesis contributes to SE for AI by focusing on the reuse of ML and DL models in specialized applications, with an emphasis on accessibility for non-expert users, robust quality assurance, and trustworthy outcomes.

The first contribution is a pragmatic imaging protocol for reusing quality-assured DL-based denoising techniques in high-speed fluorescence microscopy. This protocol emphasizes systematic practices inspired by SE principles, addressing key challenges such as the trade-off between imaging speed and signal-to-noise ratio (SNR), producing reliable results, and accessibility for biologists. The protocol uses unsupervised DL methods, such as Noise2Void (N2V), and integrates quality assurance metrics like structural similarity (SSIM) and Fourier ring correlation (FRC). Applied to live imaging of zebrafish embryos, this protocol achieved a fivefold increase in imaging speed while maintaining accuracy and reliability. It enabled the discovery of novel biological processes, showing its value in dynamic biological studies where high-speed and trustworthy results are critical.

The second contribution is a framework for weakly supervised ML segmentation, designed to simplify the selection, reuse, and optimization of segmentation algorithms with minimal user input. By integrating interactive user input with active learning, the framework iteratively refines segmentation settings, enabling effective segmentation without requiring annotations. The design uses SE practices such as quality assurance, and iterative refinement to adapt to the variability of biological data. Through a real case study in zebrafish embryos, the framework demonstrated its ability to handle diverse datasets and achieve high accuracy with minimal user intervention.

Together, these contributions bridge the gap between advanced ML/DL technologies and their practical application in biological imaging. By incorporating SE principles, this thesis provides reusable, accessible, and reliable approaches for biologists.

Contents

List of Figures

List of Tables

Part I.

Prologue

1. Introduction

In this chapter, we explore the essential steps in acquiring, processing, and analyzing biological images. These steps include image acquisition, pre-processing, object detection, feature extraction, and the analysis of extracted features. Each stage plays a critical role in generating reliable results that can support biological discoveries.

Next, we discuss how ML methods, including both supervised and unsupervised approaches, can enhance the accuracy and speed of biological image analysis. While these methods offer potential, their reuse comes with challenges—particularly in ensuring the trustworthiness and reliability of the results. This challenge is amplified in the absence of tailored quality assurance measures, which are often necessary due to the domain-specific complexities of biological data.

We then examine the unique difficulties in applying ML methods, focusing on the contrasting challenges of supervised and unsupervised approaches. Supervised methods require large annotated datasets, which are often unavailable in biological imaging, while unsupervised methods, such as denoising and segmentation, can introduce errors or artifacts (fake structure) that compromise data reliability. Additionally, we address the specific issues associated with light microscopy, such as managing trade-offs between imaging speed, image quality, and photodamage, which further complicate the reuse of ML models. These factors make it difficult to generalize solutions across tasks and increase reliance on users with advanced programming expertise.

To address these challenges, this chapter outlines our research goals, which focus on making ML models accessible and trustworthy for biologists with limited technical expertise. This is achieved by incorporating principles from software engineering for AI (SE 4 AI), emphasizing the need for tailored quality assurance mechanisms [67]. General approaches to quality assurance often fall short in domain-specific contexts like biological imaging, where variability in experimental conditions demands adaptable and robust solutions.

Finally, we provide an overview of the thesis, outlining its structure and the key contributions it makes to enhance the reuse of ML methods in biological image analysis.

1.1. Biological image acquisition, processing and analysis

Biological image processing and analysis involve a series of complex steps to extract useful information from images. Before any processing and analysis, images must first be acquired. The spatial distributions of physical matter are converted into images that represent these distributions. These images can be stored on film or printouts, or more commonly nowadays, in the form of digital data that can be digitally processed.

In this thesis, we will focus on processing images obtained by fluorescence microscopy, a key technique in molecular, cellular, and organismal biology as well as clinical diagnostics. Fluorescence microscopy allows for the selective assessment of molecular species of interest by labeling them with fluorescent markers, capturing their spatial and temporal distribution. The resulting image data typically consist of numerous 2D matrices containing numerical values that indicate the local concentration of labeled material. These digital data are analyzed to uncover organizational patterns and test specific hypotheses quantitatively. The process involves technical tasks that require the application of mathematical and data science concepts, often performed using specialized software tools. These tasks include controlling microscope hardware, transferring and storing data, processing images, filtering, detecting, and classifying objects. Integrating machine learning methods, particularly deep learning, has become essential in enhancing these tasks, especially in image processing and analysis [28].

The next section will discuss the various tasks involved in analyzing biological images and the commonly employed pipeline models used in practical applications. These tasks entail several steps, from capturing images to extracting features and analyzing the data obtained.

1.2. Understanding the pipeline: main tasks in biological image acquisition and analysis

Biological image analysis involves several sequential steps for extracting valuable insights from microscopy data, and while this is a general pipeline model, the steps and algorithms used can vary depending on the specific biological question being studied and the imaging technique used. However, the recent advancements in the field of ML can assist in all these steps. For example, ML can potentially optimize image acquisition, enhance image quality through preprocessing, automate segmentation, and even improve resolution.

1.2.1. Image acquisition

The process begins with image acquisition, where several imaging techniques, such as fluorescence microscopy, electron microscopy, confocal microscopy, etc., are used based on the type of the biological specimen and the experimental requirements [122].

Image acquisition can be classified from different perspectives. One key distinction is between time-lapse imaging and single-time point imaging. Time-lapse imaging involves capturing a sequence of images over a period of time, which is particularly useful for studying live specimens and observing dynamic processes. In contrast, single time point imaging captures a single snapshot at a specific moment, often used for fixed specimens where dynamic changes are not a concern [51].

One of the most critical parameters in time-lapse imaging is exposure time, which refers to the duration the camera or detector's sensor is exposed to light. Longer exposure times allow more light to be captured, resulting in higher quality images with an increased Signal-to-Noise Ratio (SNR). However, longer exposure also increases the risk of overexposure, where the image becomes too bright and loses detail and photodamage, which can harm the sample due to prolonged light exposure.

The optimal exposure time depends on several factors, including the type of sample and the specific dyes or markers used to stain or label the sample. These dyes or markers have varying brightness and stability under light, which affects the required exposure time for clear imaging. Therefore, careful optimization of exposure time is essential for each experiment to balance image quality and sample integrity. Images obtained through live imaging typically have low SNR. Often, low SNR is encountered as a problem in live imaging because the need to minimize photodamage and overexposure leads to shorter exposure times, resulting in less light being captured and thus a lower SNR.

Fixed imaging is more permissive for longer exposure times because the sample does not move, allowing for extended exposure without motion blur. This stability enables more light capture, increasing the SNR and resulting in higher-quality images. Higher quality is needed for cellular structures to ensure detailed visualization and accurate analysis of cellular components. Single time point acquisition is often used to study these structures or analyze the expression of specific proteins or genes [90].

1.2.2. Pre-processing

Microscopy images often suffer from noise and variations in brightness and contrast due to limitations in the imaging system and sample preparation techniques. Noise can obscure important details, making it challenging to identify and analyze relevant features.

To mitigate these issues, various pre-processing techniques are employed. Background subtraction helps in removing unwanted noise from the image. Deep denoising algorithms can reduce random noise, enhancing the SNR. Contrast enhancement methods highlight relevant features, making them more discernible for further analysis. These pre-processing steps collectively aim to improve the accuracy and reliability of the subsequent analysis by providing a clearer and more accurate representation of the sample [79].

1.2.3. Object detection

Segmentation is one way to detect objects, involving dividing images into distinct regions or objects of interest. Various algorithms are available, ranging from conventional threshold-based methods to advanced ML and DL techniques. Threshold-based methods are simple and efficient but may struggle with images with varying intensities or noise. In contrast, DL-based techniques provide greater accuracy by leveraging large datasets and computational power to learn the characteristics of biological structures. These techniques can handle variability in image data and often yield more robust segmentation results. Segmentation accuracy is essential, as subsequent analysis relies on segmented objects. [89].

1.2.4. Feature extraction

Various types of features are considered in this step:

Intensity-based features: These features analyze the pixel intensity values within the segmented object. Common intensity-based features include mean intensity, standard deviation, and the intensity histogram. These features help in understanding the brightness and contrast variations within the object.

Shape-based features: Shape-based features describe the geometric properties of the segmented object. Key features include perimeter, area, eccentricity, circularity, and aspect ratio. These features are crucial for distinguishing objects based on their geometric characteristics and can be used to identify and categorize different shapes within the sample.

Texture-based features: Texture-based features capture patterns and variations in the pixel intensities that represent the object's surface characteristics. Common texture features include measures of surface roughness, smoothness, and granularity.

1.2.5. Analysis of the extracted features

This analysis can involve several approaches, depending on the specific goals.

For example, features can be used to monitor dynamic changes in sample characteristics, such as alterations in size, shape, texture, intensity, or spatial distribution of the segmented objects over time. This is particularly useful in longitudinal studies where tracking the evolution of these characteristics can provide insights into biological processes and disease progression [45].

Additionally, the extracted features can be employed to develop predictive models. ML algorithms, such as support vector machines, random forests, or neural networks, are often used to build models that can predict outcomes based on the features extracted from the data [97]. These models can then be applied to new, unseen data to make predictions [57].

Furthermore, statistical analysis of the features can reveal significant correlations and patterns within the data. Techniques such as principal component analysis (PCA) and cross-correlation analysis can be used to identify the underlying trends and relationships that might not be immediately apparent, thereby enhancing our understanding of the biological system under study [80, 182].

1.3. Applications of ML in biological image analysis

This section aims to objectively explore the role of ML in enhancing these stages.

1.3.1. Supervised ML methods

Supervised ML is a type of ML where the algorithm is trained on a labeled dataset. This means that each input image in the training set is paired with the correct output label, which could be the identification of biological structures or features of interest. During training, the algorithm learns to make predictions based on the input data by adjusting its parameters to minimize the difference between its predictions and the actual labels. This process involves a loss function, quantifying the error, and an optimization algorithm, which iteratively updates the model to reduce this error. Supervised ML methods rely heavily on labeled datasets. A few examples of the applications of supervised ML methods in biological image analysis include [109]:

Image restoration (image denoising): Microscopy images can suffer from various factors that negatively impact their quality, such as blurring, uneven illumination, artifacts, and noise. This can lead to inaccurate analysis of the images. Image restoration techniques are used to address this issue. For example, the state-of-the-art denoising method CARE (Content-Aware Image Restoration), based on supervised DL introduced by Weigert et al. [176], improves microscopy image quality by training a network on pairs of noisy and high-quality images. Once the network is properly trained, it can restore new, unseen images by applying what it learned during training. The network processes the noisy input image and outputs a restored image, effectively reconstructing the high-quality version by removing noise and correcting distortions. This reconstruction process involves applying the learned weights and biases from the training phase to the new data, enhancing image quality.

Image segmentation: Supervised image segmentation uses annotated training data to locate and extract specific regions of interest within an image. This approach involves an expert manually segmenting training images to create a ground truth dataset. The resulting dataset is then used to train a ML algorithm to segment new images automatically.

The U-Net architecture is a convolutional neural network designed specifically for biomedical image segmentation. It consists of a contracting path to capture context and a symmetric expanding path that enables precise localization, making it highly effective for segmenting cells and tissues in microscopy images [146].

The Region-based Convolutional Neural Network (R-CNN) is another popular object detection and segmentation architecture. It generates region proposals and then classifies each region, making it suitable for identifying and segmenting multiple objects within an image [105]. R-CNN has been widely adopted for various applications because it can handle complex scenes with multiple objects.

The Segment Anything Model (SAM) [84] is a recent innovation that aims to provide a versatile and generalized segmentation solution across diverse image types. SAM is designed to be highly flexible and capable of segmenting any object in any image without requiring extensive domain-specific training. However, while SAM's generalization capabilities are impressive, they might not be ideal for microscopy images. Microscopy images often contain very specific details requiring specialized knowledge to accurately segment. Studies have shown that while SAM performs well on general image segmentation tasks, its performance on medical and microscopy images does not match the specialized models like U-Net or R-CNN, primarily due to its generalist nature and lack of domain-specific training [117].

1.3.2. Unsupervised ML methods

Unsupervised ML methods have many applications in biological image analysis. Here are some examples used in image restoration and image segmentation:

Image restoration / Image denoising: Various unsupervised techniques are employed for denoising microscopy images, including conventional and DL approaches. Conventional denoising methods include filters like median filters, Gaussian filters, and wavelet denoising [16], which are widely used for their simplicity and effectiveness in reducing noise. On the other hand, several unsupervised DL techniques have been developed specifically for denoising microscopy images without need of the ground truth, such as Noise2Void [91], Noise2Self [10], and Noise2Atom [171]. These unsupervised DL approaches, which are called one-shot denoising, directly learn the noise characteristics from noisy input images, eliminating the need for clean reference images. These approaches are noteworthy in microscopy image analysis because they can operate on a single noisy image, making them highly applicable in scenarios where obtaining clean images is challenging.

Image segmentation: Unsupervised segmentation in microscopy refers to automatically separating objects or regions of interest in microscopic images without manual annotations or predefined labels. These unsupervised methods mainly employ conventional ML techniques. One standard method is thresholding [39, 119], which involves setting a threshold value and separating pixels based on their intensity values. Initially, pixels are classified as above or below the threshold, which helps detect potential objects. After this initial detection, further processing is done to group these pixels into distinct regions or objects.

Another method is clustering-based segmentation [169], where similar pixels are grouped into clusters, defining regions within an image. This technique involves algorithms like k-means or Gaussian Mixture Models (GMM) to identify clusters of pixels with similar characteristics, effectively segmenting the image into meaningful regions.

Additionally, edge-based segmentation methods [53], such as the Canny edge detector, can identify edges or boundaries between different structures within an image. This method focuses on detecting the contours of objects by identifying significant changes in pixel intensity, which helps outline the distinct regions within the image.

It is important to note that DL-based segmentation methods, while highly effective, typically rely on supervised learning, which requires large amounts of labeled data for training [146]. As such, they do not fall under the category of unsupervised segmentation methods discussed here. Although techniques like transfer learning can reduce the amount of labeled data required by leveraging pre-trained models on diverse datasets and fine-tuning them with smaller labeled microscopy images [167], they still require some labeled data. Therefore, these deep learning approaches remain within the domain of supervised learning and are not applicable in the context of unsupervised segmentation [44].

1.4. Challenges

As discussed in the previous section, ML methods can potentially help biologists in image processing. However, the widespread use of these techniques is limited due to challenges explained in this section.

1.4.1. General challenges in reusing ML techniques in biological research

Reusing ML techniques in biological research can be approached through conceptual, adaptation, and deployment reuse, as outlined in [38]. Each of these approaches presents unique challenges that can benefit from a software engineering (SE) perspective.

- Conceptual reuse: Conceptual reuse means using ideas, algorithms, or designs from existing ML models described in research papers or reports and applying them to new projects. From a SE perspective, this kind of reuse needs good documentation, standard formats, and modular designs to help biologists who may not have strong programming skills. Without clear design patterns or reusable software tools, it can be hard for biologists to turn complex mathematical ideas into useful tools for their work. Also, the lack of compatible frameworks makes it harder to use these ideas in research processes [38].
- Adaptation reuse: Adaptation reuse means adjusting pre-trained models to work on new tasks, often using transfer learning. One challenge from a SE perspective is making sure pre-trained models are built in a way that makes them easy to modify and extend. Without clear design and interfaces, it can be very difficult to adjust complex DL architectures, settings, or fine-tuning methods. Another big challenge is the lack

of labeled data in biology. This makes it important to have strong data processing systems and tools that can help make the adaptation process easier and faster [38].

- Deployment reuse: Deployment reuse refers to the operationalization of pre-trained models in new environments. One challenge is ensuring the reliability of the origins of these models, including where they were developed, the datasets used for training, and the algorithms applied. Biologists often worry about errors or biases that might come from how the model was originally trained, especially if they don't fully understand the algorithms or training data. They also may not have the tools or processes to ensure the models are reproducible, free of bias, and work well with different software or hardware setups. Using SE practices in the deployment process can help solve these problems and make the models more reliable for real-world use [38].

Despite these general challenges in reusing the ML models, some specific challenges are discussed as follows.

1.4.2. Application of supervised methods

Supervised methods, though developed for microscopy image analysis, present several challenges. Even pre-trained models often require fine-tuning to adapt to new data, which can be complex [54].

Apart from the technical challenges for biologists in reusing these models, the scarcity of high-quality labeled datasets and the variability in microscopy images present major limitations. Creating accurate annotations is labor-intensive and requires expert knowledge. At the same time, differences in lighting, staining techniques, and other factors make it difficult for models trained on specific datasets to generalize well to new conditions [167, 108, 146, 110].

Interpreting microscopic images often necessitates domain expertise, and the subjective nature of annotations can introduce inter-observer variability, leading to inconsistencies in labeled datasets [159]. Given these challenges, our proposed approach focuses exclusively on unsupervised/weakly-supervised methods. These techniques bypass the need for large, labeled datasets and address the challenges of variability and complexity in microscopy images without relying on supervised learning. By doing so, we seek to develop solutions that are more adaptable and accessible to users without the burdens of manual annotation and model fine-tuning.

1.4.3. Application of unsupervised methods

Unsupervised or weakly-supervised methods are particularly useful when high-quality labeled datasets are scarce. In such cases, they are preferred over supervised approaches, which require annotated data for training. Unsupervised microscopy image analysis methods are mainly based on conventional ML techniques and statistical methods, such as k-means clustering and watershed segmentation. These methods rely on manually engineered features, making them suitable for datasets that don't require advanced feature engineering [39, 169].

Variability in biological images Conventional unsupervised methods, while computationally efficient, face significant challenges when dealing with the diversity of biological images. Variations in microscope parameters, sample preparation protocols, and structural variations in biological images introduce noise and inconsistencies. These variations make it difficult for conventional ML methods to categorize relevant patterns accurately across different datasets [110]. The inherent diversity in biological data complicates the development of robust and generalizable models for different conditions.

Amplification of noise in unsupervised DL On the other hand, unsupervised DL approaches offer the advantage of automatic feature learning without manual feature engineering. However, the absence of labeled data for quality assurance introduces new challenges. One key issue is the amplification of noise or distortions during the deep learning process, potentially resulting in artifacts [146]. These artifacts can compromise the reliability of the analysis and require careful quality assurance mechanisms. Without proper checks, the unsupervised nature of DL can lead to unreliable results, especially when applied to complex and noisy microscopy datasets.

Complexity of reuse for biologists While conventional unsupervised methods are more accessible for biologists due to their reliance on well-established algorithms, reusing unsupervised DL models in microscopy image analysis is more difficult. One key issue is the lack of clear guidelines or imaging protocols to assist biologists in effectively applying these DL and ML approaches. For example, when using models like Noise2Void (N2V), biologists need detailed instructions on properly recording images and understanding how this impacts other workflow steps. Without such guidelines, it is difficult for biologists to integrate these tools into their pipelines with confidence [167].

Selecting and fine-tuning ML approaches In addition to the lack of guidelines, biologists face the challenge of selecting the most appropriate ML approach for their specific data. The wide range of ML models available can be overwhelming, especially since many models are designed for different data types and purposes. Even after selecting a model, reusing and fine-tuning it to suit unique datasets often requires a deep understanding of both the model and the underlying biological data. For researchers without advanced ML expertise, this process of selection, reuse, and fine-tuning adds another layer of complexity to their research workflows.

1.4.4. Challenges of light microscopy that can be addressed by ML methods

Light microscopy is widely used in areas like cell biology, neuroscience, materials science, and engineering [163]. One important application is visualizing fluorescently labeled molecules, but this comes with challenges when designing experiments.

Trade-off between acquisition parameters and image quality A major challenge in light microscopy is managing the trade-offs between exposure time, acquisition speed, and SNR. When trying to improve one aspect, it often affects another negatively. This makes it hard to get high-quality images without damaging the sample or slowing down the imaging process [136, 152].

Potential for artifacts in reconstructed images Improving image quality post-acquisition through computational methods is one way to recover the SNR. Traditional methods, like projection techniques [156], deconvolution filters [143, 138], and denoising methods [19, 56], help to a certain extent but often struggle with very noisy, low-SNR images. DL methods have recently shown better performance in these tasks. Supervised DL methods need large, high-quality datasets to train the models, which are not always available. These datasets are often hard to create because getting high-quality images can be difficult or impossible. While newer methods like Noise2Noise (N2N) [100, 101] and N2V) [91] reduce the need for high-quality data, but they introduced a new challenge, which is the possibility of creating artifacts in the reconstructed images. These artifacts are errors that show up when there is no high-quality image to compare with the processed one [13, 120]. The absence of a ground truth image to guarantee quality can undermine trust in image analysis, particularly in unsupervised approaches.

1.5. Research objectives

Many ML techniques hold promise for solving challenges in bioimaging pipelines, yet reusing them seamlessly into routine tasks raises obstacles. This research aims to achieve two primary objectives: 1) reusing unsupervised deep denoising with quality assurance into the image processing pipeline and 2) addressing the reusing challenges of ML methods in segmentation tasks.

Unsupervised DL denoising tools that operate without ground-truth data have the potential to tackle the challenge of needing annotations. When used effectively, these tools can speed up image capture and enhance signal quality after acquisition. However, it is essential to ensure that the processed images are of high quality and free from artifacts. Achieving this requires a thorough and well-designed imaging protocol, which can support biologists reusing deep denoising with quality assurance.

Regarding segmentation challenges, the focus is on selecting appropriate ML algorithms, reusing them and optimizing the relevant parameters without relying on annotations while addressing the diversity in biological data. These choices influence segmentation accuracy and pose a challenge for biologists without a computer science background. A comprehensive solution is needed to automatically choose and adapt algorithms to the specific characteristics of the image data with minimal user input and no annotations.

1.5.1. Research goal and questions

Research Goal
Our goal is to develop an approach that allows biologists to seamlessly integrate advanced ML and DL techniques into their bioimaging workflows with built-in quality assurance, all without the need for manual annotation. This approach will empower biologists, regardless of their technical expertise, to effectively reuse existing ML and DL methods, enhancing their research capabilities and enabling them to achieve more accurate results. Ultimately, this will contribute to significant advancements in the field of biology

Providing protocols to reuse ML and DL and simplify the selection and fine-tuning of parameters and hyperparameters without annotations offers two benefits:

- Access to ML techniques for non-experts: SE practices such as automation and usability-focused design will ensure that biologists with limited programming expertise can use these tools effectively, lowering technical barriers and increasing accessibility.
- Streamlined optimization processes with quality assurance: The research will develop modular and flexible workflows, incorporating quality assurance—a critical aspect of SE—to ensure reliability and reproducibility. By embedding quality assurance mechanisms, the processed data will be dependable, enabling biologists to trust the results and confidently use them in their research.

The research aims to address challenges in two stages of the biological image processing pipeline. The first stage focuses on image acquisition and denoising, while the second stage targets image segmentation and analysis. The solutions will integrate practical SE principles, such as adaptability and modularity, to ensure that ML techniques can be reused effectively for diverse bioimaging tasks.

Reusing quality-assured DL denoising

Conducting experiments in light microscopy requires balancing rapid data acquisition, maintaining high SNR, and avoiding specimen damage. Reducing exposure time speeds up acquisition but can lower SNR, compromising image quality. While unsupervised DL denoising techniques can address the need for annotation and help restore SNR during post-acquisition processing, they can potentially introduce errors in the reconstructed images.

This presents a dilemma: DL networks can only be practically applied without ground truth data, yet ground truth data is essential to ensure error-free reconstruction. Therefore, determining the optimal exposure time to increase imaging speed, optimize DL denoising performance, and address concerns about artifact-free images is challenging.

To guide our investigation, we have formulated the following research questions:

- How can biologists effectively reuse DL denoising techniques in the image processing pipeline to improve the SNR of light microscopy images? (RQ 1.1)
- What methods can be employed to assure the quality and reliability of DL-based fluorescence image reconstruction when ground truth data is unavailable? (RQ 1.2)
- What methods can suggest the ideal exposure time for the DL denoising-supported experiment? (RQ 1.3)

By addressing these questions, we aim to reuse DL denoising techniques into light microscopy, assuring accurate and reliable image reconstruction without ground truth data.

Ease of reusing segmentation

Obtaining annotations for bioimaging data can be expensive or impossible. Annotation-free ML algorithms are more practical for such projects. Unsupervised conventional ML approaches do not require annotations and often involve a separate feature extraction phase, making them suitable for datasets that do not need advanced feature engineering.

However, a one-size-fits-all approach may not suit all object types in bioimaging data. Developing effective approaches for selecting, reusing, and parameter adjustment ensures optimal performance across diverse datasets. This part of our research addresses these challenges by offering practical solutions for selecting and reusing ML segmentation techniques for the specific needs and characteristics of biological imaging data. Our research questions for this phase are:

- What methods can identify the most suitable annotation-free segmentation ML algorithms for the given data? (RQ 2.1)
- How can we employ strategies to reuse and fine-tune the parameter configurations of the selected ML algorithms to achieve precise and efficient segmentation outcomes? (RQ 2.2)
- How can we develop adaptive parameter adjustment techniques to ensure optimal performance across diverse object types? (RQ 2.3)

We aim to create dependable and flexible techniques for segmenting bioimaging data without relying on annotations. This will result in more effective and accessible image analysis for biological research and medical applications.

1.5.2. Contributions

Modern scientific software, particularly in domains such as biological image analysis, is highly diverse and often requires specialized solutions tailored to the unique demands of the field. Previous studies, such as those by Wilhelm Hasselbring [67, 48], emphasize that research software must balance generalizability with the specific needs of its domain. Our work builds on this understanding by contributing to SE for AI, specifically addressing the challenge of reusing ML models for domain-specific applications. Below are the key ideas and contributions for each research question:

Pragmatic imaging protocol for quality-assured deep denoising in live imaging studies

In this phase, we propose a comprehensive imaging protocol to ensure the quality-controlled reuse of deep denoising techniques in high-speed fluorescence microscopy [66]. Building on ideas often used in SE, the protocol emphasizes systematic and reproducible practices to address challenges in image reconstruction. This protocol specifically targets annotation-free DL denoising techniques and encompasses several critical steps:

- Capturing a small dataset of high-quality reference images for each object of interest.
- Optimizing exposure times to balance imaging speed and SNR.
- Assuring reconstruction quality to ensure reliable image recovery during full-time-lapse acquisition.

This imaging protocol adopts a SE-inspired approach to tackle key challenges: 1) the complexity of the reuse for biologists, 2) the trade-off between acquisition parameters and image quality, and 3) potential artifacts in reconstructed images.

We applied this imaging protocol to analyze fluctuations in molecular clusters within live zebrafish embryos. Our real-life case study demonstrated the approach's effectiveness, achieving a fivefold increase in imaging speed and unveiling a novel biological process [66]. This process highlighted a close coordination between post-translational modifications of RNA polymerase II (Pol II) and concurrent 3D shape changes in these clusters over a brief time frame.

By integrating SE principles such as quality assurance, this protocol offers general guidance for microscopists reusing ground-truth-free image reconstruction methods. It ensures the accuracy and reliability of live imaging analyses and is particularly valuable in scenarios requiring high-speed data acquisition to capture dynamic biological processes.

Framework for weakly supervised reuse of ML with interactive user input

Our research introduces a reusing framework [65, 64] for weakly supervised ML segmentation, integrating interactive user input to refine segmentation algorithms with minimal user intervention. This framework is particularly beneficial in scenarios where training data and objective functions are unavailable, yet a quality-approximating metric is definable.

This work contributes to SE for AI, specifically in the domain of biological image analysis, by addressing the challenge of reusing ML solutions for highly specialized tasks. The framework adopts a systematic approach to streamline the selection and optimization of ML segmentation solutions through a two-step process:

1. Searching for existing annotation-free segmentation solutions and optimizing parameter settings with limited user input.
2. Reducing user input requirements simplifies the algorithm's reuse in the imaging pipeline.

SE practices, such as modularity, quality assurance, and iterative refinement, are reflected in the framework's design. By correlating object-morphology feature space with specific segmentation algorithms and parameters, our approach enables biologists to discover, implement, and adapt segmentation methods tailored to their data variations. This approach simplifies reusing existing solutions for diverse biological data, enhancing usability and efficiency. The proposed framework will address the challenges of 1) variability in biological images, 2) complexity of the reuse for biologists, and 3) selecting and fine-tuning ML approaches.

We demonstrated the effectiveness of our framework through the example case of RNA Polymerase II (Pol II) segmentation in 2D super-resolution microscopy images. Given the highly variable morphologies and diffuse boundaries of these clusters, this case serves as an ideal model to showcase the versatility and robustness of our approach.

1.5.3. Expected benefits

The expected benefits of our contributions are as follows:

1. Enhanced image reconstruction quality: The quality-controlled deep denoising-supported live imaging protocol improves the reconstruction quality of high-speed fluorescence microscopy data. By enabling the reuse of DL-based denoising methods in new experiments, researchers can achieve more accurate and reliable results. This is achieved by recording high-quality data for each object and focusing on compromised data for time-lapse acquisition.
2. Insights into molecular clusters: The application of the protocol to analyze fluctuations in molecular clusters within live zebrafish embryos provides valuable insights into the coordination between post-translational modifications of RNA polymerase II (Pol

II) and changes in the 3D shape of these clusters. Such discoveries can deepen our understanding of dynamic biological processes and molecular interactions.

3. Guidance for microscopists: The presented quality-controlled image reconstruction method offers guidance to other microscopists working on live imaging studies. By adopting this approach and reusing DL denoising methods without requiring ground-truth annotations, researchers can ensure the quality of their denoising processes, leading to more reliable results.
4. Efficient and user-friendly ML segmentation: The proposed framework for weakly supervised ML with interactive user input streamlines the selection and reuse of ML segmentation solutions. It allows scientists to utilize existing annotation-free segmentation techniques and optimize parameters with minimal user interactions, making it more accessible and practical for researchers with varying levels of ML expertise.
5. Time and resource savings: With fewer user interactions, biologists can save valuable time and resources in implementing ML segmentation solutions for their bioimaging data. The framework's ability to find and adapt segmentation methods without annotations accelerates the analysis process.
6. Customized segmentation algorithms: The framework enables researchers to fine-tune segmentation algorithms for specific object types in their biological imaging data. This customization ensures that the algorithms are optimized for accurately capturing the unique characteristics of various biological structures, leading to more precise and meaningful results.
7. Wider application of ML in bioimaging: By offering a user-friendly and practical approach to applying ML algorithms in bioimaging, our contributions encourage more biologists to embrace ML techniques in their research. This broader adoption can lead to more sophisticated and automated image analysis methods, advancing the field of bioimaging and facilitating discoveries.

1.6. Thesis outline

The thesis is structured as follows:

Chapter 2, also called Foundation, establishes the foundational knowledge required to understand the report. This chapter covers microscopy image analysis, machine learning techniques, and concepts such as image denoising, segmentation, and active learning strategies.

Chapter 3 is the first part of the author's contribution, called "Speeding up live microscopy with DL denoising." This chapter discusses a novel approach that enhances live microscopy imaging using DL denoising techniques. It includes the motivation behind the study, the proposed methodology, evaluation, discussions, and related work.

Chapter 4, "Active learning to define segmentation of unknown objects," is the second part of the author’s contribution. This chapter presents a novel framework that uses active learning strategies for segmenting unknown objects in biological images. It covers the introduction, scientific use case, proposed approach, evaluation, discussions, and related work.

The epilogue contains the conclusion, summary, and future work. The conclusion summarizes the key findings of the thesis, discusses implications, and suggests avenues for future research.

2. Foundation

The Foundation chapter provides a comprehensive overview of essential concepts and methodologies related to microscopy image analysis and machine learning, which are necessary to understand this thesis. It begins by exploring the fundamental challenges and noise models inherent in microscopy images and then discusses various imaging modalities. The chapter then examines the critical balance between acquisition speed, image quality, and photodamage. It also introduces the concept of SNR and its significance in imaging.

Afterward, the chapter moves on to the basics of machine learning and deep neural networks. It discusses different machine learning techniques, the structure and function of neural networks, and the fundamental principles of deep learning. The chapter also explores the practical application of these concepts in microscopy image denoising. It emphasizes both deep learning-based denoising techniques and the metrics used to evaluate denoising performance.

The chapter further explores microscopy image segmentation, detailing thresholding-based methods, watershed segmentation, and deep segmentation approaches. Finally, it addresses active learning and different strategies, particularly focusing on active learning query strategies to enhance the efficiency and effectiveness of the learning process. Each section is designed to provide the reader with a solid grounding in these foundational topics.

2.1. Foundations of microscopy image analysis

This section introduces the essential aspects of microscopy image analysis. We will discuss key factors such as contrast, noise, and the availability of annotated data, which influence the quality and interpretability of microscopy images.

2.1.1. Understanding the nature of microscopy images

Analyzing microscopy images can be difficult due to several factors:

Noise and low SNRs: Microscopy images often pose challenges in segmentation due to low contrast, noise, artifacts, and varying intensity. The low contrast in these images complicates feature identification, which results from the sample's inherent properties, such as low optical density or refractive index variations, and limitations in the imaging system's dynamic range [153]. This insufficient contrast hinders accurate segmentation and quantitative analysis. Additionally, various types of noise and unwanted artifacts can arise from imaging technology limitations. Noise sources include photon shot noise, electronic noise from the camera sensor, and environmental factors like temperature fluctuations. Artifacts may originate from optical aberrations, specimen preparation techniques, image acquisition settings, and image processing techniques, impacting image fidelity and interoperability. Furthermore, in fluorescence microscopy, fluorescent chemical compounds are used to label biological structures, and the signal intensity can vary across the image due to uneven staining or photobleaching. Fluorophores can exhibit distribution variations because of inconsistent labeling or sample preparation and in excitation light intensity due to uneven illumination [61]. These variations lead to spatial changes in signal intensity, complicating quantitative analysis and making it challenging to distinguish true biological signals from background noise or artifacts.

Multi modality nature of microscopy images Data diversity is crucial for developing robust microscopy analysis models as it at least encompasses diversity from 3 channels: cell origins, staining methods, and cell morphologies.

The source of cells in microscopy images can differ significantly, arising from a variety of tissues or existing within cell cultures that are subject to different conditions. This leads to substantial variation, as cells in tissues are typically closely packed and arranged in a specific manner, while cells in culture are often spread out and positioned randomly. Secondly, the selection of staining techniques, such as using specific antibodies in fluorescent microscopy, adds to the variety by highlighting different cellular structures or proteins. Cell morphologies vary significantly across different cell types. While most cells are round, some may have elongated or irregular shapes. In microscopy images, structures often have complex and irregular shapes, making analysis more challenging. This irregularity may stem from biological differences, such as cell morphology or architecture, as well as technical factors, like specimen preparation techniques or imaging conditions. Accurate analysis and feature extraction require advanced algorithms capable of handling shape variations and boundary ambiguities [112].

Object edges in microscopy images might appear blurry or indistinct, especially in low-resolution images or those using specific staining techniques. Diffraction effects, optical aberrations, or limitations in the imaging system's spatial resolution can blur object boundaries, complicating accurate delineation. Ambiguous boundaries hinder quantitative analysis and lead to inaccuracies in measuring object size, shape, or spatial distribution [112].

Limited annotated data The success of DL models hinges on the similarity between the training data and the user's own data. While the exact definition of 'sufficient similarity' varies depending on the methods and architectural choices of the model, it is clear that creating foundational models capable of analyzing a wide array of biological images necessitates a diverse and comprehensive training dataset. This dataset should encompass a broad range of microscope modalities, various imaging conditions, and numerous distinct cell types to ensure the model can be generalized effectively.

However, creating accurate annotations for microscopy images is both time-consuming and requires specialized expertise. This scarcity of annotated data complicates the analysis and validation of image processing algorithms. The manual annotation involves labor-intensive tasks such as outlining cell boundaries or labeling subcellular structures. Furthermore, the availability of annotated datasets for specific research questions is often limited, posing significant challenges for developing and evaluating algorithms. This limitation underscores the need for innovative approaches to generate and utilize annotated data more efficiently, thus enhancing the overall effectiveness of DL models in biological image segmentation.[108, 43].

2.1.2. Noise models in microscopy images

Understanding noise in microscopy is crucial for accurately interpreting captured images. This section explores key characteristics of noise in fluorescence microscopy images:

1. **Random nature of noise** At the pixel level, noise appears as randomly generated positive or negative values, adding variability to the pixel's 'true value.' This randomness can arise from various sources, including photon emission and detection processes, electronic noise in the imaging sensor, and environmental factors like temperature fluctuations.
2. **Pixel-independent noise** The noise at each pixel is unaffected by its location or the noise at other pixels, emphasizing its independence. This independence allows noise to be modeled and treated separately at each pixel during image processing, simplifying noise reduction techniques.
3. **Specific noise distribution** Each noise value follows a distinct distribution, and their collective histogram reflects the underlying pattern. Common noise distributions in microscopy include Poisson, Gaussian, and uniform distributions, representing different sources and characteristics of noise.

Digital imaging primarily means noise sources such as dark noise, photon noise (shot noise), and readout noise [162, 155]:

- **Dark noise** Arising from thermal agitation of electrons, dark noise occurs even without incident photons and is often represented by a Poisson distribution, $P(\lambda_d)$, where λ_d represents the average dark flux. Dark noise varies with exposure time, temperature, and sensor characteristics, significantly affecting low-light imaging conditions.

- **Photon noise** Resulting from fluctuations in detected photons during exposure, photon noise introduces stochasticity and is described by a Poisson distribution with an average photon flux λ_p. Photon noise is inherent in all optical signals and can dominate the noise budget in low-light microscopy, affecting image quality and signal-to-noise ratio.
- **Readout noise** Caused by imperfections in the output amplifier, readout noise is modeled as an additive component with a zero-mean Normal distribution. Readout noise arises from electronic components in the imaging sensor and readout circuitry, contributing to the overall noise level in the captured image.

The noise level is influenced by exposure time, temperature, and the fluorescence properties of imaged structures. Longer exposure times increase photon counts and may exacerbate photon noise, while higher temperatures can increase electronic noise levels, impacting overall image quality [30].

Microscopic image denoising relies on three main models:

1. **Additive white Gaussian noise (AWGN)** This model treats readout noise as an additive component with a zero-mean Gaussian distribution, commonly used in image processing literature. Mathematically, the AWGN model can be expressed as $y = x + n$, where y is the observed image corrupted by noise, x is the true image, and n is the AWGN with zero mean ($E[n] = 0$) and constant variance ($Var[n] = \sigma^2$). The AWGN model simplifies noise modeling and allows for applying efficient denoising algorithms [30].
2. **Poisson noise** Commonly used to represent photon noise in fluorescence microscopy, this model assumes photon counts at each pixel follow a Poisson distribution, suitable for low-light conditions. Mathematically, the Poisson noise model can be formulated as $y = \text{Poisson}(x)$, where y denotes the observed image and $\text{Poisson}(x)$ represents the Poisson distribution with parameter x as the true image [30].
3. **Mixed Poisson-Gaussian (MPG) noise** This realistic model combines Poisson and Gaussian noise components, considering dark noise, shot noise, and readout noise simultaneously. Mathematically, the MPG noise model is represented as $y(s) = \gamma P(\lambda_s) + n$, where γ is a gain constant, P denotes a Poisson random process, and n represents a zero-mean Gaussian random variable [61, 30].

2.1.3. Imaging protocol

An imaging protocol is a step-by-step guide to help researchers get clear and accurate images of biological samples that have been labeled with fluorescent dyes or antibodies. These dyes attach to specific proteins or molecules in the sample and glow when exposed to certain light, allowing scientists to see where these molecules are located in the cells or tissues.

The imaging protocol starts with preparing the sample, including fixing the cells or tissues to keep their structure intact and then staining them with the right amount of fluorescent markers. It also explains how to adjust the microscope settings, like choosing the right light for the dyes, setting the magnification, focusing, and adjusting the exposure time to capture bright and clear images without too much background noise.

After capturing the images, the protocol may provide guidelines for post-processing, such as enhancing image contrast, removing background noise, or using software tools for quantitative analysis.

2.1.4. Imaging modalities

Biological research employs various imaging modalities to study specimens at different levels of detail and under different conditions. The common approaches in microscopy include fixed imaging, live imaging, and several advanced microscopy techniques. Fixed imaging means examining chemically preserved or "fixed" specimens at a specific point in time, while live imaging entails observing and recording biological specimens in real time as they undergo dynamic processes. Advanced microscopy techniques encompass a broad range of methods, including light microscopy, electron microscopy, fluorescence microscopy, confocal microscopy, super-resolution microscopy, and atomic force microscopy. Each technique offers unique capabilities and advantages for visualizing biological samples and studying cellular structures and dynamics.

2.1.4.1. Fixed Imaging

Fixed imaging means the examination of specimens that have been chemically preserved or "fixed" at a specific time. This preservation process, achieved through applying chemicals like formaldehyde, halts ongoing biological processes, stabilizes cellular structures, and prevents decay. By capturing the unchanging or stabilized condition of cells and samples, scientists can use microscopes to study these specimens. Fixed specimens can be stored for later analysis, retaining their structural integrity at the time of fixation. This method is particularly valuable for investigating cell structure. Since the sample is fixed and protected from damage during extended exposure, longer exposure times can capture high-quality images, clearly visualizing the sample's structure [141, 55].

2.1.4.2. Live Imaging

Live imaging [45] means the real-time observation and recording of biological specimens undergoing dynamic processes. Specimens are maintained alive or in conditions closely resembling their natural environment, allowing researchers to monitor and analyze biological activities and changes as they occur. Live imaging provides valuable views into dynamic processes, such as cell division and signaling, which cannot be captured through fixed imaging alone. However, using laser power in live imaging can potentially damage samples, necessitating short exposure times to minimize this risk. Consequently, live imaging often yields lower image quality than fixed imaging, but it offers unique opportunities to study biological phenomena in real-time [45].

2.1.4.3. Microscopy Techniques

Microscopy is a technique used to observe small structures within biological specimens. There are different types of microscopy available, including:

1. **Light Microscopy:** Utilizes visible light to magnify and observe small structures within biological specimens. It includes techniques such as brightfield microscopy, phase-contrast microscopy, differential interference contrast (DIC) microscopy, and fluorescence microscopy [95, 55].
2. **Electron Microscopy (EM):** Uses a beam of accelerated electrons to illuminate a specimen, providing extremely high-resolution images. It includes transmission electron microscopy (TEM) and scanning electron microscopy (SEM) [41, 55].
3. **Confocal Microscopy:** Employs a pinhole aperture to eliminate out-of-focus light, resulting in optical sectioning and improved resolution compared to conventional widefield microscopy. It is particularly useful for imaging thick samples and studying three-dimensional structures within biological specimens [42, 55].
4. **Super-Resolution Microscopy:** This technique surpasses the diffraction limit of light, allowing for the visualization of structures at the nanoscale level. Techniques include structured illumination microscopy (SIM), stimulated emission depletion microscopy (STED), and single-molecule localization microscopy (SMLM), such as stochastic optical reconstruction microscopy (STORM) and photoactivated localization microscopy (PALM) [15, 55].
5. **Atomic Force Microscopy (AFM):** Scans a sharp probe tip over the surface of a specimen, measuring interactions between the tip and the sample surface to generate high-resolution images. It is particularly valuable for imaging biological samples in their native state, including proteins, DNA, and cell membranes, under physiological conditions [25, 55].

2.1.5. Signal-to-noise ratio

SNR is a measure commonly used in science and engineering to assess the strength of the desired input signal relative to unwanted background noise. A higher SNR value indicates a better output, reflecting a favorable ratio of useful information (signal) to undesired data (noise). Calculating SNR can be done using various methods, with the general formula expressed as:

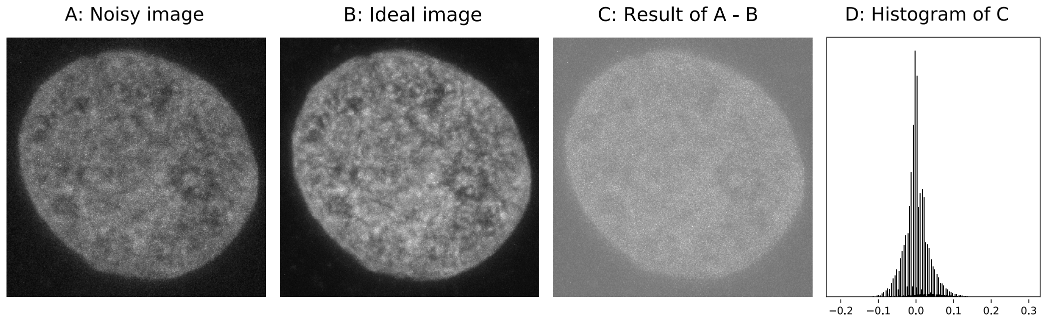
$$\mathrm{SNR} = \frac{P_{\mathrm{Signal}}}{P_{\mathrm{Noise}}}$$

Here, P_{Signal} represents the power of the signal, and P_{Noise} represents the power of the background noise.

Power of the signal (P_{Signal}): For microscopic images, the signal typically represents features of interest that require analysis or study, such as structures within cells, particles, or biological molecules. One common approach to estimating the signal's power is to measure the intensity or brightness of these features. This may mean selecting regions of interest (Region Of Interest (ROI)s) corresponding to these features and measuring their pixel intensities. After obtaining intensity measurements, one can calculate the average or total intensity within the ROIs, representing the strength of the signal in the image.

Power of the noise (P_{Noise}): Noise in microscopic images can stem from various sources, including electronic noise in the imaging system, photon shot noise, and background noise. Estimating noise power can be more challenging than for simple electronic signals. It often means analyzing regions of the image without features of interest (background regions). One common approach is to measure the standard deviation of pixel intensities in these background regions, providing an estimate of the image's variability or noise level.

for a clearer perspective Fig. 2.1 provides a visual representation of noise in imaging. In panel A, a noisy image is presented, highlighting the random variations that typically affect image quality. Panel B shows the ideal, noise-free version of the same image, serving as a reference for comparison. The difference between these two images is calculated and displayed in panel C, where the residual noise is isolated by subtracting the ideal image from the noisy image. This difference map illustrates the characteristics of the noise present. Finally, panel D presents a histogram of the noise values from panel C, revealing a Gaussian (normal) distribution centered around a mean of 0. This indicates that the noise consists of both positive and negative variations, with most values concentrated near zero.



A: Noisy image

B: Ideal image

C: Result of A - B

D: Histogram of C

Figure 2.1.: Visual Representation of noise in Imaging This figure contrasts a captured noisy image (A) with the desired noise-free image (B). The residual 'noise' is calculated by subtracting one from the other (C). The histogram (D) exhibits a normal (Gaussian) distribution, indicating that the noise comprises both positive and negative values, centered around a mean of 0.

2.1.6. Balancing acquisition speed, SNR, and photo damage

Researchers must consider certain limitations when conducting experiments using fluorescent labels [88]. These limitations create a delicate balance between three key factors: acquisition speed, SNR, and avoiding photodamage to the specimen [88]. Unfortunately, optimizing these parameters individually is impossible, leading to a complex trade-off (see Fig. 2.2).

Increasing data acquisition speed often necessitates reducing exposure time, which in turn lowers the SNR. Conversely, enhancing the SNR can be achieved by increasing the intensity of light used to stimulate fluorescence in the sample. However, this enhancement comes with an increased risk of photodamage to the living specimen. Therefore, finding the optimal balance between exposure time and SNR is crucial for obtaining meaningful data while preserving the specimen's well-being.

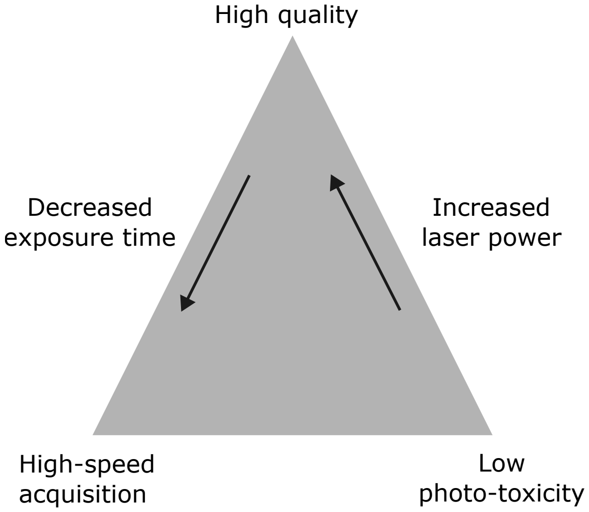


Figure 2.2.: Balancing factors in biological imaging: striking the right equilibrium The illustration represents the balance researchers face during biological imaging experiments. The triangular balance showcases three crucial factors: "High-speed acquisition," "High quality (SNR)," and "Low photo-toxicity." Achieving optimal imaging conditions requires navigating the trade-offs among these factors. At the center of attention is the central symbol, representing the microscope or camera lens, signifying the ultimate goal of obtaining accurate data. Arrows connect each factor to the central symbol, symbolizing the interdependence of these critical elements. On the left, "Decreased exposure time" is depicted, offering the advantage of higher acquisition speed but potentially reducing the SNR. On the right, "Increased laser power" is shown, which enhances SNR but may lead to increased photodamage to the living specimen.

In the context of biological imaging, the trade-off between these parameters is often influenced by the specific requirements of the experiment. For instance, rapid processes such as calcium signaling or neuronal activity necessitate high-speed acquisition to capture fast dynamics, often at the expense of SNR. In contrast, structural imaging, which aims to resolve fine details, benefits from higher SNR, typically achieved through increased exposure time or laser power, risking greater photodamage.

Ultimately, the goal is to tailor the imaging conditions to the specific needs of the experiment, balancing the competing demands of acquisition speed, SNR, and photodamage to obtain high-quality, biologically relevant data.

2.2. Foundations of machine learning and deep neural networks

This section discusses machine learning and deep learning techniques, particularly focusing on the methods and concepts essential for understanding this thesis report.

2.2.1. Machine learning and different techniques

Machine learning is within Artificial Intelligence (AI) area and focuses on creating algorithms and models that empower computers to learn from data and formulate predictions or choices. Rather than being directly coded for a particular job, a ML setup utilizes patterns and information from the data it encounters to enhance its effectiveness progressively.

In general, ML contains the following key concepts:

1. Data: Machine learning algorithms require data to learn from. This data can come in various forms, such as text, images, numerical values, etc.
2. Training: During the training phase, the ML model is presented with labeled data, where the desired outcome or output is known. The model learns the underlying patterns and relationships in the data.
3. Feature extraction: Features are specific characteristics or attributes extracted from the data that the model uses to make predictions or decisions.
4. Model building: Machine learning algorithms construct mathematical representations or models based on the training data. These models capture the patterns and relationships learned from the data.
5. Testing: The model is tested using new, unseen data to evaluate its performance after training. This helps ensure the model can generalize its learning to make accurate predictions on new data.
6. Prediction or decision making: Following training, the model becomes capable of predicting outcomes, categorizing new data, or making choices based on its acquired patterns.

In machine learning, various techniques are employed to train models, each tailored to different scenarios:

Supervised learning: Supervised learning is a technique in which a model is trained on labeled data, meaning that each input is paired with a corresponding output label. By learning from these labeled examples, the model can make accurate predictions or classifications on new, unseen data. In microscopy image analysis, this approach is often used for tasks like cell classification, where the model learns to distinguish between different cell types based on labeled training data. Other examples include segmenting cell structures and detecting specific biomarkers within samples.

Unsupervised learning: Unsupervised learning is a technique that trains models on data without pre-defined output labels. Instead, it concentrates on identifying patterns, associations, or structures inherent in the data. In microscopy image analysis, unsupervised learning techniques are often used to cluster similar cells or tissue types based on their morphological features without knowledge of cell categories. For example, clustering can help identify new subtypes of cells within a heterogeneous sample. Dimensionality reduction techniques are also applied to reduce the complexity of high-dimensional image data while preserving essential information, making it easier to visualize and interpret patterns in cellular structures.

Semi-supervised learning: Semi-supervised learning is a technique that trains a model using both labeled and unlabeled data. In microscopy image analysis, this approach uses the information from many unlabeled images to improve model performance, which is especially useful when labeling data is costly, time-consuming, or requires expert knowledge. For example, in cell segmentation tasks, a limited set of labeled images (where cells are manually outlined) can be combined with a larger set of unlabeled images to enhance the model's ability to identify cell boundaries more accurately. This method is also valuable for detecting rare cell types or variations in cell morphology, where obtaining a comprehensive labeled dataset can be challenging.

Weakly supervised learning: Weakly supervised learning is a technique for training models with data that has coarse labeling or noisy annotations. In microscopy image analysis, this approach is useful when obtaining precise labels for every feature in the image, whether difficult or expensive. For instance, instead of requiring detailed pixel-level annotations of cell structures, a weakly supervised model might learn to classify cell types or detect specific structures using only image-level labels that indicate the presence or absence of a feature. This technique allows the model to use partially labeled or noisy data, making it particularly effective in scenarios where high-resolution annotations are impractical.

Active learning: Active learning is a technique in which the algorithm actively selects the most informative samples to label from a pool of unlabeled data. In microscopy image analysis, this approach helps optimize model performance with fewer labeled instances, focusing on labeling the most challenging or ambiguous images. For example, when analyzing cell images, the algorithm might prioritize labeling samples with uncertain cell boundaries or rare cell types to improve the model's accuracy in these difficult cases. This strategy is particularly valuable in microscopy, where expert annotation is often costly and time-consuming.

2.2.2. Neural networks

Neural network (NN)s are powerful tools for tackling various challenges in science and engineering, including microscopy image analysis. These networks are inspired by how the human brain works—processing information in parallel, learning from experiences, and adapting to new situations. The goal is to create systems that can handle complex problems, similar to how the brain processes visual information.

2.2.3. Understanding concepts through a simple neural network

Imagine a simple three-layer NN designed to analyze microscope images in shades of gray (Fig. 2.3).

Input layer: In the first layer, each neuron receives the intensity value of a single pixel in the image. Therefore, all the pixels in the image contribute information to the network.

Hidden layer: In the middle layer, the network begins to grasp important features from the input. Each node in this layer is connected to all nodes in the input layer. The network combines these connections using weights, aggregates them, and then applies an activation function (Fig. 2.3). This function introduces non-linearity to the model, allowing it to learn non-linear patterns and complex aspects in the image.

Output layer: The final layer depends on the specific analysis task. For example, if the objective is to classify different cell types or structures within the microscopic image, the output layer might consist of several nodes, each representing the likelihood of the image belonging to a specific cell type or structure.

Objective function and backpropagation: After the network produces the output (prediction), the backpropagation process occurs, during which the network's weights are updated based on the calculated error. Typically, an objective function is utilized to quantify this error. In many cases, the objective function includes a loss function that directly measures the difference between the predicted output and the actual label, as well as a regularization term to enhance learning generalization and reduce overfitting risks.

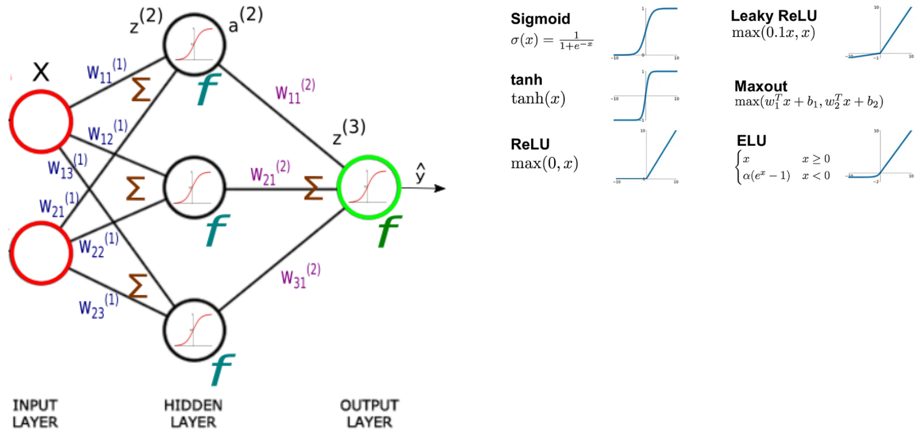


Figure 2.3.: Left: Illustration of a basic three-layer neural network. Right: Various activation functions. In the depicted neural network, the Sigmoid activation function is employed.

Consider a scenario where a 3-layer neural network classifies three classes, and the chosen loss function is cross-entropy. The network's error is computed using the gradient of the Cross-Entropy loss function [71], which is then propagated backward.

$$L = -\sum_{i=1}^{3} y_i \log(\hat{y}_i)$$

where $\hat{y}_i$ represents the predicted probability for the ith cell type and y_i is the actual label indicating whether a cell belongs to that type. The objective function in this context typically incorporates this loss function to evaluate the network's performance and optimize the learning process

During backpropagation, this error is propagated backward through the network layers, adjusting the weights to minimize the objective function. The choice of an appropriate loss function, such as cross-entropy, plays a crucial role in defining the objective function, impacting the overall learning performance.

This simple network takes the intensity of each pixel in a microscope image, identifies important features in the middle layer, and generates a result in the last layer based on its learning. Despite its simplicity, this neural network is effective at discerning image details.

2.2.4. Deep learning

Deep learning, an area within machine learning, can be viewed as automating human learning processes. It begins with raw input—such as a grayscale microscopic image—and gradually builds towards a complete understanding of the structures within that image. This process involves multiple layers in a neural network, each contributing to different levels of learning. The terms "deeper" or "deepest" learning refer to the number of layers used; "deepest" learning is fully automated from start to finish, while "deeper" learning combines human insights with computer processing to achieve the final result [71, 98].

In image processing applications, each layer in a DL model transforms input data into more abstract representations. The first layer might detect basic features like edges from pixel values. The second layer combines these edges to identify simple shapes. As the model progresses, later layers recognize more complex structures, such as cells or other microscopic patterns. Eventually, higher layers can spot specific types of cells or smaller details within the image. This hierarchical learning mirrors how humans interpret visual information. Unlike traditional ML, which requires separate feature extraction before learning, DL integrates feature extraction and learning into one seamless process. This enables the system to automatically find and adapt to complex patterns in the data [98].

Many DL architectures exist, but this thesis's central emphasis is on a U-architecture network. The subsequent section will explore this particular network.

2.3. Microscopy image denoising

Many strategies have been proposed to eliminate noise and recover signals. Traditional methods like Gaussian filtering smooth images by blurring them to reduce noise often result in a loss of detail. More advanced techniques include Non-Local Means (NLM) [20], Block-Matching 3D [37], and wavelet transforms [111], which effectively preserve details while reducing noise.

Recently, DL has emerged as a powerful denoising tool [12, 121], overcoming the blurring drawbacks of traditional methods. These DL networks are trained on example data to fine-tune millions of parameters, improving the denoising process.

2.3.1. Deep learning-based denoising

Various DL-based denoising methods exist. Initial attempts were standard deep networks, necessitating training data sets including low-quality and high-quality images. The networks were then trained on a reference dataset with high SNR as the "ground truth" to restore matched images with low SNR, referred to as "noisy data" [176]. This network is based on U-net architecture introduced in [146]

The loss function that is used in this network is the Mean Squared Error (MSE) loss function, denoted as L_{MSE}:

$$L_{\text{MSE}} = \frac{1}{n}\Sigma_{i=1}^{n}\left(f_\theta\left(x_i\right) - y_i\right)^2 \tag{2.1}$$

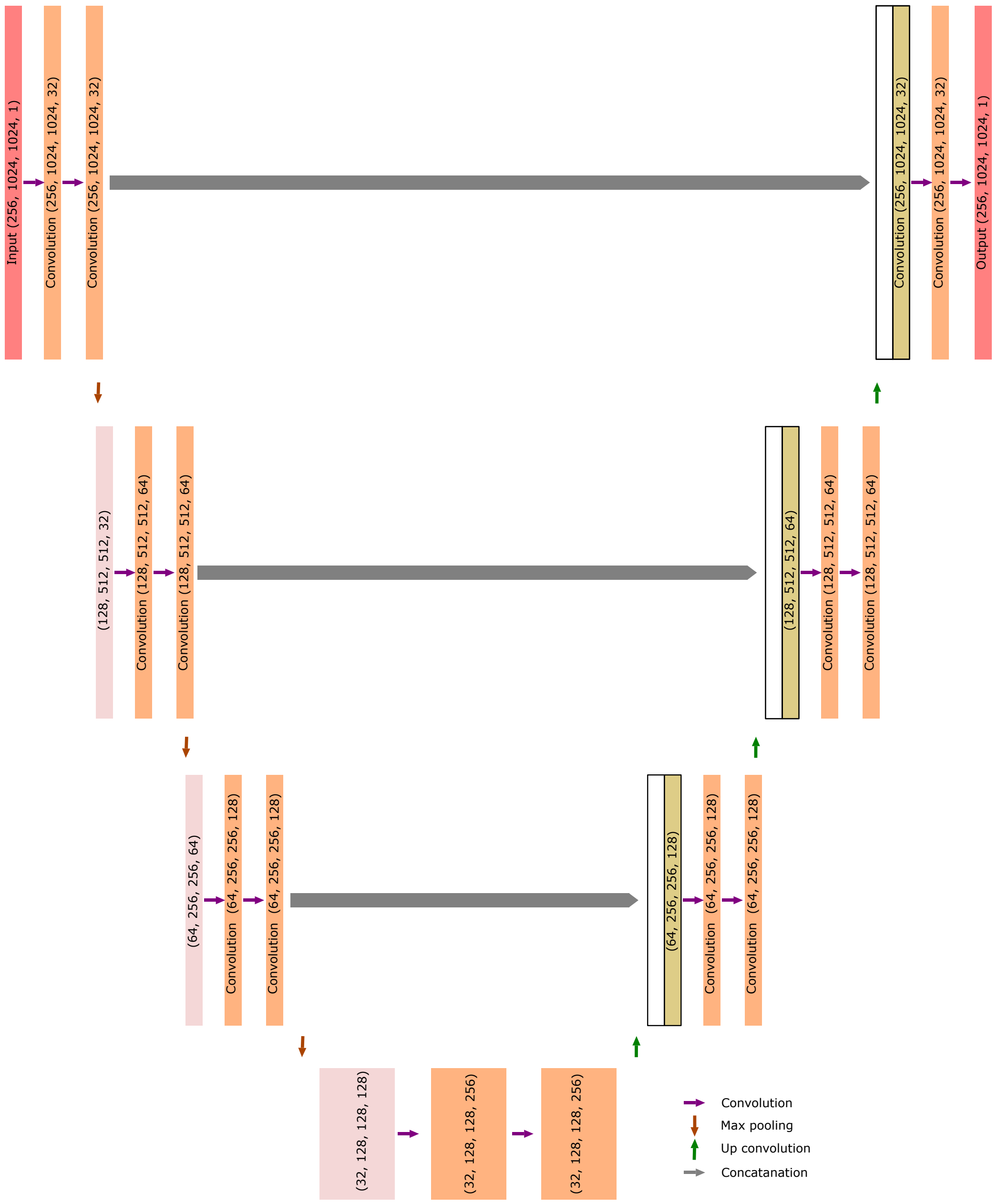


Figure 2.4.: U-net architecture based on [146]

Where:

- L_{MSE} is the Mean Squared Error loss function.
- n is the number of elements in the vectors $f_\theta\left(x_i\right)$ and y.
- $f_\theta\left(x_i\right)$ represents the predicted value for the i-th element of the vector $f_\theta\left(x\right)$, which is the denoised image obtained from the noisy input x.
- y_i is the i-th element of the vector y, which is the clear (ground truth) image.

- θ represents the parameters of the denoising model.

The loss function calculates the average squared difference between each pixel of the denoised image $f_\theta(x_i)$ and the corresponding element of the clear image y. Minimizing this loss function during training ensures that, on average, the denoised image is as close as possible to the clear image. The goal is to reduce the overall difference between the predicted denoised image and the actual clear image, making the denoising model learn to produce outputs that closely resemble the ground truth.

However, a significant challenge hindering the widespread application of denoising networks is the requirement for matched high-quality training data [13, 120]. Obtaining such data can be labor-intensive or, in some cases, even unattainable with sufficient compatibility with noisy data.

To address this issue, Noise2Noise (N2N) techniques provide a powerful alternative by enabling the training of deep networks using only pairs of noisy images, without needing clean images [100, 101]. If a noisy image can be represented as

To address this issue, N2N techniques provide a powerful alternative by enabling the training of deep networks using only pairs of noisy images, without needing clean images [100, 101]. If a noisy image can be represented as $s + n$ where s denotes the signal and n denotes the noise, then the MSE loss function for N2N network is expressed as:

$$L_{\text{MSE}} : \|f_\theta(x_i) - y\|_2^2 \Rightarrow f_\theta = E[y] = E[s + n] \xRightarrow{E[n]=0} f_\theta = E[s]$$

The network learns to map noisy images to cleaner ones by minimizing this loss function. The assumption here is that the two noisy images must share the same underlying content while having independent noise, a requirement that can still be difficult to fulfill in practice.

To overcome this limitation, a practical approach called N2V has been proposed, allowing the use of a single noisy image as both the input and training target for the denoising network. This method, as demonstrated by Krull et al. [91], eliminates the need for matched noisy image pairs or clean reference images. Instead, it focuses on training the network to predict missing or corrupted pixels in the noisy image itself, making it highly useful in scenarios where obtaining multiple noisy samples or clean data is not feasible.

The objective of empirical risk minimization can be represented as follows:

$$\theta^* = \arg\min_\theta \sum_j \sum_i L\left(f\left(x_j^{\text{RF}(i)}; \theta\right), x_j^i\right)$$

Where:

- θ represents the parameters of the neural network.
- j indexes the training examples.

- i indexes the pixels within each example.
- $\mathrm{RF}(i)$ represents a random field surrounding pixel i, but excludes the value of the central pixel.
- $f\left(x_j^{\mathrm{RF}(i)}; \theta\right)$ is the output of the neural network for the noisy input patch $x_j^{\mathrm{RF}(i)}$.
- x_j^i is the noisy target pixel taken from the image x_j at position i.
- L is the loss function (mean square error), which typically measures the difference between the predicted output and the target value.

However, a potential drawback arises when denoising techniques, such as N2V, unintentionally prioritize the central pixel excessively. In these methods, there's a tendency for the model to heavily weigh the central pixel's information, potentially leading to a 'blind spot' where contextual details from the surrounding areas are overlooked. Nevertheless, the need for precisely matched images is eliminated. In this scenario, noise learning and removal are based on a single set of noisy image data [91]. Denoising with only one noisy dataset also allows individualized training for each image, addressing variations such as day-to-day changes in fluorescence labeling or minor adjustments to optical components.

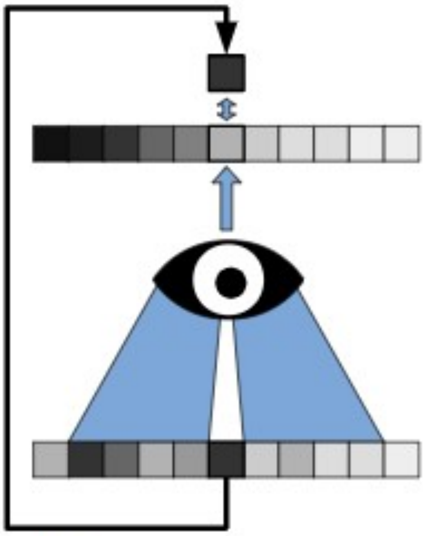


Figure 2.5.: Visualization depicting the blind spot network, showcasing the weighted sum of surrounding pixels compared to the value of the center pixel.

2.3.2. Metrics for denoising

Metrics measure image quality and accuracy when assessing denoising algorithms. The following subsection explains several common metrics, which evaluate the effectiveness of denoising algorithms in preserving image details while reducing noise.

2.3.2.1. Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is a common metric to measure how similar two images are. The formula is given by:

$$\mathrm{SSIM}(x, y) = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\right) \times \left(\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}\right) \times \left(\frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}\right) \tag{2.2}$$

In this formula, the terms c_1, c_2, and c_3 are constants, often set to 0.01, 0.03, and $c_2/2$ respectively. The SSIM formula has three parts: the Luminance term, the Contrast term, and the Structural term.

The Luminance term looks at the similarity in brightness between the two images. It considers each image's average brightness (μ) (x and y) and their interaction, using c_1 for stability.

The Contrast term measures the similarity in contrast, evaluating the variability (σ) in the pixel values of the individual images (x and y) and their interaction. c_2 is introduced for stability, similar to the Luminance term.

The Structural term analyzes the structural information in the images, looking at how pixels are related through covariance (σ_{xy}) and variance (σ_x^2 and σ_y^2). Introducing c_3 constant for stability further improves the assessment of structural similarity.

In simpler terms, the SSIM score comprehensively measures both luminance and structural similarities between images. A higher SSIM value indicates a closer resemblance, with value 1 representing a perfect match. This formula is particularly useful for comparing images in various applications.

To show how SSIM works, in Fig. 2.6, we employed a Leica Stimulated Emission Depletion (STED) microscope, running in full 3D depletion mode, to capture detailed images of DNA distribution within the nuclei of fixed zebrafish embryos. We took five images, each with a different dwell time—ranging from 0.6 ms to 10 ms, doubling with each step. Dwell time, representing the laser exposure period, is a crucial parameter in STED microscopy that significantly influences image quality. We aimed to examine how varying dwell times affect image similarity by comparing each to a reference image taken at the maximum dwell time of 10 ms.

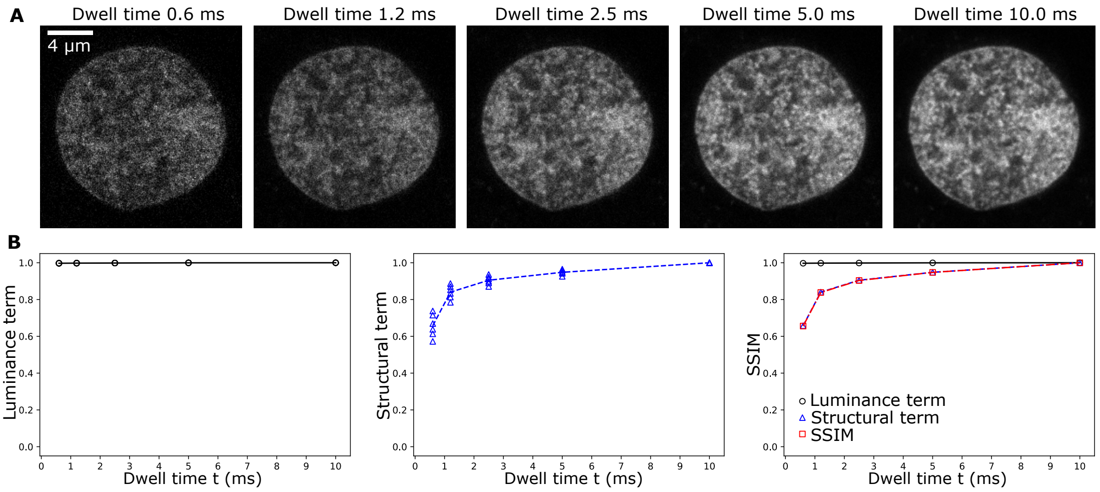


Figure 2.6.: Figures show the DNA distribution in a mid-section of a nucleus in a fixed zebrafish embryo obtained by stimulated emission depletion (STED) microscopy. A) To illustrate how the different terms of the SSIM metric depend on the duration of photon collection in the averaging detector type, DNA images were acquired with the detector left open to collect photons at each pixel for longer times (dwell time t), then the photon count is normalized by the dwell time t. B) In the averaging detector case, the luminance term is constant and close to 1.0. Only the structural term changes with increasing t, so the overall SSIM values reflect structural reliability for a given t. SSIM values were calculated based on $n = 6$ images obtained for a given t, combined with a matching image recorded with the highest $t = 10$ ms. Individual values are shown with the mean.

The images are fixed biological samples, so the contrast differences tend to be minimal; the sample is stable and uniformly illuminated. Unlike live imaging, where dynamic changes in contrast can be significant, fixed samples generally do not show substantial contrast variation. This further diminishes the relevance of the contrast term in this analysis. That is why we only focus on the Luminance term and the Structural term of SSIM value to measure the similarity between each image and a reference image, which was recorded with a 10 ms dwell time. 2.6 shows that the luminance part stays consistent across different dwell times due to the uniform illumination provided by the STED microscope. However, the structural term increases with increasing dwell time. This suggests that higher SSIM values indicate a greater structural resemblance, implying potential improvements in image quality.

2.3.2.2. Fourier ring correlation (FRC)

Spatial frequency is a term used in image processing to describe the rate at which visual patterns or structures change across a given space. In the context of microscopic grayscale images, it measures how quickly intensity values (shades of gray) vary across different image regions. In spatial frequency analysis, images are examined based on the frequency of these intensity changes rather than the specific pixel values.

In a microscopic image, high spatial frequency components represent rapid changes in intensity, which typically capture fine details such as sharp edges or small structures. Low spatial frequency components, on the other hand, represent slower changes corresponding to broader features and smoother regions of the image.

A Fourier transform is commonly used to analyze the spatial frequency content of grayscale images. This method decomposes an image into its component frequency parts, allowing detailed analysis of the image's structural characteristics.

To evaluate the similarity between two independent representations of the same microscopic object in frequency space, Fourier Ring Correlation (FRC) is employed. The basic idea behind FRC is to compare how well the information in corresponding frequency components of the two representations matches. This is done by comparing the Fourier transforms of the images in groups of frequency components that correspond to specific levels of detail in the image.

The FRC curve is generated by plotting the correlation values against spatial frequency. A critical point on this curve, often defined by a specific threshold (commonly 1/7), determines the resolution limit. This resolution limit signifies the point at which the two representations are no longer consistent, indicating that the object can no longer be reliably resolved in the images beyond this frequency.

In practical terms, the threshold value of 1/7 corresponds to the point where the correlation between the images declines significantly due to noise, rather than reflecting a true signal. Although other thresholds can be chosen depending on specific experimental conditions or requirements, 1/7 is widely accepted in this field as a reasonable balance between sensitivity and noise tolerance [87].

$$\text{Resolution} = \frac{1}{\text{Spatial Frequency}}$$

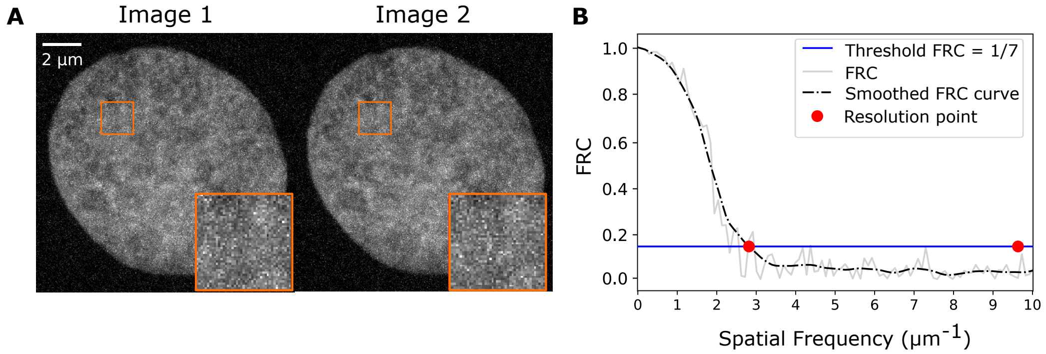


Figure 2.7.: A) Representative micrographs of the DNA distribution in a nucleus in a fixed zebrafish embryo, recorded with a stimulated emission depletion (STED) super-resolution microscope. The same image plane was recorded twice. B) FRC curves are calculated based on the image pair.

Fig. 2.7 shows the distribution of DNA in the nucleus of a fixed zebrafish embryo, captured using stimulated emission depletion (STED) microscopy. This technique allows us to see very small details that normal microscopes cannot resolve. The figure has two parts: the images of the nucleus and a graph showing how the resolution was calculated.

In panel A, two images (Image 1 and Image 2) of the same part of the nucleus were taken using STED microscopy. These images reveal fine details of how the DNA is distributed. The orange boxes highlight specific areas where we can see the structures more clearly. These areas help to demonstrate how much detail the STED microscope can resolve.

Panel B shows the results of a method called FRC, which was used to measure the resolution of the images. The black curve represents the FRC, which compares the level of detail in the two images at different scales. The blue horizontal line marks the resolution threshold (FRC = 1/7), which is a standard value used to separate real image details from noise. The point where the black curve crosses the blue line is called the resolution point, shown here with a red dot. This happens at a spatial frequency of about 3 μm^{-1}, which means the resolution of these images is about 333 nanometers (calculated as 1/spatial frequency in µm).

2.3.2.3. Local SSIM

Local SSIM enhances the SSIM approach by assessing structural similarity variations across distinct partitions of an image. Instead of deriving a single SSIM value for the entire image, local SSIM divides the image into smaller, non-overlapping partitions. The structural aspect of the SSIM is independently computed for each corresponding pair of partitions in the original and reconstructed images. This process allows for a more detailed analysis of the quality of the reconstruction. Fig. 2.8 illustrates the map comparing reconstructed and high-quality images. We see the reconstructed image in panel A and in panel B, the reference image. Panel C shows the local SSIM map, which compares the two images. The map uses a scale from 0 to 1, where 1 (black areas) means the images are very similar in that region, and 0 (white areas) is different. This helps to see which parts of the images are alike and where they vary.

Partitioning of images: The initial step is dividing the original and reconstructed images into smaller, non-overlapping partitions.

Calculation of local SSIM: Following partitioning, the structural component of the SSIM is calculated for each corresponding pair of partitions in the original and reconstructed images.

Creation of SSIM map: Next, an SSIM map is generated to illustrate the spatial distribution of structural similarity values throughout the image. This map effectively highlights areas that exhibit higher or lower structural similarity.

Detection of faults: The SSIM map is then examined to identify regions where local SSIM values consistently register as low. Such areas may indicate portions of the reconstructed image susceptible to distortion, noise, or other imperfections.

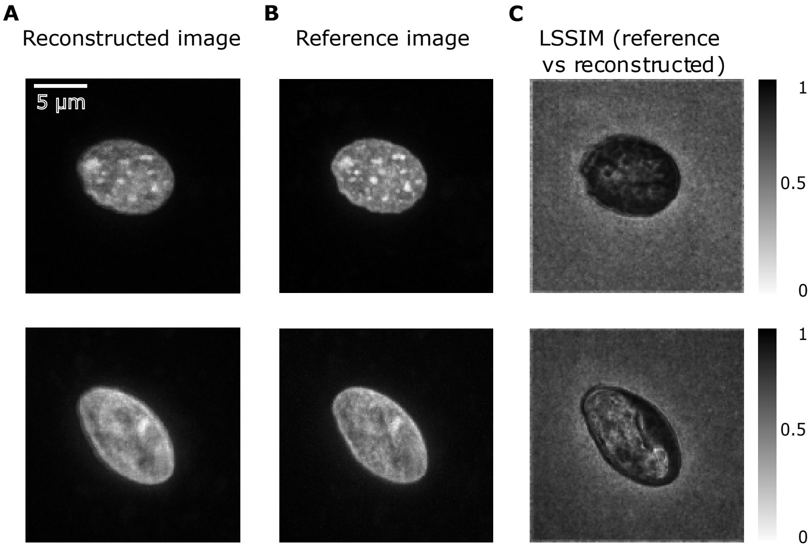


Figure 2.8.: A) Representative micrographs of nuclei of human cheek cells in which DNA was labeled by Hoechst 33342. Images are reconstructed. B) The reference images are acquired at the same position but with a long exposure time. D) Local structural similarity index metric (SSIM) map for comparing reconstructed and reference images.

2.4. Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a measure used to assess the quality of a reconstructed or denoised image compared to the original reference image. It quantifies the ratio between the maximum possible power of a signal (in this case, the image) and the power of corrupting noise that affects the fidelity of its representation. Higher PSNR values generally indicate that the reconstructed image is closer to the original, implying better quality.

These are the steps to calculate PSNR:

1. Mean Squared Error (MSE) calculation:

The MSE between the original image I and the reconstructed image K is calculated as:

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

where:

- $m \times n$ are the dimensions of the images.
- $I(i,j)$ is the pixel value at position (i,j) in the original image.
- $K(i,j)$ is the pixel value at position (i,j) in the reconstructed image.

2. PSNR computation

Once the MSE is calculated, the PSNR can be determined using the following formula:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}}\right)$$

where MAX_I is the maximum possible pixel value of the image. For an 8-bit grayscale image, this value is 255.

2.5. Microscopy image segmentation

Segmentation means automatically outlining objects like cells, an important step before measuring their characteristics. Accurate segmentation enables scientists to analyze cell features, but it remains an ongoing and complex problem due to various cell types and imaging methods.

Advancements in segmentation are mostly driven by DL methods. However, DL approaches may not always be the best choice, especially when dealing with tasks that require large annotated datasets and present challenges in terms of interoperability. Therefore, our exploration begins with developments in traditional ML processing techniques.

2.5.1. Thresholding-based segmentation methods

The most basic thresholding techniques generate a mask by assigning a value of 0 to pixels with intensity values below a specified threshold intensity, denoted as T. Conversely, pixels with intensity values exceeding the threshold are assigned a value of 1.

While users can manually set the threshold, several automated strategies exist to calculate it. These varied strategies include histogram shape analysis, clustering, and entropy-based methods. Histogram analysis studies the distribution of intensity values; clustering techniques aim to group similar pixels, and entropy-based methods assess the information content in the image. These diverse strategies offer users flexibility in choosing an approach.

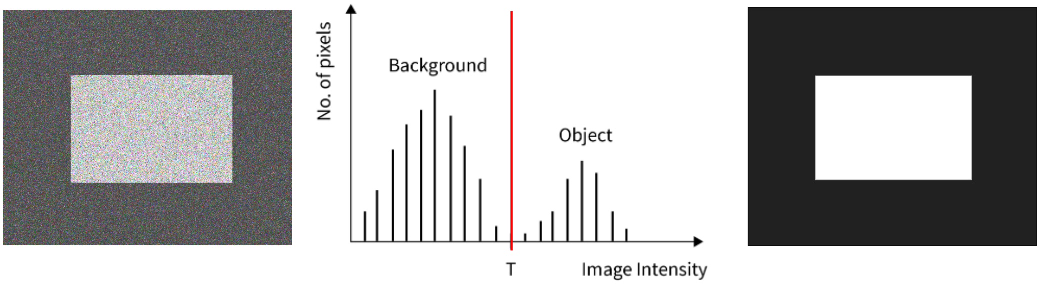


Figure 2.9.: on the left is the original image, showcasing the visual representation of the data. In the middle, the histogram visually portrays the distribution of pixel intensities within the image. The right image illustrates the mask, a binary representation generated through thresholding.

The choice of thresholding technique is important to determine the accuracy and effectiveness of image analysis. Different thresholding techniques have their strengths and limitations. Selecting the appropriate method depends on factors such as image complexity, noise levels, and the desired outcome. Therefore, it is essential to consider the selection carefully and to conduct experimentation to ensure optimal results in image processing tasks. Here, we explore some thresholding techniques.

Global thresholding Global Thresholding is a widely used method where a single threshold value is applied evenly across the entire image. This method works well when the foreground and background have easily distinguishable intensity distributions. However, it may not work as effectively when dealing with complex intensity patterns or significant changes in lighting throughout the image. Additionally, it might struggle to accurately define objects or areas with overlapping intensity values.

Adaptive thresholding techniques have become important in addressing these issues. These techniques dynamically adjust the threshold values locally, considering the specific characteristics of each pixel's surroundings. This adaptive approach is particularly useful when illumination varies across different parts of the image.

Local (adaptive) thresholding strategies Local thresholding responds to global thresholding's limitations. Examining smaller regions of the image, this technique calculates personalized threshold values based on local features like mean or median intensity. This approach improves adaptability to various intensity distributions. However, it's important to note that local thresholding may require more computational resources and could be sensitive to noise or uneven illumination, potentially affecting the overall performance of the segmentation algorithm.

Mean and Gaussian adaptive thresholding are widely used techniques. Mean adaptive thresholding calculates the threshold value for each subregion by averaging the intensity of all pixels within that area. Meanwhile, Gaussian adaptive thresholding takes a weighted average of pixel intensities, prioritizing pixels near the center of the subregion.

Otsu Otsu's Method for Automatic Threshold Determination is a commonly employed technique to automatically find the best threshold value in image segmentation. It determines the threshold by maximizing the variance between classes of pixel values, which can be expressed mathematically as follows:

$$\sigma^2_{\text{between}}(t) = \omega_1(t) \cdot \omega_2(t) \cdot [\mu_1(t) - \mu_2(t)]^2$$

Where:

- $\sigma^2_{\text{between}}(t)$ is the between-class variance at threshold t.

- $\omega_1(t)$ and $\omega_2(t)$ are the probabilities of the two classes separated by threshold t.
- $\mu_1(t)$ and $\mu_2(t)$ are the means of the pixel values in the two classes.

This effectively distinguishes foreground and background regions. This method proves especially good when working with images featuring bimodal or multimodal intensity distributions, as it precisely identifies the threshold that most effectively separates various objects or regions in the image.

2.5.2. Whatershed segmentation

The watershed algorithm is also a powerful segmentation among classic image segmentation techniques. This algorithm conceptualizes an image as a topographical landscape, creating ’catchment basins’ and ’watershed ridge lines’ to delineate various objects within the image. To simplify, imagine a grayscale image as terrain where heightened intensity represents peaks and hills, while lower intensity signifies valleys.

Although the Watershed algorithm is conceptually straightforward and effective, it occasionally encounters challenges, such as over-segmentation. Over-segmentation occurs when an object is excessively divided into multiple segments. However, optimizing the algorithm and integrating pre-processing steps can refine its performance and address such challenges. Fine-tuning means adjusting its parameters and incorporating additional steps before segmentation to ensure more accurate and satisfactory results.

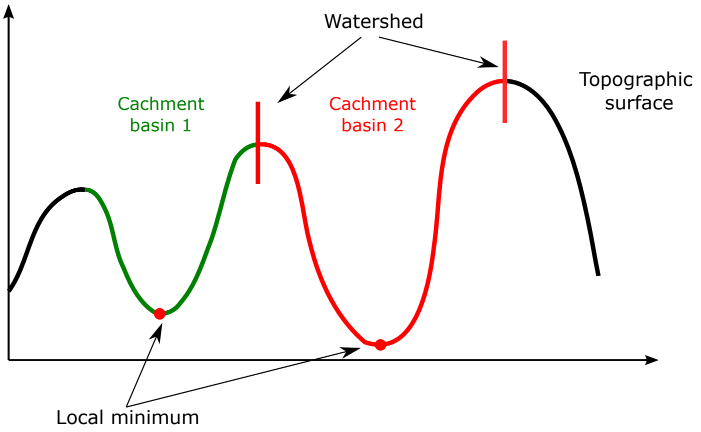


Figure 2.10.: Illustration depicting the watershed concept, featuring the watershed, local minima, and catchment basins within the topographic curve.

2.5.3. Deep segmentation for microscopy images

Semantic segmentation assigns a role to every pixel in an image, such as identifying them as part of the main subject or background. The Fully Convolutional Network is a widely used technique for performing segmentation. It is a network type designed explicitly for semantic segmentation without needing fully connected layers. Other methods, such as SegNet[6], PSPNet [188], and U-Net [146], are built on the foundation of FCN. U-Net, in particular, is well-known for its success in analyzing microscopy images due to its unique U-shaped design.

U-Net works in two main steps: first, it zooms out to capture the big picture, extracting features and ensuring it doesn't miss anything important. Then, it zooms back in to refine its understanding and make detailed decisions. This process helps it handle challenges like crowded cell images. Another advantage of U-Net is its ability to learn well even with limited examples, which is particularly useful in microscopy, where data can be scarce. Researchers have also modified the U-Net model to better suit tasks. However, microscopy images pose their unique challenges, such as poor contrast or irregularly shaped cells, which can be problematic for U-Net.

2.5.3.1. Generalization challenges and solutions

Before undergoing fine-tuning, deep learning models often struggle with new, unseen data, referred to as out-of-distribution data. These datasets differ significantly from the original training data due to variations in acquisition parameters, staining methods, or imaging techniques. This difference in style can be addressed using a style transfer process, which means training a model to convert images from one style to another while preserving semantic content. For example, a style transfer model can change brightfield images to fluorescence images or convert annotation masks into their corresponding images.

The NucleiAlzer model [73], introduced by Hollandi et al., demonstrates a strategy to improve its ability to generalize to out-of-distribution data. It uses style transfer to convert annotation masks into images, making it more applicable for various data styles. Despite being in existence for over three years, this model continues to be at the leading edge of cell segmentation models [99]. To make it more user-friendly, a plugin for CellProfiler 3 was developed by McQuin et al. [118], allowing users to conduct inference through a graphical user interface (GUI). However, the manual installation of NucleiAlzer and its dependencies can be challenging for non-technical users. Additionally, while a web interface is available, it requires data to be uploaded to a central server, which makes it less convenient for processing large batches of images.

StarDist, an older model than nucleiAlzer, continues to receive updates and improvements [113]. StarDist employs a star-convex shape model, achieving high accuracy in segmenting objects that fit this geometric constraint. Its comprehensive documentation and compatibility with various environments and tools, such as Fiji/ImageJ, Napari[1], CellProfiler, and KNIME [49], contribute to its widespread use. Even though pre-trained models are available, StarDist is only suitable for specific data types, such as fluorescence or histology stains, and is designed for star-convex objects. This means it may not work well for irregularly shaped objects. Star-convex shapes are defined in a way that limits their flexibility for complex

---

[1] napari.org

biological structures, as any line segment from a boundary point to an interior point must lie entirely within the object.

Cellpose [132] aimed to create a generalist model by training on a large dataset of manually segmented images from diverse modalities. The preprocessing method of this model converts input data into spatial gradients, allowing it to work with various shapes. This preprocessing captures important gradient information for defining boundaries in different cell types and shapes. While the model is designed for general use, it often requires fine-tuning for specific datasets. Cellpose 2.0 [132] introduced multiple pre-trained models, a human-in-the-loop pipeline for fine-tuning custom models with small datasets, and improved graphical software tools. This pipeline allows iterative model improvement through user feedback during training.

Although Cellpose supports training, this feature is unavailable in Cellpose 2.0, which focuses on human-in-the-loop fine-tuning. Both versions offer a custom GUI and extensive documentation, with broad support through plugins for many of the same tools as StarDist. This facilitates integration into existing workflows and leverages familiar tools for segmentation tasks.

Mesmer, a deep learning pipeline trained on the largest public sample dataset of annotated nuclei and whole cells, TissueNet [58], offers remote access to its model via a web portal and plugins for Fiji/ImageJ. Mesmer's extensive training in diverse sample types allows it to perform well across various histological images. It also provides a Docker container for self-hosted use, accessible via Jupyter Notebook or CLI, enabling users to run Mesmer in their computing environments, ensuring data privacy and control over computational resources.

2.5.4. Metrics for segmentation

Metrics for evaluating image segmentation can be categorized based on various criteria. One criterion is whether a metric is subjective or objective [33]. Subjective metrics rely on human visual assessments and are not formally defined by nature. These metrics often mean qualitative judgments made by human observers, such as assessing the visual similarity between segmented regions and ground truth annotations. Since subjective metrics are based on human perception, they may vary depending on individual observers and their subjective interpretations. To mitigate this variability and achieve a more generalizable result, many observers should be involved to provide diverse perspectives. Additionally, subjective metrics may lack formal mathematical definitions, making them less suitable for automated quantitative analysis. Despite these limitations, subjective metrics play a important role in evaluating image segmentation algorithms, as they provide valuable information into how well the segmentation results align with human perception and visual understanding [33].

Another criterion is whether the evaluation of a segmentation approach is analytical or empirical [187]. In empirical evaluations, accuracy and precision are evaluated based on example data in a supervised manner. On the other hand, analytical evaluation does not rely on example data but instead examines the theoretical properties of the segmentation. This means analyzing the underlying principles, algorithms, and mathematical models used in the segmentation process. By studying these theoretical aspects, analytical evaluation aims to assess the segmentation algorithm's performance based on its inherent capabilities and characteristics rather than its performance on specific example data. This approach provides information into the algorithm's behavior and performance under various conditions, helping to understand its strengths, weaknesses, and theoretical limitations [173].

As a third criterion, the evaluation of image segmentation can be supervised (with annotated images) or unsupervised (without annotated images). Supervised evaluation benefits from directly comparing the segmentation results and the reference annotated image. The evaluation can be based on per-pixel comparison (True Positive (TP) and True Negative (TN), Matthews Correlation Coefficient (MCC), and F-measure), region-based comparison (Bipartite Graph Matching (BGM) [166] and Segmentation Covering (SC) [40]), or distance-based evaluation (Hausdorff Distance, directional Hamming distance, and Mahalanobis distance [137]). Unsupervised metrics evaluate the quality of segmentation by directly calculating the feature parameters of the result without using the annotated reference image. These metrics are ideal, especially when there is no ground truth. They are mostly based on mathematical indicators showing the quality of segmentation results, such as Peak Signal-to-Noise Ratio (PSNR) [186], calculating the inside and outside contrast of the segmentation area, or measuring the foreground and background variances [186].

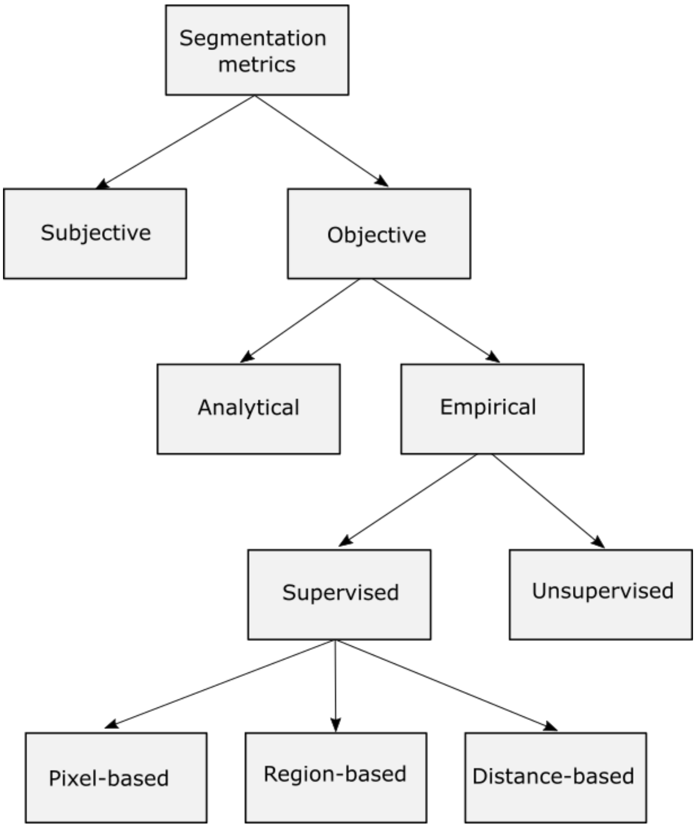


Figure 2.11.: Diagram illustrating various segmentation metrics categorized by different aspects.

2.6. Active learning and different strategies

Active learning is a specialized branch of machine learning that addresses the challenge of labeling data. Obtaining labeled data can be costly and time-consuming in many applications. Active learning aims to minimize this cost by intelligently selecting the most informative data points to label, thus improving the learning process more efficiently.

Principles of active learning

- Selective querying: Instead of labeling all available data, the algorithm selectively queries a subset of data points expected to be the most informative for the model. This selective querying is based on certain criteria that measure the potential impact of the data points on the model's performance.
- Iterative process: Active learning is an iterative process where the model is trained on the initially labeled data, queries the most informative data points, updates the labeled dataset, and retrains the model. This cycle continues until a stopping criterion is met, such as reaching a desired accuracy level or exhausting the labeling budget.
- Feedback loop: The algorithm continuously learns from the labeled data and adjusts its querying strategy based on the feedback. This adaptive process helps in focusing on areas where the model is uncertain or where the data distribution is sparse.

2.6.1. Active learning query strategies

Active learning query strategies determine which data points should be selected for labeling to improve the model's performance most effectively. These strategies can be broadly categorized into stream-based selective sampling, pool-based sampling, and query synthesis methods. Additionally, these queries can be based on various informativeness measures that assess the potential value of each data point.

2.6.1.1. Stream-based selective sampling

Stream-based selective sampling means evaluating data points as they arrive in a sequential manner. The decision to query a label for a data point is made immediately based on certain criteria, such as uncertainty or diversity. This method is particularly useful in scenarios where data arrives continuously and decisions must be made in real time.

2.6.1.2. Pool-based sampling

Pool-based sampling operates on a large pool of unlabeled data, where the model iteratively selects the most informative data points for labeling. This approach is suitable for batch processing and allows for a more comprehensive analysis of the data pool to identify the most valuable samples. Common strategies in pool-based sampling include uncertainty sampling, where the model queries the data points about which it is least confident, and diversity sampling, where the model aims to cover a wide range of data variability.

2.6.1.3. Query synthesis methods

Query synthesis methods generate new data points that are considered most informative for the model. Unlike stream-based and pool-based methods, which rely on existing data, query synthesis creates hypothetical data points that can challenge the model's current understanding. These methods are particularly powerful in scenarios where the data space is sparse or when exploring edge cases not well represented in the existing data.

2.6.1.4. Query based on informativeness measures

Queries based on informativeness measures assess the potential value of a data point by evaluating its contribution to improving the model. Informativeness can be measured through various metrics, such as uncertainty, expected model change, or information density. By selecting data points that maximize these measures, active learning strategies aim to optimize the learning process and achieve better performance with fewer labeled samples.

2.6.1.5. Example of active learning in microscopy image segmentation

For example, Rombaut, Roels, and Saeys [145] presents BioSegment, a platform that facilitates large 3D electron microscopy image segmentation through active learning. Electron microscopy requires detailed and labor-intensive segmentation to analyze high-resolution 3D structures. Traditional DL segmentation methods demand extensive labeled datasets, which are costly and time-consuming.

BioSegment [145] incorporates active learning strategies that prioritize the most informative samples for labeling, thus reducing the effort needed to annotate data. The paper compares five different active learning strategies, including maximum entropy sampling, which showed improved segmentation quality (up to 15% increase) compared to random sampling.

Using active learning, BioSegment achieves segmentation quality comparable to full supervision with only 25% of the annotation budget. This makes the tool especially valuable for efficiently handling large and complex Electron Microscope datasets. The platform was validated on three public Electron Microscope datasets, demonstrating that active learning strategies, particularly maximum entropy and least confidence sampling, perform better than random sampling. It supports interactive annotation, large datasets, 3D data, and remote processing.

However, the paper has some weaknesses that make it difficult for users without computer skills to use. Setting up BioSegment requires tools like Docker and managing APIs, which can be confusing for non-technical users. Understanding how active learning works and fine-tuning models also requires some knowledge of ML. The software also needs powerful computers to handle large datasets, which may not be available to everyone. These challenges make BioSegment difficult for users who do not have experience in technology or data science.

2.7. Foundational concepts in biology

To help readers understand the biological terms used in this study, here is a simple explanation of the key concepts:

2.7.1. Macromolecular clusters

Macromolecular clusters are groups of large molecules, such as proteins or RNA, that gather together in specific areas inside cells. They play important roles in how cells function, like helping make proteins, sending signals, or organizing the cell's structure. Think of them like little workstations in a factory, each focused on a specific job.

2.7.2. Pol II (RNA Polymerase II)

Pol II is an enzyme that helps cells make messenger RNA (mRNA), which carries instructions from DNA to build proteins. During its work, Pol II goes through different "states" based on chemical changes called phosphorylation. Two key states are:

- **Pol II Ser5P**: This state happens when Pol II is just starting its job and getting ready to copy the DNA.
- **Pol II Ser2P**: This state happens when Pol II is actively making RNA and moving along the DNA.

2.7.3. Cluster morphology

Cluster morphology refers to the shape and structure of macromolecular clusters. By studying their shape in 3D, scientists can understand how these clusters work and how they change while performing their functions.

2.7.4. Fabs (Fluorescently Labeled Antibody Fragments)

Fabs are small pieces of antibodies specially designed to stick to specific molecules in cells. In this study, these Fabs are attached to glowing markers (fluorescent labels), so scientists can see where certain molecules are and what they are doing under a microscope. For example, Fabs in this study are used to detect Pol II in its different states without harming the cells or the embryos being studied.

2.7.5. Lateral resolution

Lateral resolution refers to the ability of a microscope or imaging system to distinguish between two points that are close to each other in the horizontal plane (the X-Y plane). It determines how sharp and detailed an image appears across the surface being observed. For example, if two small structures are very close together, high lateral resolution means the system can clearly show them as two separate objects.

2.7.6. Sphere-stage Zebrafish embryos

Sphere-stage zebrafish embryos refer to a specific developmental stage in zebrafish, occurring approximately 4 hours after fertilization. At this stage, the embryo appears as a spherical mass of cells, with roughly 1000 cells forming a ball-like structure. This stage is part of the early development process when the cells are dividing rapidly and the basic body plan is beginning to form.

Zebrafish embryos are commonly used as a model in biological studies because they develop quickly, are transparent, and allow researchers to observe and manipulate cellular processes in real-time. The sphere stage is particularly useful for imaging studies because the embryo is still small and relatively uniform, making it ideal for high-resolution microscopy and live imaging techniques.

Part II.

Contributions

3. Speeding up live microscopy with DL denoising

3.1. Motivation

Light microscopy is a crucial tool in biological research [163]. Biologists use it to observe cellular processes and test hypotheses, often by visualizing fluorescently labeled molecules. However, designing these experiments comes with limitations, such as preventing damage to the specimen, acquiring data quickly, and maintaining a high SNR [75, 96, 83]. These parameters are interdependent, meaning optimizing one can negatively affect the others. For example, reducing exposure time can increase acquisition speed but lower SNR, while increasing light power to improve SNR can cause photodamage to the specimen [136, 152].

While strict trade-off relationships constrain experimental parameters during acquisition, the computational processing of images post-acquisition can significantly enhance image quality. This allows for reduced exposure times, followed by computational reconstruction of low-SNR images. Traditional methods for reconstructing low-SNR images include projection techniques [156], deconvolution filters [143, 138], and various traditional denoising methods [19, 56].

In the last decade, DL has emerged as a powerful tool in image processing, often outperforming these traditional methods [126]. It has proven particularly effective in biological microscopy for different tasks such as image restoration [93, 165, 129, 81, 11, 22, 92, 144, 123]. Initially, DL techniques utilized standard networks trained with datasets composed of low-quality and corresponding high-quality images. Networks trained on such reference datasets could restore images from noisy data to a high-SNR "ground truth" [176]. However, a major challenge has been the need for precisely matched high-quality training datasets, which are difficult and sometimes impossible to obtain [13, 120].

To address this limitation, self-supervised denoising methods like N2N and N2V have gained prominence. These methods enable denoising without high-quality ground truth data [101, 91]. N2V specifically trains on single noisy datasets, making it adaptable to day-to-day variations in experimental setups, such as changes in fluorescence labeling or optical components. Despite these advancements, reusing DL methods in image reconstruction introduces a critical challenge: ensuring quality without inadvertently creating artifacts or false structures [13, 120].

From a Software Engineering for AI perspective, this challenge highlights the importance of tailored solutions for quality assurance. Unlike general-purpose imaging tasks, biological microscopy involves highly variable experimental conditions, including sample dynamics, fluorescence labeling, and optical setups. Standardized quality benchmarks often fail to address these domain-specific challenges. Therefore, an adaptable and systematic approach is needed to ensure the reliability of DL-based workflows while maintaining the flexibility required for diverse applications.

This chapter proposes a pragmatic imaging protocol inspired by SE principles, enabling biologists to reuse DL methods like N2V with quality assurance. The protocol addresses three key questions:

- How can biologists effectively reuse DL denoising techniques in the image processing pipeline to improve the SNR of light microscopy images? (RQ 1.1)
- What methods can be employed to assure the quality and reliability of DL-based fluorescence image reconstruction when ground truth data is unavailable? (RQ 1.2)
- What methods can suggest the ideal exposure time for the DL denoising-supported experiment? (RQ 1.3)

The proposed imaging protocol begins by recording a small dataset comprising high-quality data to ensure precise control over the reconstruction quality for every acquired view of a given sample. This is followed by the complete time-lapse acquisition of data recorded under conditions that prioritize speed or efficiency, resulting in reduced quality compared to the initial high-quality dataset. The practicality of this imaging protocol was then demonstrated through actual case studies, specifically by analyzing fluctuations in molecular clusters within live zebrafish embryos. Notably, the analysis yielded significant results, showing a fivefold increase in speed compared to workflows without the proposed methodology [66]. This study allowed the observation of a previously unseen biological phenomenon involving rapid changes in the three-dimensional (3D) shape of molecular clusters within seconds.

Furthermore, an alternative experimental approach, described in [66], was conducted to validate the findings. This previous study, published in PNAS Nexus, independently observed the same phenomenon reported here, providing further confirmation of the results. My role in this study centered on addressing the challenges of reusing DL denoising techniques with quality assurance in the image processing pipeline. Specifically, I contributed by designing the imaging protocol to balance rapid data acquisition, SNR, and specimen safety, and by determining the optimal experimental parameters to ensure DL denoising was applied effectively. Additionally, I contributed to formulating methods to evaluate the quality of DL-reconstructed images in the absence of ground truth data. These contributions are directly tied to the research objectives of improving imaging speed, minimizing artifacts, and optimizing DL-based post-processing in light microscopy experiments.

This imaging protocol is designed to be adaptable and generalizable and provides a guideline for microscopists interested in the quality-assured reuse of ground-truth-free image reconstruction methods. Notably, the imaging protocol is intended to be flexible and generalizable to various fluorescence microscopes, regardless of their specific specifications.

It can be seamlessly implemented on microscopes commonly used for time-lapse recordings without requiring extensive software development or hardware modification beyond standard functionality.

The chapter outlines the discussion as follows: we begin by discussing the motivation behind the study. Section 3.1 highlights the critical challenges faced in live microscopy, emphasizing the need for high-speed imaging and the limitations of existing denoising techniques. Following this, Section 3.2 presents the proposed approach for addressing these challenges. This section is further broken down into four key components: evaluating structural reliability with SSIM (3.2.1) (introduced in Section 2.3.2.1), quantifying effective resolution with FRC (3.2.2) (introduced in Section 2.3.2.2), guiding exposure time selection in high-speed time-lapse imaging (3.2.3), and implementing a two-phase acquisition imaging protocol for high-speed time-lapse imaging (3.2.4). Section 3.3 evaluates the proposed method by comparing it with conventional deconvolutional network objectives (3.3.1) and conducting a case study conducting a case study on the real-time visualization of molecular dynamics (3.3.2). The discussion in Section 3.4 explores the implications and potential improvements of the approach. Section 3.5 extends the generalizability of the DL denoising models to other applications. Section 3.6 addresses potential threats to the study's validity, ensuring robustness and reliability of the results. Finally, Section 3.7 reviews related work, situating our approach within the broader context of existing research.

3.2. Proposed approach

In this section, we introduce the evaluation metrics used to ensure the quality and accuracy of reconstructed images, as well as the proposed imaging protocol that is helpful for biologists aiming to reuse unsupervised DL denoising with quality assurance in their workflows. Using metrics that effectively capture structural reliability and spatial resolution is important for accurate analysis, as they help maintain the precision of the data. We also explain why focusing on structural reliability and spatial resolution is more practical than emphasizing other aspects.

The concept of structural reliability is about how well the reconstructed image preserves the structure of the original. This is critical because any inaccuracies in the reconstructed image could lead to misunderstandings of biological processes. This directly supports RQ 1.2, ensuring the quality and reliability of DL-based reconstructions. Meanwhile, spatial resolution measures the level of detail or sharpness in the reconstructed image, which is essential for observing fine structures. This aspect aligns with RQ 1.1, focusing on an imaging protocol that improves the overall SNR in light microscopy images and enables researchers to reuse unsupervised DL denoising with quality assurance.

To thoroughly evaluate both structural reliability and spatial resolution, we use three key metrics: SSIM, local SSIM, and FRC. SSIM [174, 18] measures the similarity between the original and reconstructed images, with values from 0 to 1, where higher values indicate a closer match. This metric helps us understand how changes in image acquisition settings, like reduced exposure time, impact the reconstruction, directly addressing RQ 1.2. Local SSIM provides a finer assessment by checking specific regions of the image for any localized differences introduced during denoising. This provides a more comprehensive understanding of structural reliability and supports researchers reusing unsupervised DL denoising with quality assurance.

FRC measures effective resolution by analyzing image consistency in the frequency domain, making it an important tool for assessing the sharpness and clarity of reconstructed images. This metric works by comparing the frequency components of two images, such as raw and denoised versions, to evaluate how well fine details are preserved. It provides a quantitative measure of spatial resolution, helping researchers determine whether the reconstructed image retains the necessary level of detail for accurate analysis. This is particularly relevant for RQ 1.1, as it demonstrates how the proposed imaging protocol can enhance spatial resolution and improve image quality.

After introducing these metrics (Section 3.2.1 and Section 3.2.2), we provide guidance on optimizing exposure time for image acquisition (Section 3.2.3). This section addresses RQ 1.2 by showing how to use the metrics to find the best exposure settings that balance image quality and speed. Proper exposure optimization ensures that even when capturing images quickly, the quality remains high enough for effective DL denoising and subsequent analysis, which helps in the reuse of unsupervised DL denoising with quality assurance in time-sensitive experiments.

Finally, we present a detailed imaging protocol (Section 3.2.4) for image acquisition and reconstruction, which aligns with RQ 1.1. This imaging protocol provides a practical guide for researchers interested in reusing unsupervised DL denoising with quality assurance. It includes step-by-step instructions for image capture, denoising, and quality evaluation, helping researchers achieve the best possible outcomes in various imaging scenarios while maintaining reliability and accuracy in their work.

3.2.1. Evaluating structural reliability with SSIM

In Fig. 3.2, we investigate how SSIM (introduced in Section 2.3.2.1) value changes with altering microscope settings and how these settings affect the quality of the images. Sample images are the DNA distribution in the nucleus of a zebrafish embryo obtained via STED microscopy.

Influencing image quality with SSIM To establish a robust imaging protocol for quality-assured DL denoising, we analyzed how the duration of photon collection (essentially the time spent gathering light) affects image quality, measured using SSIM. While many parameters could influence image quality, photon collection duration serves as a representative example to demonstrate the relationship between imaging settings and structural reliability.

In one approach, the total amount of light collected is summed without adjusting for the time taken, resulting in changes to how structural details are captured as the photon collection time

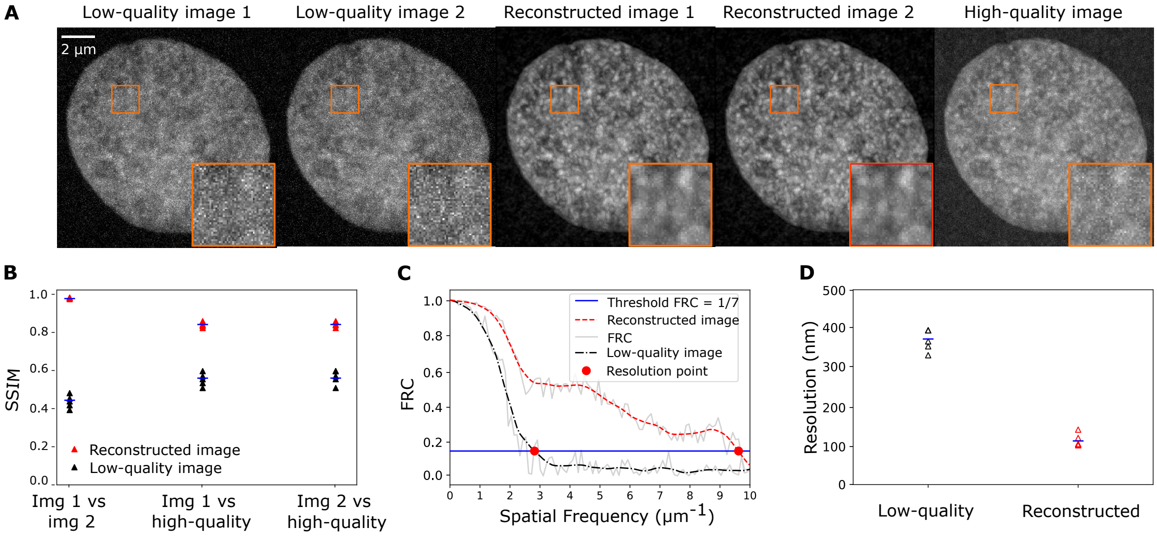


Figure 3.1.: Metrics for the reliability and effective resolution in Noise2Void-reconstructed images. A) Representative micrographs of the DNA distribution in a nucleus in a fixed zebrafish embryo, recorded with a stimulated emission depletion (STED) super-resolution microscope. The same image plane was recorded twice at low quality, once at high quality, and two Noise2Void-reconstructed images were prepared from the low-quality images. The panel shows examples of low-quality images, reconstructed images, and high-quality images used for comparison. B) SSIM values for pairwise comparisons are shown. The first column represents comparisons between Image 1 and Image 2 across 5 different nuclei (black triangles). The second and third columns show comparisons between low-quality images and their corresponding high-quality images (black triangles), as well as reconstructed images and high-quality images (red triangles), across the same 5 nuclei. C) FRC curves are calculated using a low-quality image pair and the corresponding reconstructed image pair. D) FRC-based effective resolution for low-quality images and their corresponding reconstructed images, averaged across 5 nuclei.

increases (Fig. 3.2B). In another approach, the total light is averaged over the collection time, so brightness is normalized, and structural details improve more steadily as the collection time increases (Fig. 3.2C). The first method captures structural variations more clearly, while the second method ensures smoother improvements by balancing light collection over time.

By understanding how these two methods of light collection affect SSIM, we show how to optimize imaging settings for better quality, which is crucial for creating reliable workflows in image denoising and reconstruction.

Assessing the impact of denoising on image quality As part of validating our proposed imaging protocol for quality-assured DL denoising, we evaluated how N2V denoising can enhance the structural quality of images. This step is particularly relevant for ensuring that denoising methods can be effectively reused in the image processing pipeline to improve the SNR of low-quality microscopy images.

In Fig. 3.1A,B, we show how effective denoising is in improving the structural quality of images. We applied N2V denoising to enhance the quality of super-resolution microscopy images. SSIM is also a useful metric for checking if reconstructions of low-quality images, taken under conditions like low exposure times, closely match high-quality images.

Figure 3.1A presents micrographs showing DNA distribution in the nucleus of a zebrafish embryo. These images were captured using a STED super-resolution microscope. We intentionally recorded the images twice because the FRC metric, which we use to measure spatial resolution, requires two images of the same object. The two low-quality images were then reconstructed using N2V.

Figure 3.1B shows SSIM values for pairwise comparisons (image 1 vs. image 2) and comparisons with the high-quality image (image 1 vs. high-quality, image 2 vs. high-quality). These comparisons are made for both the low-quality images and the N2V-reconstructed images. The results indicate that N2V processing improves structural similarity.

Augmenting SSIM with local SSIM However, relying solely on SSIM to assess image reliability might overlook localized differences between images, especially those introduced at edges during denoising procedures. To overcome this limitation, it is important to complement SSIM with local SSIM, specifically designed to detect and evaluate these localized discrepancies [172].

In Fig. 3.3, we compare reconstructed images obtained using N2V processing with high-quality images of nuclei from human cheek cells labeled with Hoechst 33342.

- **Panel A** displays representative micrographs showing nuclei with DNA labeling acquired at different exposure times (t_{exp}).
- **Panel B** shows the corresponding Noise2Void-processed images derived from the data in Panel A.
- **Panel C** exhibits high-quality images acquired at the same position but with a longer exposure time (t_{exp} = 200 ms), serving as a reference for accurate image quality.
- **Panel D** presents the Local SSIM map, illustrating the comparison between the reconstructed and high-quality images.
- **Panel E** shows the Local SSIM map for comparing two high-quality images acquired at the same position, indicating no structural mismatch in the area of interest.
- **Panel F** highlights an example of a faulty reconstruction, indicated by a structural mismatch within the area of interest.

In **Panel G**, the average SSIM values are calculated based on the 5% lowest local SSIM values of the 5% brightest pixels. This approach focuses on the most relevant regions of the images, such as areas with high intensity and detailed structural features, while ignoring less important background pixels. For exposure times of 20 ms and 40 ms, the "Reconstructed vs high-quality" values are generally higher than the "Low-quality vs high-quality" values. This indicates that Noise2Void (N2V) denoising improves the structural similarity of low-quality images, making them more closely resemble the high-quality reference.

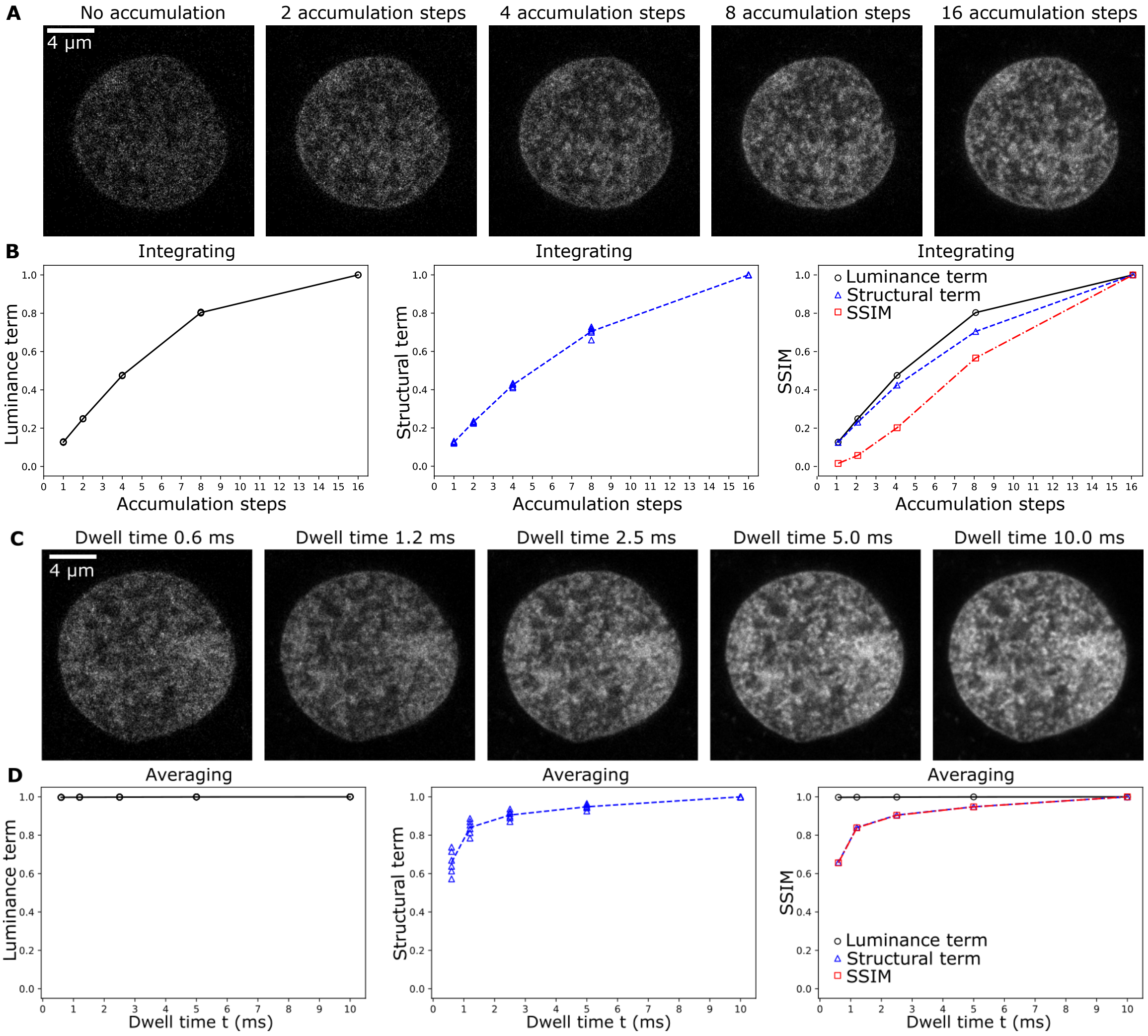


Figure 3.2.: Structural similarity is a metric of proper structure. Figures show the DNA distribution in a mid-section of a nucleus in a fixed zebrafish embryo, obtained by stimulated emission depletion (STED) microscopy. A) Depending on the microscope detector type and settings, photons can be accumulated over time without normalization (integrating detector) or with normalization (averaging detector). To illustrate how the different terms of the SSIM depend on the duration of photon collection in the integrating type detector, images were acquired with line-repeat scans in the accumulation mode, using increasing numbers of line repeats $(1, 2, 4, 8, 16)$. B) In the integrating detector case, the luminance (mean) as well as the structural (covariance) term of the SSIM depend on the number of accumulation steps. Overall SSIM values contain contributions from both terms, so that changes in overall image intensity would obscure an assessment of structural reliability. SSIM values were calculated based on $n = 6$ images obtained by a reduced number of accumulation steps and a reference image obtained with the highest number of accumulation steps (16). Individual values are shown with the mean. C) To illustrate how the different terms of the SSIM metric depend on the duration of photon collection in the averaging detector type, DNA images were acquired with the detector left open to collect photons at each pixel for longer times (dwell time t), then the photon count is normalized by the dwell time t. D) In the integrating detector case, the luminance term is constant and close to 1.0. Only the structural term changes with increasing t, so the overall SSIM values also directly reflect structural reliability for a given t. SSIM values were calculated based on $n = 6$ images obtained for a given t, combined with a matching image recorded with the highest $t = 10$ ms. Individual values are shown with the mean.

However, the observation that at 70 ms, the "Reconstructed vs high-quality" values are not higher than "Low-quality vs high-quality" suggests that the denoising algorithm may reach a limit in its ability to enhance image quality when the original low-quality image is already relatively clear. In such cases, further denoising provides minimal improvement because the structural information in the low-quality image is sufficient for accurate comparison with the high-quality reference.

The results demonstrate that Noise2Void (N2V) reliably enhances structural similarity, particularly for shorter exposure times (e.g., 40 ms). By improving the match between reconstructed and high-quality images, local SSIM confirms that the denoising process can produce reliable and accurate reconstructions while maintaining detailed structures.

3.2.2. Quantifying effective resolution with FRC

Another crucial aspect of performance in microscopy is the effective image resolution. The effective image resolution is determined by the optical resolution of a given imaging instrument and by the ratio of photons emitted by the structure of interest over polluting photons, often referred to as SNR. This effective resolution can be quantified via FRC (introduced in Section 2.3.2.2) [170, 9]. FRC evaluates the similarity of a pair of images in frequency space to determine the spatial frequency up to which the images are consistent.

In Fig. 3.4A, representative micrographs depict the DNA distribution within mid-sections of nuclei in a fixed zebrafish embryo, captured using stimulated emission depletion (STED) microscopy. Two low-quality images (1 and 2) were acquired with identical acquisition settings. Subsequently, two reconstructed images (1 and 2) were generated using N2V reconstruction from the corresponding low-quality images. Additionally, a high-quality image was obtained in the same scanning sequence as the low-quality images but included accumulation by repeated line-scanning to enhance image quality.

Analyzing effective resolution enhancement Fig. 3.4B shows the results of FRC analysis, which evaluates the improvement in effective resolution of reconstructed images compared to unprocessed low-quality images. FRC works by analyzing the similarity of two images in frequency space and identifies the spatial frequency at which the correlation between the images falls below a defined threshold. This spatial frequency marks the limit of reliable resolution, and its inverse provides the effective spatial resolution.

The FRC analysis applied to the super-resolution microscopy data in Fig. 3.4B demonstrates that N2V denoising enhances effective resolution. By restoring fine structural details that are otherwise unclear in low-quality images.

Taken together, SSIM and FRC assess image structural reliability and effective resolution.

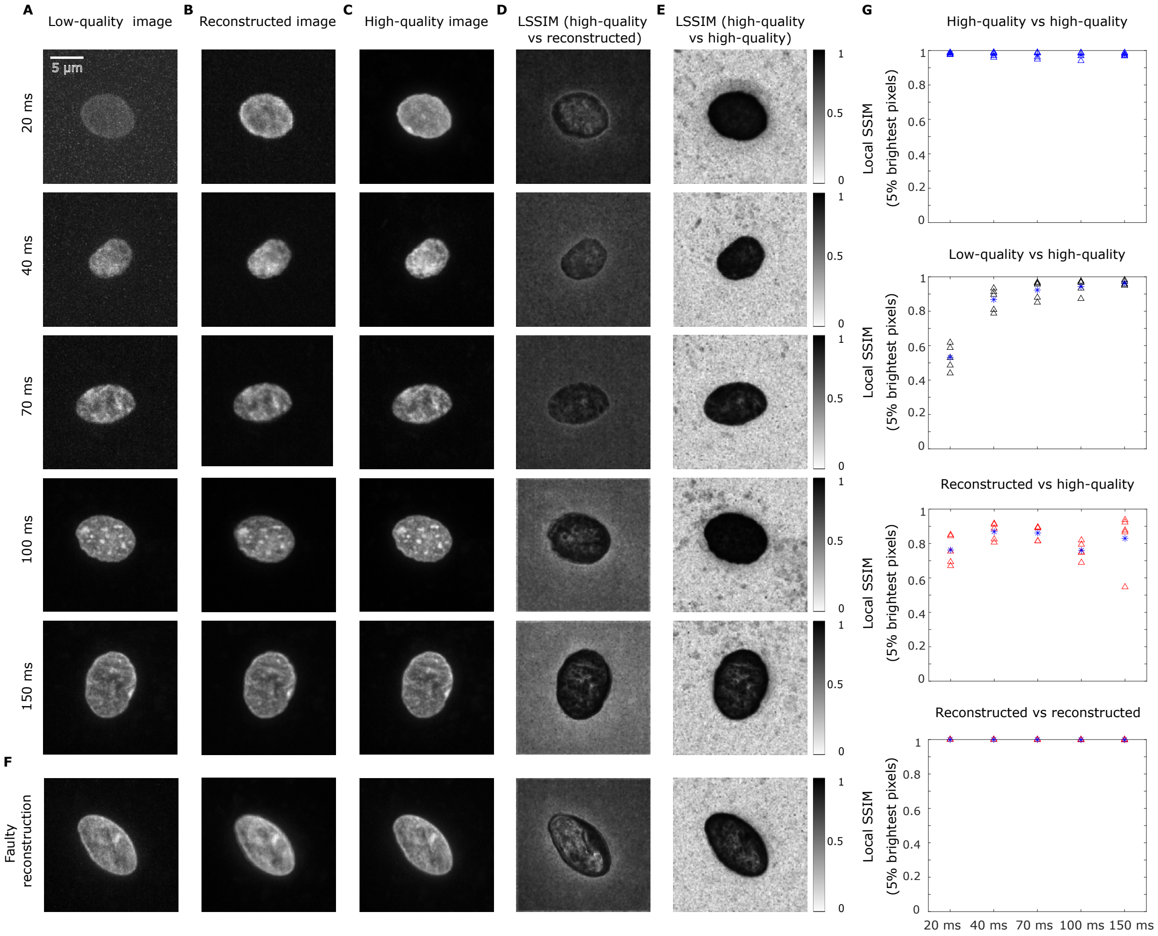


Figure 3.3.: Local structural similarity index metric can show faulty reconstructions. A) Representative micrographs of nuclei of human cheek cells in which DNA was labelled by Hoechst 33342. Images are maximum intensity projections of full volumetric stacks acquired with different exposure time (t_{exp}) as indicated B) Noise2Void-processed images corresponding to panel A. C) High-quality images acquired at the same position, but with $t_{exp} = 200$ ms. D) Local SSIM map for comparing reconstructed and high-quality images. E) Local SSIM map for comparing two high-quality images acquired at the same position, suggesting no structural mismatch in the area of interest. F) An example of a faulty reconstruction, indicated by a structural mismatch inside the area of interest. G) Average SSIM values based on 5% lowest local SSIM value of the 5% brightest pixels. $n = 4, 5, 5$ values from $N = 5$ different nuclei.

3.2.3. Guiding exposure time selection in high-speed live imaging

A key challenge in high-speed live imaging is finding the balance between capturing sufficient signal intensity and minimizing phototoxicity. To address this, the proposed imaging protocol combines SSIM and FRC metrics to determine the optimal exposure time, ensuring both structural similarity and effective resolution in denoised images (RQ 1.3). We illustrate this process through a case study, demonstrating how these metrics guide the selection.

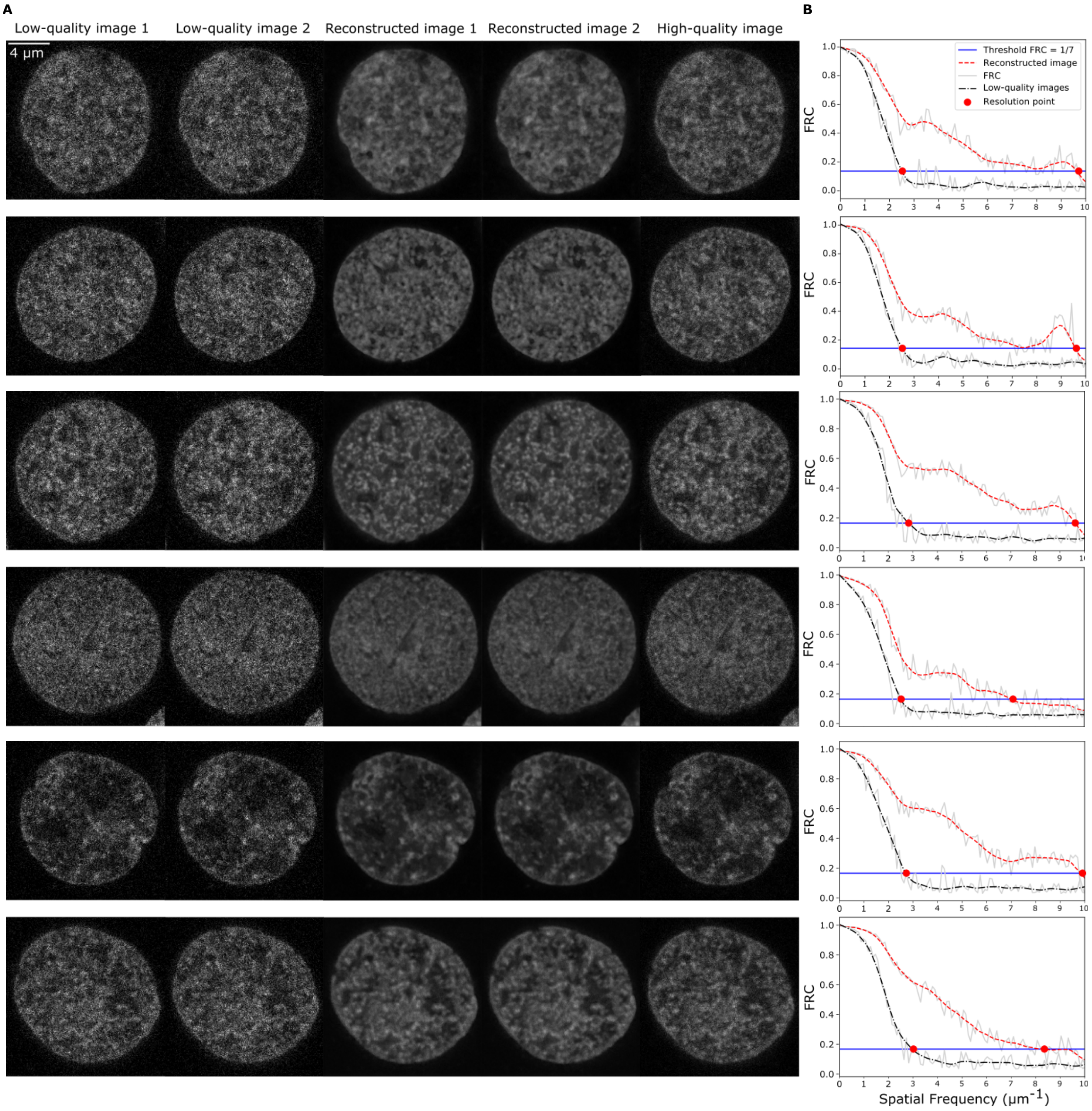


Figure 3.4.: Fourier ring correlation can quantify improvements in effective image resolution obtained by Noise2Void reconstruction. A) DNA distribution in mid-sections of nuclei in a fixed zebrafish embryo, obtained by stimulated emission depletion (STED) microscopy. Low-quality images 1 and 2 are acquired with identical acquisition settings. Reconstructed images 1 and 2 are obtained by Noise2Void reconstruction from the low-quality images 1 and 2, respectively. The high-quality image was acquired in the same scanning sequence as the low-quality images but included accumulation by repeated line-scanning to improve image quality. B) Fourier ring correlation (FRC) analysis to determine the improvement in effective resolution of reconstructed images relative to the unprocessed low-quality images.

Fig. 3.5 shows live sample microscopy of human cheek cells recorded for a range of exposure times, specifically $t_{\mathrm{exp}} = 20, 40, 70, 100, 150\,\mathrm{ms}$, to determine which exposure time should be selected.

First, for each exposure time, a separate N2V network was trained on the images to assess the effective resolution of the reconstructed images. As depicted in Fig. 3.5A-B, the FRC analysis revealed that an exposure time of 70 ms or longer achieved an effective resolution of approximately 200 nm (corresponding to a spatial frequency of $5\,\mu\mathrm{m}^{-1}$, shown in Fig. 3.5C). Shorter exposure times, such as $t_{\mathrm{exp}} = 20\,\mathrm{ms}$, did not achieve this level of resolution, while increasing the exposure time beyond 70 ms provided no additional resolution enhancement. This indicates that 70 ms is the optimal exposure time for maximizing effective resolution in this experiment.

In addition to resolution, the structural reliability of the reconstructed images was evaluated using local SSIM metrics. The analysis in Fig. 3.3A-G revealed that for $t_{\mathrm{exp}} = 20\,\mathrm{ms}$, noticeable reconstruction errors occurred, indicating insufficient exposure time to maintain structural similarity. At $t_{\mathrm{exp}} = 40\,\mathrm{ms}$ or higher, local SSIM values confirmed that the reconstructed images closely matched the high-quality reference images, demonstrating structural reliability. Interestingly, the local SSIM of the 5% brightest pixels in reconstructed images for $t_{\mathrm{exp}} = 70\,\mathrm{ms}$ was comparable to that of the low-quality images, suggesting no further improvement in structural similarity beyond 40 ms.

To conclude, FRC analysis showed that $t_{\mathrm{exp}} = 70\,\mathrm{ms}$ was necessary to achieve maximal effective resolution of 200 nm, while local SSIM indicated that exposure times as low as 40 ms provided sufficient structural reliability. Thus, experimenters can choose $t_{\mathrm{exp}} = 40\,\mathrm{ms}$ for faster acquisition with reliable structural similarity or $t_{\mathrm{exp}} = 70\,\mathrm{ms}$ for higher resolution, depending on their experimental requirements.

3.2.4. Two-phase acquisition protocol for high-speed live imaging

We propose an acquisition protocol to streamline the reuse of quality-assured N2V-processed images into high-speed live data recording. This imaging protocol consists of two distinct phases executed at each position within a sample. It is important to note that this imaging protocol allows substituting any other DL-based denoising method without requiring changes.

The two-phase acquisition protocol, depicted in Fig. 3.6, shows how we synchronize the acquisition of quality control images with high-speed live imaging.

- **Panel A:** Image data acquisition at multiple positions within a sample, capturing various viewpoints containing objects of interest, such as nuclei, marked by circles. These positions span different depths along the z-axis (*z positions*), which refer to the vertical positions within the sample.
- **Panel B:** The sequence of the two acquisition phases for each position.

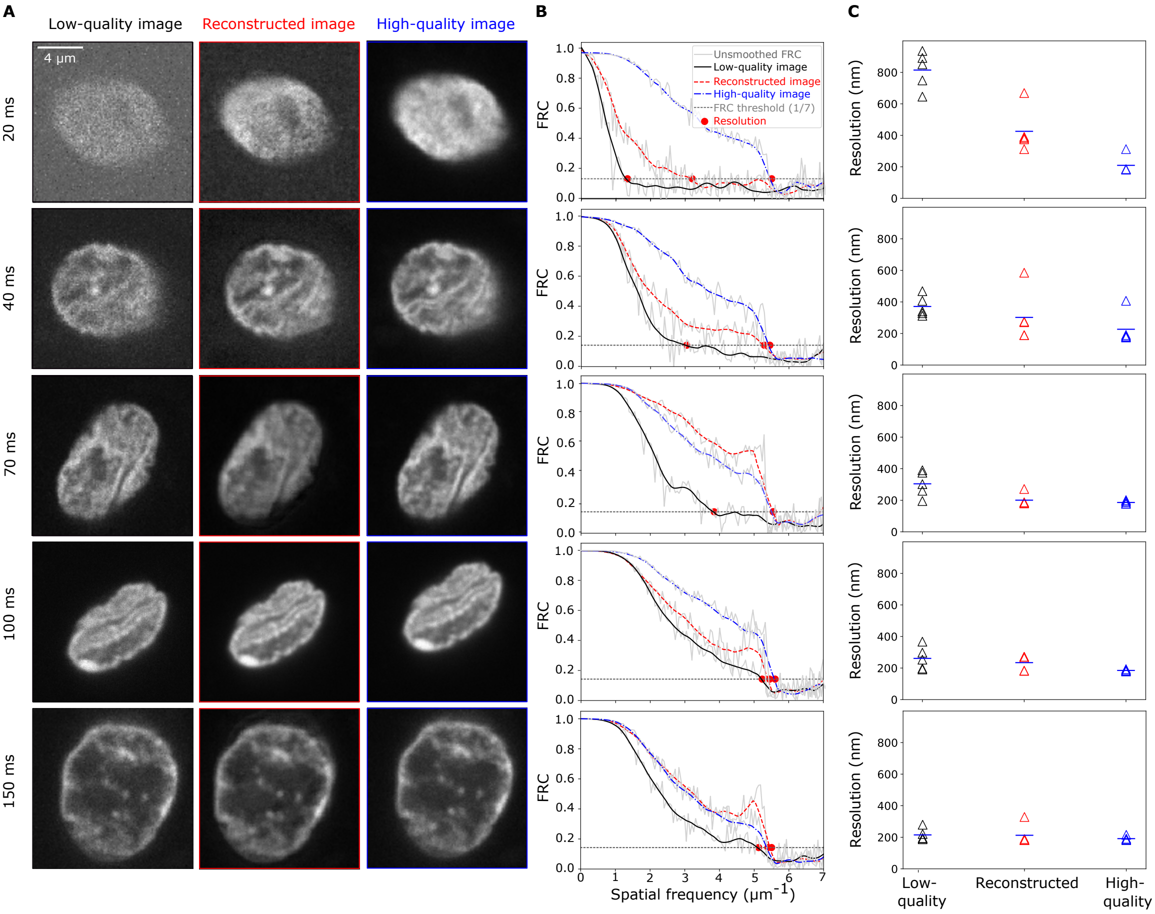


Figure 3.5.: Metric-based estimation of how far image quality can be compromised while still allowing recovery of effective resolution by denoising. A) Representative micrographs of nuclei of human cheek cells for different camera exposure times (t_{exp}, as indicated), all high-quality images were acquired at the same position but with an exposure time of 200 ms. Images are maximum-intensity projections, DNA was labelled by Hoechst 33342. B) FRC curves calculated from a pair of matched low-quality, reconstructed, and high-quality images for the different t_{exp}. C) Effective resolution for the indicated t_{exp}, $n = 5$ nuclei per t_{exp}, values are shown with mean.

In **Phase A**, images for each z position are captured. This includes a low-quality image taken at a shortened exposure time ($t_{\mathrm{exp},i}$, which may vary for different experimental conditions), two high-quality reference images captured at a standard exposure time (t_{ref}), and two additional low-quality test images. These acquisitions are important for generating data to train N2V models and to assess image quality using SSIM and FRC metrics (Fig. 3.6B).

In **Phase B**, a high-speed time-lapse series is recorded exclusively with low-quality images acquired at shortened exposure times ($t_{\mathrm{exp},i}$). These images are not processed immediately but are reconstructed later based on the optimal exposure time determined from Phase A. While Phase A provides all the data required to evaluate image quality across different exposure times, Phase B prioritizes rapid acquisition for live imaging. All data for both phases are recorded upfront without interruption, ensuring continuous acquisition during live imaging.

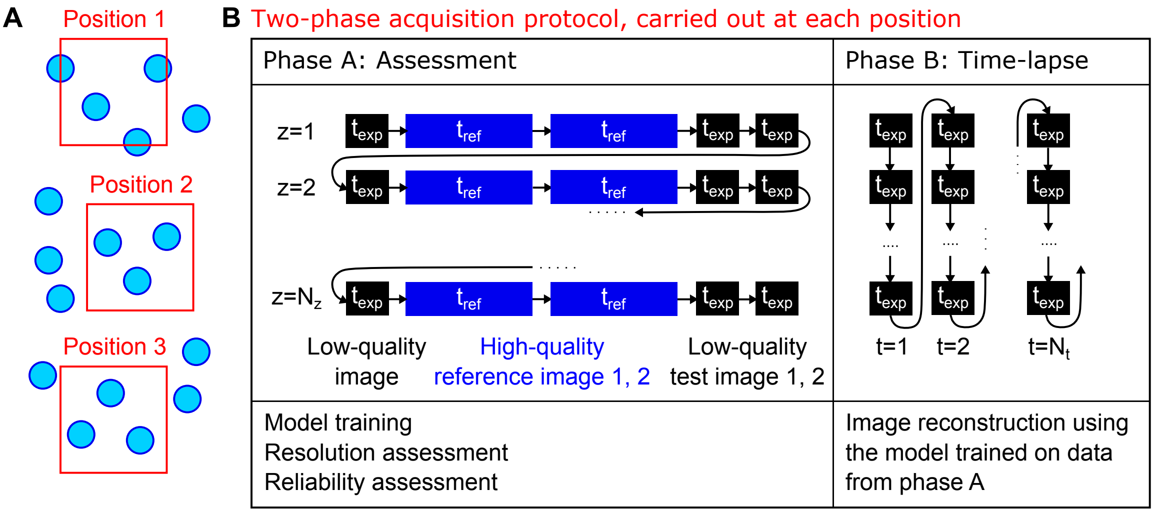


Figure 3.6.: A two-phase acquisition protocol to combine the acquisition of quality control images with high-speed live imaging. A) Image data were acquired at multiple positions in a sample, spanning different depths along the z-axis (*z positions*), to obtain multiple viewpoints containing several objects of interest (nuclei, indicated as circles). B) A sequence of two acquisition phases is carried out for each position. In Phase A, for each z position, a low-quality image, two high-quality reference images, and two low-quality test images are recorded. Low-quality images are recorded at shortened exposure times ($t_{\mathrm{exp},i}$), and high-quality images are recorded at a reference exposure time (t_{ref}). Phase A obtains the images required for Noise2Void model training and assesses effective resolution and reconstruction errors across various exposure times. In Phase B, a high-speed time-lapse series of low-quality images is recorded exclusively at shortened exposure times ($t_{\mathrm{exp},i}$). These Phase B images are reconstructed later using the optimal exposure time determined from the Phase A analysis.

After the experiment, we process the Phase A data to train the N2V network, assess the effective resolution with FRC, and evaluate reconstruction accuracy with SSIM and local SSIM. Once the optimal exposure time ($t_{\mathrm{exp,opt}}$) is identified, the N2V network is applied to reconstruct only the Phase B time-lapse stacks corresponding to $t_{\mathrm{exp,opt}}$. This approach focuses on keeping the processed data reliable and accurate while working within the limits of high-speed live imaging.

3.3. Evaluation

We evaluate the proposed imaging protocol and its components through a two-part approach.

1. How much improvement can biologists achieve by adopting the proposed imaging protocol over traditional methods such as the 3D Richardson-Lucy deconvolution algorithm?

   Biologists often hesitate to use unsupervised DL methods, fearing that such approaches may introduce artifacts or fail to reliably reconstruct structures of interest. Without the proposed protocol, biologists would rely solely on traditional methods, such as the

   3D Richardson-Lucy deconvolution algorithm, which they perceive as more reliable due to their long-standing use and clear integration into microscopy tools despite their limitations. This evaluation demonstrates how much improvement biologists could achieve by using our protocol, compared to relying only on traditional methods.

2. How does the complete imaging protocol enhance the analysis of molecular dynamics in a real case study?

   In the second evaluation, we apply the complete imaging protocol in a real case study, where it is used to study molecular dynamics and cluster morphology in Pol II enriched regions. This case study demonstrates the broader utility of the protocol in live imaging scenarios, supporting novel biological discoveries.

3.3.1. Evaluation against traditional deconvolution techniques

A challenge in the adoption of unsupervised DL denoising methods in microscopy is establishing trust in their reliability. In the absence of the proposed imaging protocol, biologists tend to rely on traditional methods like the 3D Richardson-Lucy deconvolution algorithm, which they consider safer due to its long-standing use and clear integration into microscopy tools. However, these conventional approaches have limitations, particularly in achieving high resolution. This comparison highlights the advancements that biologists could achieve by utilizing our proposed imaging protocol versus relying on traditional methods.

The purpose of this comparison is to evaluate the extent of improvement biologists could achieve by adopting the proposed imaging protocol. Specifically, we compare the performance of N2V within the protocol to the traditional deconvolution approach. For this evaluation, we use a dataset comprising low-quality images, N2V-reconstructed images, high-quality reference images, and 3D deconvolved images. The dataset consists of images from zebrafish embryos, a commonly used model for assessing microscopy methods.

To quantitatively assess the resolution improvements, we use the FRC metric, which measures the effective resolution of each set of images. Specifically, we conduct FRC analysis on low-quality images, N2V-reconstructed images, high-quality reference images, and deconvolution-processed images. For each condition, we calculate the mean effective resolution, using data from $n = 4, 4, 3$ samples across $N = 3$ different embryos.

By comparing the effective resolution results from these different approaches, we aim to demonstrate how biologists can benefit from adopting our imaging protocol. Without the proposed imaging protocol, biologists would continue to rely on traditional methods like Richardson-Lucy deconvolution due to concerns about the reliability of DL-based denoising methods. This analysis shows that by using the proposed protocol, which ensures quality-assured reuse of N2V, biologists can confidently adopt N2V and achieve better resolution compared to traditional methods.

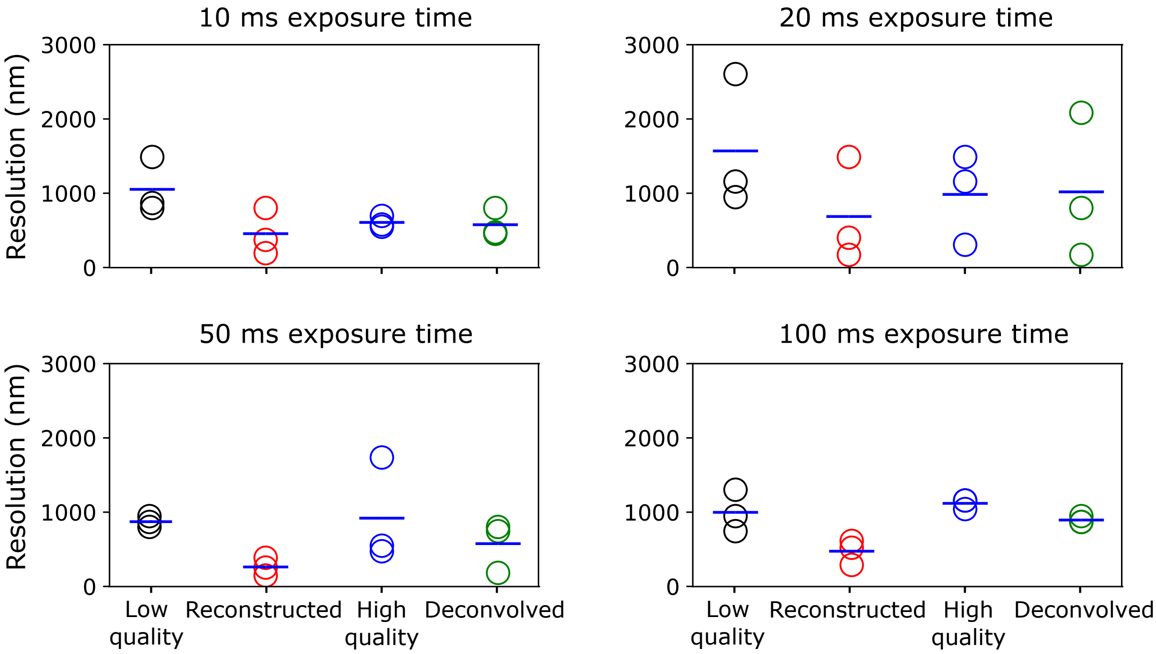


Figure 3.7.: Comparison of improvement of effective resolution by Noise2Void and conventional 3D deconvolution. Effective resolution as determined by FRC analysis for low-quality images, reconstructed images, high-quality images and 3D deconvolved images for the indicated t_{exp}. $n = 4, 4, 3$ values from $N = 3$ different embryos are shown with the mean. Deconvolved images were obtained by Richardson-Lucy 3D deconvolution.

3.3.2. Case study: real-time visualization of molecular dynamics and cluster morphology in Pol II enriched regions

We conducted a real case study to demonstrate the effectiveness of our new quality-controlled, N2V-supported live imaging protocol. In this specific study, our proposed imaging protocol made imaging tasks five times faster. We visualized changes in the molecular state and the 3D shape of macromolecular clusters enriched in Pol II, revealing completely new information in these structures. This demonstrates how our imaging protocol enables detailed and rapid understanding of complex biological structures (some foundational biological terms are introduced in Section 2.7). Here, all the steps come with details.

3.3.2.1. Experimental setup

For this purpose, microscopy imaging on live zebrafish embryos using an instant-SIM microscope is conducted [184]. To visualize Pol II recruitment to macromolecular clusters (Pol II Ser5P) or its transition towards RNA transcript production (Pol II Ser2P), fluorescently labeled antibody fragments (Fabs) is utilized. These Fabs have been validated to specifically detect changes in Pol II Ser5P and Pol II Ser2P levels in zebrafish embryos without affecting embryonic development [151, 70, 133].

3.3.2.2. Determination of exposure times

The imaging parameters are initially adjusted to reveal cluster shapes in the Pol II Ser5P channel on the microscope's live display without any processing, allowing us to determine suitable exposure times. A reference exposure time of $t_{\text{ref}} = 200\,\text{ms}$ is established, providing images of the desired quality and requiring a total time of 6 s to obtain a full 3D image stack. The experiment then proceeds with the proposed two-phase acquisition protocol, which is specifically designed for uninterrupted live imaging.

In **Phase A**, images are captured at various exposure times ($t_{\text{exp}} = 10, 20, 50, 100, \text{ms}$) to generate the data needed for training the N2V network and evaluating image quality metrics such as SSIM and FRC. Then, in **Phase B**, a high-speed time-lapse series is recorded using the same exposure times ($t_{\text{exp}} = 10, 20, 50, 100, \text{ms}$). Both phases are recorded continuously without interruption to maintain the continuity of live imaging and avoid interfering with the biological processes being studied, which is essential for observing dynamic processes in real time.

After recording both phases, the data from Phase A is analyzed to determine the optimal exposure time ($t_{\text{exp,opt}}$). Once $t_{\text{exp,opt}}$ is identified, only the images from Phase B corresponding to $t_{\text{exp,opt}}$ are reconstructed using the N2V network. This approach ensures that the most reliable and high-quality stacks from the Phase B time-lapse series are used for live imaging analysis.

Fig. 3.8 panel A shows representative micrographs of recruited RNA polymerase II (Pol II Ser5P) in live sphere-stage zebrafish embryos. Each image is a single image plane acquired with different exposure times (t_{exp}), which are indicated. Panel B displays the corresponding images after N2V-based reconstruction. This demonstrates how the N2V algorithm can recover the SNR in the images post-acquisition, even when compromised during acquisition with shorter exposure times.

Panel C exhibits corresponding high-quality reference images captured with an exposure time (t_{exp}) of 200 ms. These reference images serve as benchmarks for assessing the effectiveness of the reconstruction process.

Panel D illustrates the effective resolution, determined using FRC analysis, for three categories of images: low-quality images, reconstructed images processed with N2V, and high-quality reference images, at various exposure times (t_{exp}). For all t_{exp}, reconstructed images achieve an effective lateral resolution of 400 nm or better after N2V-based reconstruction. This is a significant improvement compared to the raw, low-quality images and is even better than the effective resolution of approximately 700 nm observed in the high-quality reference images. These results highlight the ability of N2V to enhance the resolution of noisy, low-quality data to a level that surpasses the resolution of traditionally acquired high-quality images.

Panel E demonstrates how well-reconstructed images match high-quality reference images using the SSIM as a metric. The analysis focuses on the 5% lowest SSIM values from the 5% brightest parts of the image, which are critical for analysis since they contain important details like RNA polymerase II signals. The SSIM values are compared across three scenarios: reconstructed images versus high-quality images, high-quality images versus each other, and reconstructed images versus other reconstructed images. The Y-axis of the panel represents the SSIM values, with higher values indicating better similarity, while the X-axis represents different exposure times (t_{exp}) of 10 ms, 20 ms, 50 ms, and 100 ms.

The data shows that for high-quality images compared to each other (blue triangles), SSIM values are consistently high, reflecting their excellent agreement. For reconstructed images compared with high-quality images (red triangles), the SSIM values are low at $t_{\text{exp}} = 10$ ms, indicating that N2V struggles to effectively reconstruct images taken with very short exposure times. However, at $t_{\text{exp}} \geq 20$ ms, the SSIM values for these comparisons approach those of high-quality images, showing that N2V can restore image quality well for longer exposure times.

For black triangles, two scenarios are represented: low-quality images compared to other low-quality images and reconstructed images compared to other reconstructed images. For low-quality images, SSIM values are consistently low across all t_{exp}, showing poor similarity between repeated acquisitions of noisy data. However, for reconstructed images, SSIM values are similarly high at $t_{\text{exp}} \geq 20$ ms, reinforcing the reliability of N2V in producing consistent reconstructions under these conditions.

These findings suggest that images acquired at $t_{\text{exp}} = 10$ ms are prone to reconstruction errors and are not reliable for further analysis. In contrast, images taken at $t_{\text{exp}} = 20$ ms and $t_{\text{exp}} = 50$ ms achieve reliable reconstruction with effective resolutions of approximately 400 nm and 350 nm, respectively. These images can be used to create full 3D image stacks at time resolutions of 1 s and 2 s. Panel E confirms that N2V can successfully reconstruct images taken with exposure times of 20 ms or longer, while images with a 10 ms exposure time remain unsuitable due to poor reconstruction quality.

3.3.2.3. Image analysis

In the images, certain bright spots corresponding to Pol II Ser5P proteins were observed. These bright spots, also known as clusters, remained stable throughout phase B of the experiment [133]. To analyze these clusters, the images were first processed using N2V. After denoising, the clusters were identified and separated (segmented) in the images. Each cluster was then tracked across time-lapse frames by determining how close their positions were in consecutive images (Fig. 3.9A).

For each identified cluster, the brightness levels (intensities) of Pol II Ser5P and another protein, Ser2P, were measured at every time point. The shape of each cluster was also analyzed using two key metrics: elongation, which measures how stretched out a cluster is, and solidity, which evaluates how smooth or irregular the edges of the cluster are (Fig. 3.9B). These measurements allowed us to study how the clusters changed in size, shape, and intensity over time.

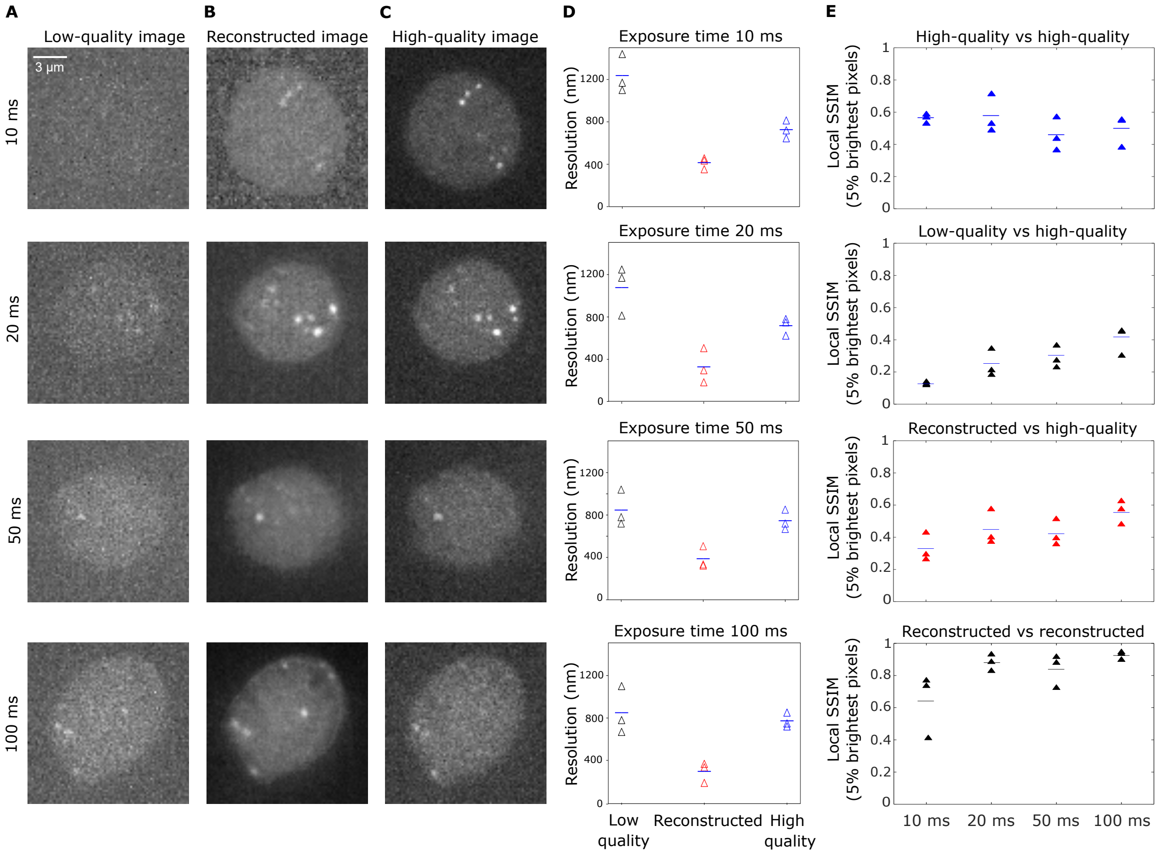


Figure 3.8.: Metric-based assessment to show how far the signal-to-noise of the image can be compromised while still allowing Noise2Void to recover signal-to-noise ratio post-acquisition. A) Example images showing RNA polymerase II clusters (proteins involved in gene expression) from zebrafish embryos, visualized using fluorescent antibody fragments (molecules that attach to specific proteins and glow under a microscope). The images were taken at different exposure times ($t_{\text{exp}} = 10$ ms, 20 ms, 50 ms, and 100 ms), which control how much light is captured per frame. Brighter, sharper images require longer exposure times but increase acquisition time. B) The same images processed using N2V. C) High-quality reference images taken with a long exposure time of 200 ms, used for comparison. D) The effective resolution of the images was quantified using FRC and is shown with the mean. The resolution is shown for low-quality, N2V-reconstructed, and high-quality images at different exposure times from $N = 3$ different embryos. E) Average SSIM values based on the 5% lowest local SSIM values of the 5% brightest pixels. $n = 4, 4, 3$ values from $N = 3$ different embryos are shown with the mean. The plot also compares the quality of high-quality images to each other (blue), low-quality images to each other (black), reconstructed images with high-quality images (red), and reconstructed images to each other (black). These results show that N2V effectively restores features in low-quality images taken at exposure times of 20 ms or longer.

The image series shown in Fig. 3.9 Panel A illustrates an example of a single RNA polymerase II cluster observed in the Pol II Ser5P channel. Each frame represents a different point in time within the time-lapse sequence, capturing the dynamic behavior of the cluster. The images were taken with an exposure time of 50 ms per frame, which enabled rapid imaging and a time resolution of 2 seconds for acquiring the full 3D volume.

Panel B provides example shapes to demonstrate how elongation and solidity are used to describe the shape of the clusters in the images.

Panel C shows time plots (time courses) for the Pol II Ser5P intensity, Pol II Ser2P intensity, elongation, and solidity of the cluster featured in the time-lapse sequence. These plots provide a detailed view of how protein intensities and cluster shapes changed over time. Cross-correlation analysis was then performed to identify any systematic relationships between these properties.

Panel D highlights the results of the cross-correlation analysis, which examines how changes in Pol II Ser5P intensity, Pol II Ser2P intensity, and solidity relate to changes in elongation over time. The gray lines show the time-shifted correlation for individual clusters, while the thick lines represent the average correlation across all clusters. The shaded gray area indicates the 95% confidence interval, showing the variability in the data.

3.3.2.4. Results and interpretation

This analysis, based on data from 30 tracked clusters in a single embryo, reveals how changes within these clusters are coordinated. The results reveal a sequence of events that happens consistently over time. First, the intensity of Pol II Ser5P increases, followed by a temporary rise in Pol II Ser2P intensity about 5 seconds later. Afterward, the Pol II Ser5P intensity decreases briefly, about 5 seconds after the Pol II Ser2P increase. At the same time, the shape of the clusters changes: the clusters initially become more rounded (indicated by an increase in solidity), followed by a temporary unfolding and elongation (indicated by a decrease in solidity) about 10 seconds later. These patterns were observed consistently across images taken with both $t_{\text{exp}} = 50, \text{ms}$ (Fig. 3.9) and $t_{\text{exp}} = 20, \text{ms}$ (Fig. 3.10), demonstrating that the findings are robust and reproducible.

Finally, Panel E provides a summary of these coordinated events, as inferred from the cross-correlation analysis. The panel depicts a typical sequence of changes: first, a brief increase in Pol II Ser5P intensity (shown in red) occurs alongside the rounding of the cluster. This is followed by a temporary increase in Pol II Ser2P intensity (shown in blue) and, subsequently, a period of cluster elongation and unfolding.

The sequence of events observed in the clusters can be explained as follows. Initially, there is an increase in Pol II Ser5P intensity, which marks the clustering of Pol II molecules in a specific region of the cell nucleus where they are preparing to activate a gene. As Pol II molecules begin their role of producing RNA (a process called transcription, where DNA instructions are copied into RNA), the Pol II Ser5P intensity decreases, and Pol II Ser2P intensity increases. This process is also accompanied by physical changes in the cluster, where the clusters stretch out and lose their rounded shape [66].

Previous studies suggest that the production of RNA by Pol II can cause these structural changes in the molecular clusters. The elongation and unfolding of clusters may result

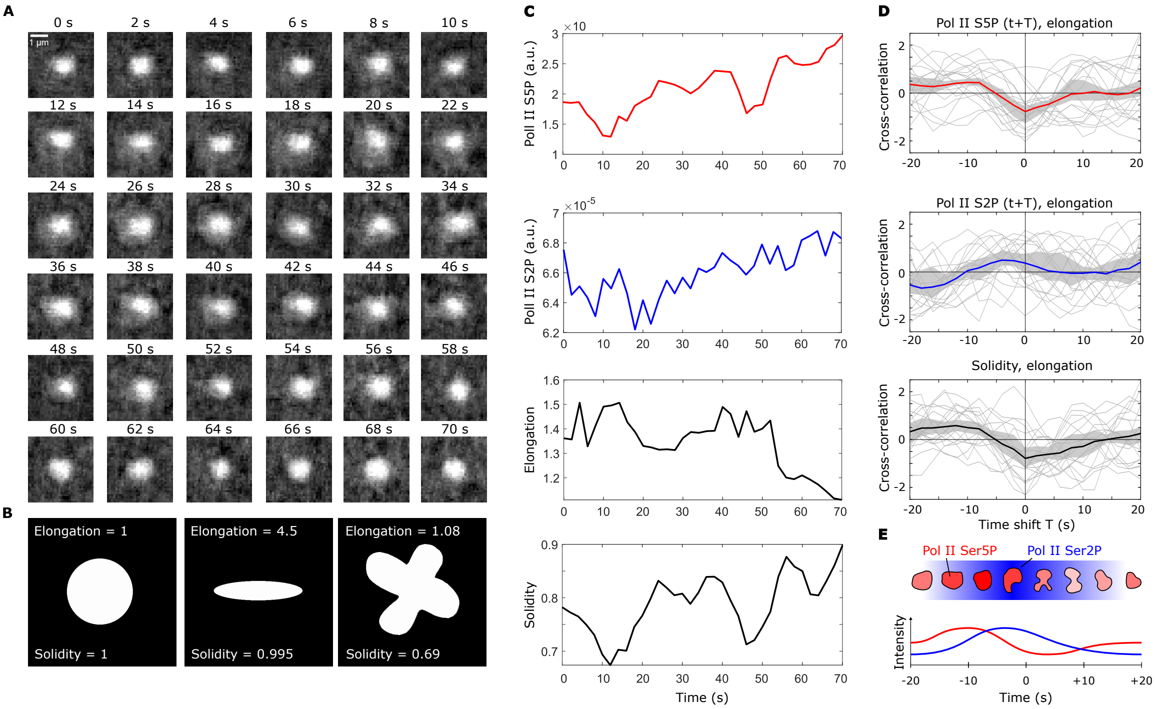


Figure 3.9.: Noise2Void-accelerated imaging reveals coordinated changes in shape and phosphorylation levels of RNA polymerase II clusters on the scale of seconds. A) Representative series of time-lapse images showing a single RNA polymerase II cluster in the Pol II Ser5P channel (single image plane from the middle z position of the cluster, exposure time t_{exp} = 50 ms, effective time resolution for full 3D volume acquisition of 2 s). The Pol II Ser2P channel is not shown because only average intensity, not shape was quantified from this channel. B) Example shapes to illustrate how elongation and solidity represent object shapes. C) Time courses of Pol II Ser5P intensity, Pol II Ser2P intensity, elongation, and solidity for the example time-lapse shown in panel A. D) Cross-correlation analysis of the temporal coordination of Pol II Ser5P intensity, Pol II Ser2P intensity, and solidity with elongation. Gray lines indicate the time-shifted correlation for single cluster time courses, thick lines indicate the mean, and the gray region is the 95% bootstrap confidence interval. Analysis based on n = 30 tracked clusters, recorded from one sphere stage embryo. E) Summary of the coordinated changes in phosphorylation and cluster shape suggested by the cross-correlation analysis. A stereotypical sequence of events can be seen: cluster Pol II Ser5P intensity transiently increases (red), and the cluster becomes rounder, then cluster Pol II Ser2P transiently intensity increases (blue), until finally, the cluster transiently unfolds and becomes elongated.

from the rearrangement of molecules within the clusters as RNA is made [70, 50, 133]. Additionally, these studies propose that changes in Pol II activity and cluster organization are linked to temporary interactions between genes and Pol II clusters. These interactions help turn genes on (gene activation), after which the genes are released from the cluster.

3.4. Discussion

In this study, we present an imaging protocol that ensures the quality-assured reuse of deep-learning-based denoising algorithms, focusing specifically on the unsupervised N2V

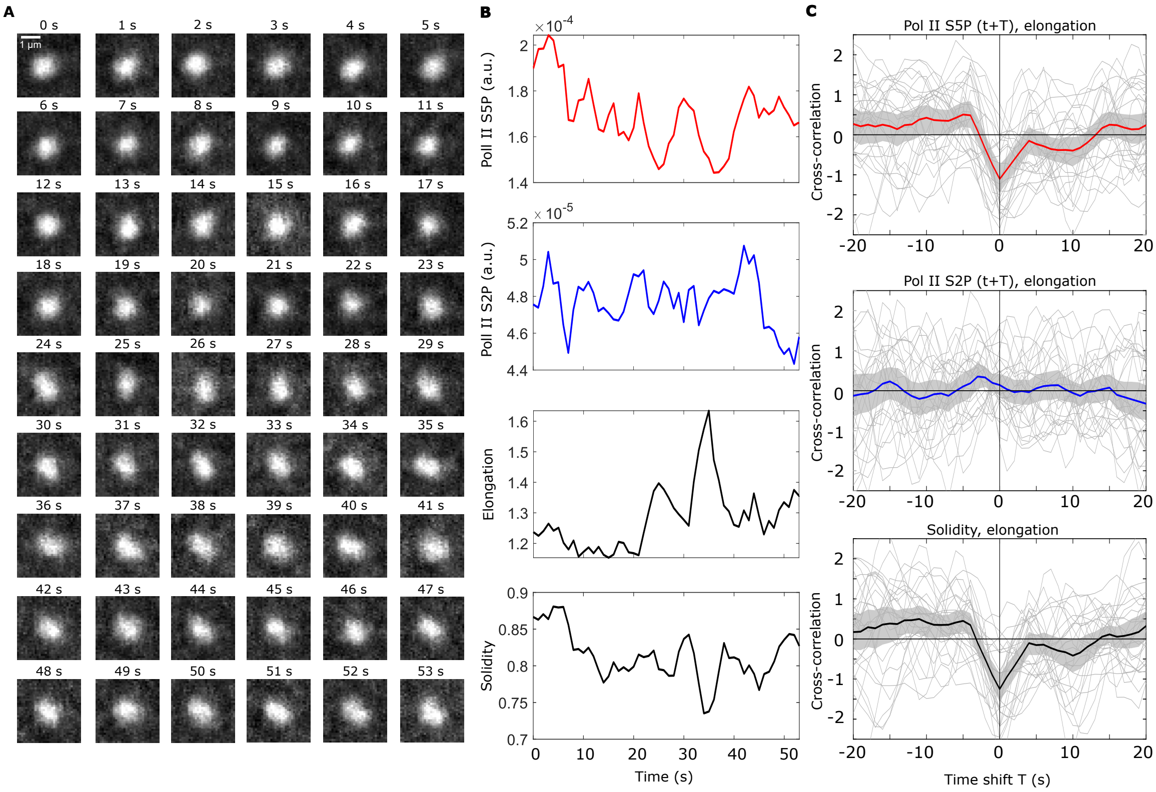


Figure 3.10.: Coordinated changes in RNA polymerase cluster phosphorylation and shape are reproduced by Noise2Void-accelerated imaging with a different exposure time. A) Representative series of time-lapse images showing a single RNA polymerase II cluster in the Pol II Ser5P channel (single image plane from the middle z position of the cluster, exposure time $t_{exp} = 20$ ms, effective time resolution for acquisition of full 3D volumes 1 s). B) Time courses of the Pol II Ser5P intensity, the Pol II Ser2P intensity, elongation, and solidity for the example track shown in panel A. C) In Cross-correlation analysis, gray lines indicate the analysis results for individual cluster time courses, thick lines the mean over all analyzed clusters, and the gray region is the bootstrap 95% confidence interval. Analysis based on $n = 27$ clusters, recorded from one sphere-stage embryo.

approach (see Fig. 3.6 for an overview). This protocol is designed to seamlessly integrate into fluorescence microscopy workflows, providing a systematic approach for optimizing imaging speed and quality. By reducing camera exposure times to accelerate image acquisition, the protocol uses N2V-based denoising to recover lost signal quality. Crucially, quality assurance steps are embedded in the protocol to validate the accuracy of reconstructed images, ensuring that the performance gains from DL are achieved with high reliability.

3.4.1. Contributions from a software engineering perspective

From a software engineering perspective, this study makes two key contributions:

Quality assurance in time-lapse image analysis: Time-lapse imaging presents unique challenges, as image quality and noise characteristics often vary between frames due to dynamic sample properties, rapid acquisition, and optical fluctuations. Our protocol addresses these challenges by providing a structured approach to assess the reliability of DL-based reconstructions. Metrics such as FRC and SSIM are used to quantitatively validate the reliability of reconstructed images.

Domain-specific challenges in quality assurance for DL: A key finding of this study is that quality assurance in DL-based reconstructions must be tailored to the specific domain of application. Unlike general-purpose imaging tasks, biological microscopy involves variable conditions, such as differences in fluorescence labeling, sample dynamics, and optical setups. These domain-specific challenges highlight why standardized quality benchmarks in SE for AI may not apply directly to scientific imaging workflows. Instead, customized quality assurance methods are essential for ensuring reproducibility and reliability in highly specialized applications.

To frame this contribution as an actionable imaging protocol, three core components are outlined:

- Dataset preparation: Noisy datasets are collected using reduced exposure times and used as inputs for self-supervised denoising models.
- Quality validation: Metrics such as FRC and SSIM quantitatively evaluate the quality of reconstructed images to ensure accuracy.
- Generalization across imaging tasks: Although this study focused on 3D fluorescence microscopy, the same quality validation methods can be adapted to other imaging techniques or dimensionalities with minimal adjustments.

This protocol tackles a major challenge in microscopy workflows: the difficulty of obtaining paired noisy and high-quality datasets required by supervised learning methods. Instead, it relies on unsupervised approaches like N2V, which reconstruct high-quality images from single noisy inputs. This flexibility is particularly important for biological imaging, where labeling techniques, sample behavior, and optical setups vary across experiments. By combining self-supervised DL with robust quality assurance, the protocol extends the usability of denoising algorithms like N2V to real-world, complex imaging tasks.

3.4.2. Findings in biological imaging

The usefulness of the proposed imaging protocol was demonstrated in the live imaging of RNA polymerase II (Pol II) clusters in zebrafish embryos. The protocol successfully recovered both molecular and shape information from noisy fluorescence microscopy datasets, discovering important details about Pol II behavior. Specifically, the high temporal resolution of this method (1-2 seconds for a full 3D stack) enabled us to capture both spatial and temporal dynamics of Pol II clusters, bridging the gap between spatial and temporal resolution that was difficult to achieve in previous studies.

Our live-sample recordings revealed a consistent sequence of events in transcriptional induction: the recruitment of Pol II to specific regions in the nucleus, followed by pause-release steps, and changes in the shape of Pol II clusters. For example, Pol II clusters initially showed an increase in Ser5P intensity, associated with clustering, followed by a decrease in Ser5P intensity and an increase in Ser2P intensity, signaling RNA production. These molecular changes coincided with cluster shape transitions from a rounded to an elongated state, emphasizing the close link between molecular and morphological dynamics during transcription.

3.4.3. Contextualizing the protocol in existing research

While previous studies have achieved either high spatial or temporal resolution, our approach integrates high resolution in both dimensions. Earlier studies of Pol II dynamics used effective time resolutions ranging from 10 seconds to 1 minute for full 3D volumes [160, 50], which limited their ability to capture rapid changes in molecular states. By comparison, our method captures 3D stacks within 1-2 seconds, providing temporal resolution sufficient to observe key transcriptional events, such as promoter escape and pause release. The kinetic models derived from our observations align with existing estimates of transcriptional dynamics, including promoter escape times (2-3 seconds) and elongation rates (approximately 1 kb per minute) [164, 82, 35]. These findings support the validity of our approach and its compatibility with previously established models of transcription.

In addition to temporal coordination, spatial relationships between Pol II clusters and nascent RNA have also been assessed. Our observations that genes undergoing elongation are located outside Pol II Ser5P clusters align with earlier studies reporting the displacement of nascent mRNA relative to transcriptional regulators [50, 104]. Unlike prior work, however, Our approach offers a detailed view of the full shape of Pol II Ser5P clusters, further advancing our knowledge of their spatial organization during transcription.

3.4.4. Extending the imaging protocol

The imaging protocol proposed here is flexible and can be adapted for a variety of imaging challenges. For example, in scenarios where photo-bleaching limits long-term image acquisition, reducing excitation light levels and recovering signals through denoising may be effective. Quality control phases could be strategically placed at regular intervals to ensure the reliability of the reconstructed data. Similarly, in experiments where sample structure or fluorescence labeling changes over time, quality control steps at both the beginning and end of the imaging process may be advisable. These extensions require no fundamental changes to the quality control methods used in this study and further broaden the applicability of the protocol.

3.5. Generalizability to other deep learning denoising models

Integration with unsupervised DL models: While our imaging protocol focuses on N2V, it can easily be adapted to work with a wide range of unsupervised DL denoising models.

The key steps and ideas in the imaging protocol are broadly applicable, no matter which specific denoising method is used. This makes the imaging protocol versatile and easy to adopt for various unsupervised techniques.

Using pre-trained models: Our imaging protocol can also work with supervised DL models that have been pre-trained on large datasets. Supervised models often deliver strong results because they are trained on extensive, diverse datasets, which can enhance their generalizability. By integrating pre-trained models into our protocol, researchers can quickly assess whether the model performs well on their specific data through the protocol's built-in quality control measures. If the pre-trained model generalizes effectively to the experimental dataset, this approach can save significant time and computational resources compared to training a supervised model from scratch. The denoising process is quicker because the model is ready to use, and our protocol ensures that any limitations in generalization are detected early, maintaining the reliability of the results.

Adapting to new DL techniques: We have designed our imaging protocol to be adaptable to new developments in DL. As new denoising methods and models emerge, they can be included in our imaging protocol. This flexible design ensures that the imaging protocol remains useful and up-to-date, even as the field of DL-based image denoising continues to grow and change.

3.6. Threats to validity

3.6.1. Threats to validity in the experimental evaluation

Algorithmic bias: The algorithms used for image processing and data analysis may have inherent biases that influence the results. For example, the N2V algorithm used for image reconstruction might favor certain patterns or structures in the data, potentially introducing biases into the reconstructed images and subsequent analyses.

Parameter sensitivity and optimization: Preprocessing algorithms often rely on parameters that must be tuned for optimal performance. If the selected parameters are suboptimal or the optimization process is biased, it could lead to inaccurate reconstructions or misinterpretations of the data.

Construct validity: Construct validity concerns whether the metrics and methods used in the evaluation accurately reflect the underlying constructs of interest. For instance, while FRC and SSIM are commonly used for assessing image quality, they might not fully capture the biological relevance of reconstructed features, potentially skewing interpretations.

External validity: External validity pertains to the generalizability of the findings. While the experimental evaluation was designed to validate the imaging protocol, the experiments were performed using zebrafish embryos under specific conditions. It is unclear whether the protocol and its findings can be directly applied to other model organisms, imaging systems, or experimental setups.

3.6.2. Threats to validity in the case study

Reliability of data collection: The case study involves live imaging of RNA polymerase II clusters, where variability in sample preparation, labeling efficiency, and fluorescence intensity may affect the reliability of the data. Small inconsistencies in these steps can introduce noise and potentially bias the results.

Subjectivity in interpretation: While quantitative metrics were used for image quality and cluster analysis, some aspects of the case study (e.g., interpreting biological relevance or morphological changes) rely on expert judgment, introducing the potential for subjective bias.

Construct validity in the case study context: The biological interpretations derived from reconstructed images assume that the denoising and quality assurance methods accurately preserve biological features. If reconstructed images fail to represent the true molecular or morphological states, the conclusions drawn from the case study could be flawed.

3.7. Related work

This section focuses on three key areas related to our study: 1) Imaging protocols for quality control in live imaging studies. 2) Strategies to bridge the gap between biologists and deep learning (DL) methods. 3) Annotation-free or weakly supervised approaches for denoising.

Imaging protocols and systematic strategies for quality control and reproducibility in live imaging

Recent research highlights the importance of imaging protocols for ensuring reliable and reproducible results in live imaging. For example, the QUAREP-LiMi initiative[1] established guidelines for quality assessment in light microscopy, focusing on standardized practices to improve data reliability [124]. However, these efforts primarily focus on hardware

---

[1] `https://quarep.org/`

calibration and standardization, leaving gaps in addressing the integration of quality control with advanced computational techniques like DL denoising.

Faklaris et al. [46] proposed simple and practical protocols to monitor the performance of light microscopes. They outlined seven methods to evaluate critical factors such as resolution, illumination stability, and stage drift. While these protocols emphasize reproducibility and reliability, they do not address computational reconstruction challenges or strategies for balancing speed and image quality, which are critical for dynamic live imaging studies.

Another study by Faklaris et al. [47] developed accessible tools and rapid acquisition protocols for assessing key performance metrics in fluorescence microscopy. While these tools enhance reproducibility, they are not designed to evaluate the artifact-free application of DL denoising methods during image reconstruction. Therefore, an imaging protocol that integrates DL-based denoising with quality assurance is needed to ensure trustworthiness in computationally reconstructed images.

Bridging the gap between biologists and DL methods

Recent efforts focus on making DL tools accessible to biologists without programming expertise. These efforts align with the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR), which aim to simplify model usage and enhance accessibility. Tools like Docker provide isolated environments with pre-installed dependencies, allowing models to be reused without complex installations. Initiatives like Multi-Modality Cell Segmentation Challenge (MMCSC) [112] further support reproducibility by sharing source code and using containers for software distribution. While these approaches simplify technical access, they do not offer specific guidance for balancing biological imaging needs, such as photodamage and acquisition speed, with the use of DL techniques.

The BioImage Model Zoo [130] exemplifies these efforts by offering a community-driven repository of pre-trained DL models for bioimage analysis. It ensures FAIR principles by standardizing model metadata and making them compatible with widely used tools like Ilastik [14], Fiji, CellProfiler [118], and QuPath [74]. The repository also supports web-based inference through the ImJoy plugin framework [131], enabling seamless access across platforms. However, some compatibility issues remain, such as models requiring unsupported software packages, which can limit usability. Additionally, these tools do not provide an integrated mechanism to assess or validate the quality of computationally reconstructed images, which our imaging protocol addresses directly.

Advances in annotation-free and weakly supervised denoising methods

Microscopy image denoising has evolved from traditional approaches to sophisticated DL-based methods. Weigert et al. [176] introduced Content-Aware Image Restoration (CARE), which trains DL models using paired noisy and high-quality images. While effective, acquiring such pairs is impractical in certain settings, such as cryo-transmission electron microscopy (cryo-TEM). To address this, Buchholz et al. [21] developed Cryo-CARE, a specialized adaptation for cryo-TEM. However, these approaches depend on the availability of specific training data, which limits their broader applicability in live imaging workflows.

Self-supervised methods such as Noise2Noise (N2N) [101], Noise2Void (N2V) [91], and Noise2Self (N2S) [10] eliminate the need for clean training images. These methods rely on statistical relationships within noisy images and have become popular for their practicality. Extensions like Self2Self (S2S) [140] use dropout mechanisms to handle complex noise types and further improve performance. However, these methods lack integrated frameworks to ensure that reconstructed images are free from artifacts or to guide their reuse in time-sensitive live imaging experiments.

Krull et al. [92] proposed Probabilistic Noise2Void (PN2V), which predicts a range of possible pixel values, offering more confidence in predictions. Additionally, Mansour et al. [115] introduced Zero-Shot Noise2Noise (ZS-N2N), a lightweight method that requires no training data or prior noise information. While these methods show promise in resource-limited scenarios, they do not address the broader integration of denoising algorithms into live imaging workflows with quality assurance measures.

Beyond these, blind zero-shot denoising algorithms [150] aim to preserve image sharpness while reducing noise without requiring clean training data. Generative models like GANs have also emerged as effective solutions. For example, Fuentes et al. [52] utilized a conditional GAN with contrastive learning to achieve high denoising performance, even with minimal training data. However, such methods often lack mechanisms for systematic quality control to ensure reliability during dynamic imaging scenarios, which our imaging protocol explicitly addresses.

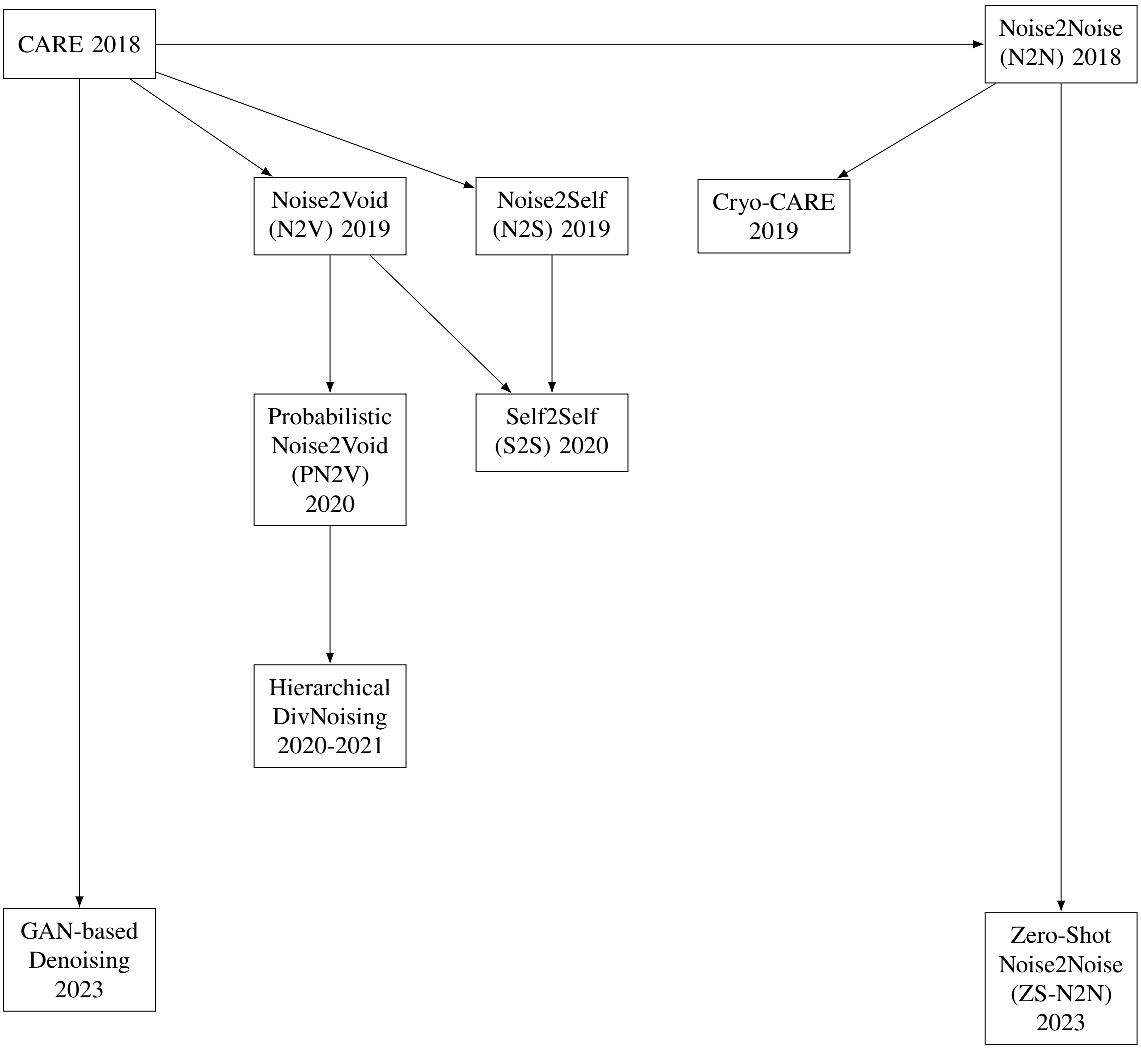


Figure 3.11.: Tree structure of DL-based denoising methods starting with Weigert et al. (2018), it shows the influence and development paths of the various methods and studies

4. Active learning to define segmentation of unknown objects

4.1. Motivation

Choosing the right ML model and reusing it in research is a challenging task. It is crucial to understand that there is no one-size-fits-all ML algorithm or parameter setup. The best solution often depends on analyzing empirical data. This challenge is even greater for researchers who are not very experienced with ML and programming. ML solutions range from flexible DL models, which can learn from raw data, to more traditional ML methods that require feature engineering without needing labeled training data [2, 78].

In the field of biological imaging, accurately separating objects from images is very important. Using ML methods for image segmentation can help scientists better understand these objects. Both traditional ML and DL-based methods have been used for this purpose. However, in many cases, getting labeled data is costly or not possible. This makes traditional ML methods more practical, especially when feature engineering is simple and doesn't need deep domain knowledge.

One effective approach for training DL models with limited labeled data is "weakly supervised" training. This approach allows scientists to label a small part of the dataset, which is then used to train the model. The trained model can then be applied to the rest of the data. Interactive weakly supervised ML tools, like WEKA [5], Ilastik [14], and nuClick [2], enable image pixel classification through user annotations using graphical interfaces. While these tools reduce the need for large training datasets compared to fully supervised DL methods, they still rely on expert labels based on visual assessments.

From a SE perspective, there is a growing need for frameworks that enable the effective reuse and adaptation of ML and DL solutions in specialized fields like biological imaging. SE for AI in biological image processing involves designing tools and methodologies that balance user-friendliness with technical robustness, making advanced ML techniques accessible to domain researchers. Addressing these challenges requires solutions that support evolving requirements, minimize manual intervention, and ensure reliable performance.

To address the challenges of limited labeled data and make reusing ML easier for researchers without programming skills, this chapter presents a new data-driven framework that uses active learning to minimize user input. This framework simplifies the training process and helps in selecting the best segmentation algorithm. It reduces the effort and expertise needed to adapt segmentation algorithms to different datasets, combining the user-friendliness of traditional methods with improved handling of complex objects.

The proposed framework starts with an unsupervised step for selecting and reusing a segmentation algorithm, followed by adjusting parameters through active learning. While the primary application is for segmenting microscopy images, this framework could also be useful in other fields where getting labeled data is difficult or expensive. Moreover, the feature engineering method used here is simple and does not require advanced algorithms or in-depth domain expertise.

The contributions of this part of the thesis include:

- **New framework for image segmentation:** This framework helps researchers reuse existing segmentation algorithms across different scientific fields, contributing to SE for AI in the domain of biological image processing [64, 65].
- **Use of an unsupervised metric:** An unsupervised metric is introduced to help with the initial selection of segmentation algorithms.
- **Mathematical model for dynamic adjustments:** A mathematical model is provided to match each object with its best segmentation setting, allowing for easier adjustments and improvements.
- **Automated clustering and quality-verified queries with user feedback:** Automated clustering combined with diverse query strategies enhances segmentation, while periodic user feedback ensures the quality and accuracy of the results.
- **Biological case study:** A detailed case study in biology demonstrates how scientists can use the framework with minimal input. This shows the framework's effectiveness in real-world applications.
- **User-focused convergence demonstration:** Additional experiments show that the framework converges effectively from the user's perspective, proving its efficiency.

By situating this work in the broader context of SE for AI, the proposed framework highlights how structured methodologies can improve the design, reuse, and adaptability of ML solutions in specialized domains. The structure of this chapter is as follows: Section 4.2 describes the scientific use case in detail, followed by a demonstration of the proposed framework in Section 4.3. Research questions, evaluation, and the experimental design are covered in Section 4.4, with results in Section 4.4.1. Possible threats to validity are discussed in Section 4.5.

4.2. Scientific use case

Biomedical image analysis involves measuring different properties of objects, like their shape, location, and intensity in experimental images. This process depends heavily on segmentation, which means identifying which image pixels (or voxels) belong to an "object"

(1) or "not-object" (0). This step is critical in the analysis workflow [178, 149]. Traditional segmentation methods, such as thresholding, use a threshold value based on the image's intensity distribution to differentiate objects from the background. In contrast, weakly supervised segmentation methods learn from partially labeled datasets provided by biologists through interactive tools [158, 14, 118, 24]. Recent advances in deep learning (DL) require large annotated datasets to achieve good results [3, 59, 31, 86, 183].

Traditional methods perform well with high-quality microscopy images where object boundaries are clear. However, they struggle with ambiguous boundaries and variations in object size, intensity, and shape (as shown in Fig. 4.1) [116, 127]. Fully supervised DL algorithms can adapt to these variations but need a lot of expert-labeled data. Weakly supervised methods reduce the need for extensive annotations but still rely heavily on expert knowledge and visual interpretation.

This chapter introduces a practical framework for customizing ML segmentation solutions for different objects. The framework starts with unsupervised segmentation methods to handle diverse object types. Then, through active learning, it adjusts (hyper-)parameters based on expert feedback for each object. This approach is particularly useful in bioimaging, where manual labeling is difficult, and reducing interaction time is important for adopting the workflow. A key component of the framework is a data-driven regressor that maps object features to segmentation algorithms and settings. This mapping adapts based on user feedback, selecting the most informative samples for further expert input.

The images shown in Fig. 4.1 are from the nuclear mid-plane of two different zebrafish embryos at the sphere stage. Both images were captured using STEDD super-resolution microscopy and show the intensity distributions of RNA Polymerase II (Pol II) Serine 5 phosphorylation (Ser5P) and Ser2P signals. The red squares in both images highlight clusters of Pol II Ser5P, which are the primary focus for detection and segmentation. Pol II Ser2P is also visible in both images but is not explicitly marked. The objects vary in shape and size, with some clusters having clear boundaries and others being much more difficult to define.

4.3. The proposed approach

This section presents a comprehensive framework (Algorithm 1) for segmenting microscopy images, leveraging active learning and data-driven methods to maximize the reuse of existing segmentation knowledge. As illustrated in Fig. 4.2, the framework outlines all key steps for achieving efficient segmentation, focusing on iterative refinement, minimal user interaction, and reuse of previously obtained insights.

The process begins with step (❶), where the object set is automatically detected and prepared by identifying object positions and extracting a bounding box around each Region of Interest (ROI), as described in Section 4.3.1. Given the variability in object shapes within the dataset, multiple segmentation settings must be considered. To handle this, the framework constructs a feature space in step (❷), as outlined in Section 4.3.2. This feature space

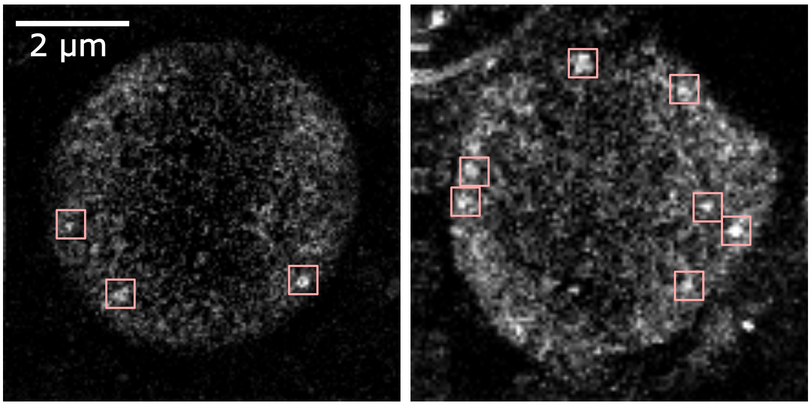


Figure 4.1.: Different object variations. The images show the nuclear mid-plane of two different zebrafish embryos at the sphere stage, recorded using STEDD super-resolution microscopy. Both Ser5P and Ser2P intensity distributions of RNA Polymerase II (Pol II) are visible in each image. The red squares highlight clusters of Pol II Ser5P, which are the primary focus for detection and segmentation. The intensity scale, adjusted from black to white, spans the 0.01-th to the 99.99-th percentile [134].

allows the framework to distinguish between different object types and facilitates the reuse of segmentation settings for similar objects.

The framework starts without any initial annotations, which means it cannot immediately select the best algorithm or settings for each object. To address this, step ❸ (Section 4.3.3) involves unsupervised iterations to automatically produce an initial estimate of suitable segmentation algorithms and settings. This estimate, however, only provides a rough approximation of segmentation quality. To fine-tune these settings, the framework uses active learning in step ❺ (Section 4.3.5), where user feedback plays a crucial role in automatically refining the segmentation approach.

Active learning in this context focuses on optimizing the reuse of user insights. It does so by selecting the most valuable data points (objects) for user feedback. The framework combines three query strategies to maximize the effectiveness of each interaction: (i) selecting samples from diverse clusters, (ii) focusing on the most informative objects, and (iii) targeting the objects with the highest uncertainty. The latter two strategies are applied during the active learning phase in step ❺ (Section 4.3.5), while clustering is performed in step ❻ (Section 4.3.6). Once feedback is gathered, the segmentation settings for those objects are automatically updated in step ❼ (Section 4.3.7), reusing the provided input to improve results across similar cases.

To ensure that objects not directly queried also benefit from user feedback, the framework uses a regression model to automate the adjustment of settings. This model creates a mapping between object features and their corresponding segmentation settings, enabling the reuse of knowledge across different objects. By using this approach, the framework generalizes user feedback to objects with similar features, making the process more efficient. After each active learning iteration, the model is retrained with updated data, allowing the refined settings to be reused for both the annotated objects and their neighbors. This iterative process continues until the framework reaches convergence, where settings stabilize, enabling efficient reuse in future tasks. Further details can be found in the "Fine Tuning" section.

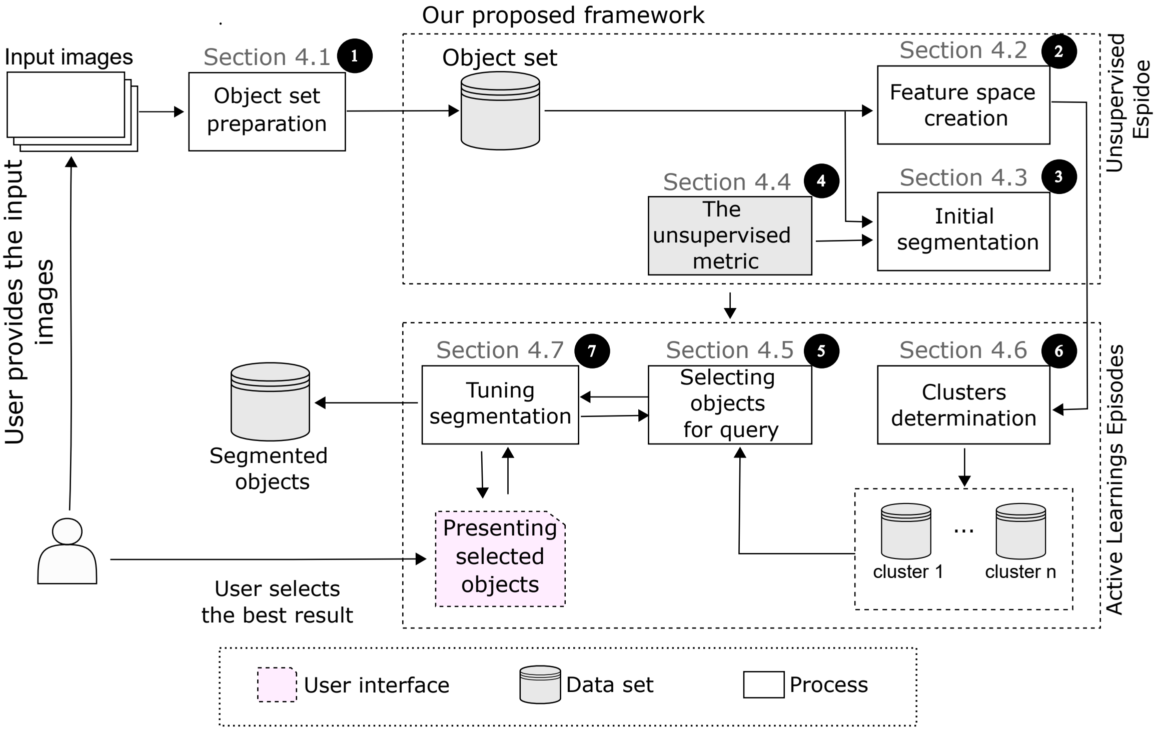


Figure 4.2.: The proposed framework consists of several automated steps. the user initially uploads the input raw images. With a few interactions, the segmented objects are given back to the user.

4.3.1. Preparing object sets

In this initial stage (step 1 in Fig. 4.2), the framework identifies specific objects for segmentation, known as Regions of Interest (ROIs), and draws bounding boxes around them. These ROIs are highlighted with red boxes in Fig. 4.1. While basic thresholding algorithms can often identify these target objects effectively, other bright subcellular components can sometimes make it difficult to detect ROIs directly (Fig. 4.1).

To overcome this, the framework first applies a blur filter to reduce noise in the image, followed by a thresholding technique to better identify the ROIs. While several types of blur filters are available, this application specifically uses a Gaussian blur filter with a standard deviation of six pixels per nucleus. After the blur is applied, a local thresholding method is used to roughly identify the ROIs. Each identified ROI is then cropped into a region of size (60×60) and saved for further analysis. Examples of these cropped objects are shown in Fig. 4.3.

Algorithm 1 The proposed segmentation framework

```
INPUT:
Input images
Base unsupervised segmentation algorithms
OUTPUT:
Optimal setting assigned to each object
Preparing object dataset according to Section 4.3.1
Initial segmentation and finding the initial estimate of the settings per object
(Section 4.3.3)
Creating the feature space according to Section 4.3.2
Finding different clusters according to Section 4.3.6
Fitting a regression model on the objects and the segmentation settings associated with
the objects (Section 4.3.7)
while not converged do
    for every cluster, select two objects (the most informative and the most uncertain
    Section 4.3.5
    Present the selected objects to the user according to Section 4.3.5
    According to the user feedback, directly adjust the segmentation settings for
    presented objects (Section 4.3.7)
    Re-train the regressor to update the associated settings for all other objects
    (Section 4.3.7)
end while
return the associated settings
```

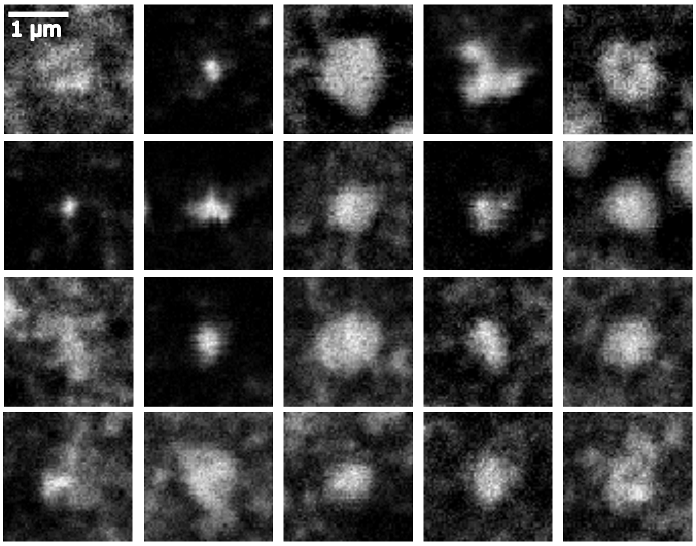


Figure 4.3.: Detail views of objects. Representative Pol II Ser5P clusters mid-plane obtained by STEDD super-resolution microscopy from a fixed sphere-stage zebrafish embryo. Each image is of size 60 × 60 pixels. The images are randomly selected from a dataset of size 148

4.3.2. Representation of object features

A feature is any measurable attribute, typically numerical, that describes an object's characteristics [157]. Each segmented object can be represented as a vector of features, and the collection of all such feature vectors forms the feature space [157]. In the context of biological imaging, objects can vary widely in shape and appearance [134]. For example, some objects may look like small dots, while others are larger and have clear boundaries. Some complex objects have unfolded shapes and lack well-defined edges, as shown in Fig. 4.3. These variations mean that different segmentation algorithms and settings may be more suitable for different types of objects.

To support the effective reuse of segmentation algorithms for different object types, the framework builds a feature space that captures these morphological variations (step 2 in Fig. 4.2). Two main categories of features are used: intensity histogram-based features and shape-related features.

- Intensity histogram-based features: These features describe the distribution and variation of pixel intensities within an object. They are particularly useful for identifying specific structures or patterns in microscopy images, as changes in the intensity histogram can signal the presence of different objects.
- Shape-related features: These focus on the shape and form of objects, including attributes like size, area, perimeter, and compactness. Such features are essential for understanding the morphology of objects, allowing for accurate comparison and analysis in microscopy studies.

The feature space includes specific features that help describe the objects:

- Local maxima and their values from the object's histogram, providing insight into prominent intensity levels.
- Local minima and their values from the object's histogram, revealing dips in intensity that may correspond to different structures.
- Object area, which measures the space occupied by the 2D shape of an object, offering a sense of its size.
- Object solidity, where a value of 1 indicates a compact object, while values less than 1 suggest irregular shapes or internal gaps.

These features were carefully selected to capture the key aspects of objects in microscopy images. Features like local maxima and minima describe intensity variations, providing a detailed profile of the object's intensity. Object area represents size, helping to assess the relative importance of objects within an image. Object solidity provides information about the regularity of an object's shape—compact objects have solidity values close to 1, while irregular ones have lower values. By focusing on these features, the framework provides a well-rounded representation of objects, which is crucial for selecting and reusing segmentation algorithms effectively.

After creating the feature space, the first step is to remove outliers (Fig. 4.4B). The Z-score method is used to calculate how far each data point is from the average. A common threshold of 3 standard deviations is applied to identify outliers. Removing these outliers ensures they do not affect the next steps and keeps the data more consistent.

Once outliers are removed, the features are normalized to a range of 0 to 1 (Fig. 4.4A). This step ensures all features are on the same scale, which makes the data easier to use during model training. Fig. 4.4 shows a 2D view of the feature space before and after outlier removal.

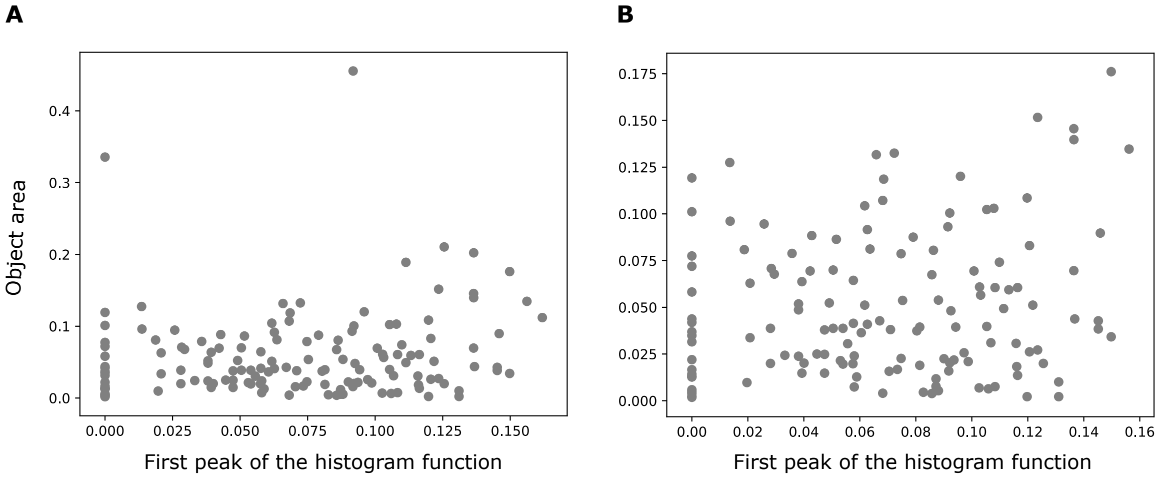


Figure 4.4.: A) Two features (area and the first local maximum of the histogram intensity function) are selected and 2D representation of the objects based on the selected features is illustrated. B) 2D representation of the object set in the feature space after outlier removal.

4.3.3. Initial selection of segmentation algorithm

In the initial segmentation phase (step 3 in Fig. 4.2), the framework uses an unsupervised process to evaluate different segmentation algorithms for each object (see Table 4.1). Although many segmentation algorithms could be used, the focus is on methods based on thresholding techniques [148] for several reasons.

First, thresholding methods are well-known for their simplicity and ease of use, making them ideal for the initial segmentation phase. They are also computationally efficient, which is important for processing large datasets quickly. Additionally, these methods have straightforward parameters, allowing easy adjustment and experimentation during this stage.

Since there are no annotated objects available for directly measuring segmentation quality, choosing the best algorithm and settings for each object is a challenge. To address this, an unsupervised metric (see Section 4.3.4) is used to estimate segmentation quality. This metric helps compare results and guides the framework in selecting an initial algorithm and its settings.

The initial results are then refined through further training with user input in the active learning steps that follow (see Fig. 4.5).

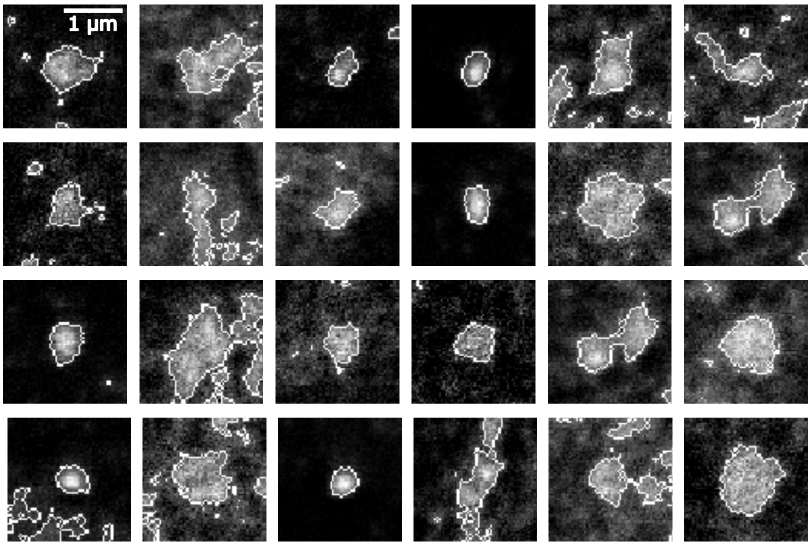


Figure 4.5.: Different objects annotated with the initial estimate of the segmentation algorithm and the settings, which is obtained by the framework through unsupervised iterations using an unsupervised metric defined in Section 4.3.4. Representative Pol II Ser5P clusters mid-plane obtained by STEDD super-resolution microscopy from a fixed sphere-stage zebrafish embryo. Each image is of size 60 × 60 pixels. The images are randomly selected from a dataset of size 148.

4.3.4. Evaluating segmentation quality with unsupervised metrics

During the process of estimating suitable segmentation algorithms and settings for each object (step 4 in Fig. 4.2), it is important to have an unsupervised metric to roughly compare segmentation results. Any unsupervised metric that can provide an initial evaluation of segmentation quality can be used here.

The proposed framework uses a metric that combines two criteria: cross-entropy [36] and 'busyness' [186]. These metrics were chosen because they each play a unique role in assessing segmentation quality.

Cross-entropy provides a numerical measure of the difference between the segmented object and its background. It helps assess how well the segmentation separates the object from the background noise. By maximizing cross-entropy, the framework aims to ensure that the boundaries between the object and the background are clearly defined.

Busyness, on the other hand, serves as a qualitative indicator. It measures the level of texture or complexity within the segmented areas. The assumption is that ideal objects and their backgrounds are relatively smooth and simple. By minimizing busyness, the goal is to produce segmentation results that emphasize clarity and simplicity, aligning with the expected appearance of the segmented objects.

Table 4.1.: These threshold-based segmentation algorithms have been selected for their diverse strengths and adaptability to various image characteristics
. These algorithms aim to create a binary mask from a grayscale image, effectively segmenting objects from a background.

Algorithm	Motivation
Isodata	This algorithm iteratively determines a threshold based on the image histogram, accommodating varying image characteristics.
Mean	Utilizes the simplicity of calculating the mean value of pixel intensities as a threshold, offering a straightforward approach for segmentation.
Minimum	Employs a histogram-based approach that iteratively smooths until only two local maxima remain, ensuring robustness in threshold determination [139].
Otsu	Maximizes the variance between two classes of pixels separated by the threshold, thereby effectively distinguishing an object from background [128].
Triangle	Utilizes a geometric method, assuming a maximum near one end of the histogram and searching towards the other end, which proves effective for various image distributions [185].
Yen	Based on Yen's thresholding method [154], offering robustness and adaptability to a wide range of image characteristics.
Li	Employs an iterative method based on Li's Minimum Cross-Entropy thresholding, ensuring accurate segmentation results even in challenging scenarios [102].
Local	Calculates each threshold value for every pixel as the weighted mean of the local neighbourhood minus an offset value, providing adaptability to local image variations.
Median	Utilizes the robustness of the median value of pixel intensities as a threshold, ensuring stability against outliers in the image data.

Combining these two metrics provides a balanced approach to evaluating segmentation quality, capturing both the quantitative accuracy of boundary separation and the qualitative simplicity of the segmented objects.

The formula for cross-entropy is:

$$\text{Cross-Entropy} = -\sum_{i=1}^{n} P(i) \times \log(Q(i/m)) \tag{4.1}$$

Where:

- i represents the intensity value of pixels.
- n and m are the number of pixels inside and outside the segmented boundary, respectively.

- $P(i)$ and $Q(i/m)$ are probability distributions of pixel intensity values inside and outside the segmented area.

A good segmentation line will maximize cross-entropy, meaning it effectively distinguishes the object signal from the background noise.

The formula for 'busyness' is:

$$\text{Busyness} = \sum |\nabla^2 I(x, y)| \tag{4.2}$$

Where:

- ∇^2 is the Laplacian operator, which calculates the second derivative of image intensity with respect to the x and y coordinates.
- $I(x, y)$ represents the intensity value of the image at coordinates (x, y).

By summing the absolute values of the Laplacian, the framework measures the overall changes in intensity, which indicate texture or 'busyness'. Lower busyness values suggest a smoother, less complex image, which is preferred for segmentation.

To improve segmentation quality, cross-entropy and busyness (Laplacian) metrics are combined. While cross-entropy effectively separates foreground and background regions, it can result in fragmented or noisy boundaries, especially in areas of high texture or noise. To address this, the Laplacian operator is used to estimate the "busyness" of different intensity thresholds. We observed that thresholds near the minimum and maximum intensity values often resulted in high busyness and poor segmentation. Based on this insight, we excluded these extreme thresholds by applying a constant offset from both ends of the intensity range. This filtered set of thresholds was then evaluated using cross-entropy to identify the optimal threshold, improving the smoothness and accuracy of the segmentation. This approach reduces noise-driven oversegmentation and improves the practical performance of the method.

4.3.5. Active learning iterations and query strategies

The metric mentioned earlier aids in estimating segmentation quality, yet additional methods are required for fully automated training. Requesting user input for every object is impractical. Hence, the combination of active learning and weakly supervised training is proposed to minimize interactions by selecting samples objectively and consulting users. A query strategy is necessary to choose samples presented to the user later (step 5 in Fig. 4.2). Three query strategies are combined in the proposed approach:

- **Selecting from diverse clusters:** To explore the feature space thoroughly, samples are chosen from various regions by clustering all samples into different groups and selecting from diverse groups within the feature space (Section 4.3.6 and Fig. 4.7).

- **Choosing the most informative objects:** Samples with the highest number of neighbors within a specified radius are identified as the most informative samples (Fig. 4.6B). Adjusting these informative objects indirectly influences many others, leading to faster convergence during active learning iterations.
- **Identifying the most uncertain objects:** Samples furthest from the center of the cluster are considered the most uncertain (Fig. 4.6C).

These query strategies are importat to efficiently guide the framework's learning process while reducing user involvement. Each strategy addresses specific challenges encountered during segmentation. By selecting samples from diverse clusters, comprehensive feature space exploration is ensured, facilitating robust segmentation across various image characteristics. Similarly, prioritizing adjustments on the most informative objects accelerates convergence during active learning iterations by influencing a broader set of similar objects. Additionally, targeting the most uncertain objects enables the framework to focus on areas with low segmentation confidence, refining accuracy where needed. Together, these strategies form a cohesive approach that balances exploration and exploitation, maximizing adaptability and efficiency in achieving high-quality segmentation results.

During each active learning interaction, multiple segmentation options for two samples from each cluster—the most informative and uncertain—are selected and presented to the user. Fig. 4.6 provides examples of the selected most informative and uncertain objects. The user is asked to evaluate and provide feedback on which segmentation option is better for each queried object. Based on this feedback, the framework directly adjusts the segmentation settings associated with the queried objects. Subsequently, an additional learning procedure updates the settings related to all other objects (Section 4.3.7). Further details on the update procedure are provided in Section 4.3.7.

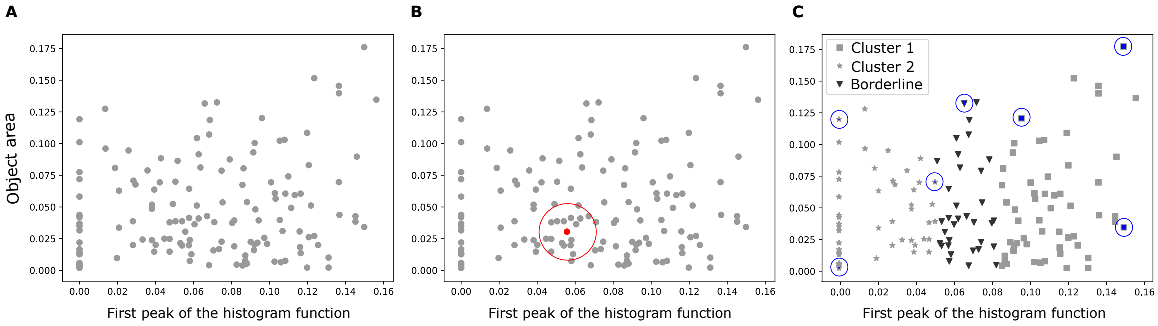


Figure 4.6.: Example results of query strategies . A) Representative 2D visualization of all samples in the feature space. B) the most informative sample with the largest number of neighbors specified with red. C) Uncertain samples with the farthest distance to the mean sample of clusters specified with blue (clusters are explained in Section 4.3.6).

4.3.6. Automatic determination of the clusters

As discussed earlier, the querying approach involves selecting samples from various clusters outlined in this section (step 6 in Fig. 4.2).

Segmenting the feature space into distinct clusters allows for diverse sample choices. While numerous clustering algorithms are available, k-means clustering was chosen [106].

K-means clustering stands out for its simplicity, efficiency, and effectiveness in partitioning data into distinct groups. Unlike more complex algorithms, k-means is easy to implement and computationally efficient, making it suitable for large datasets and real-time applications. Moreover, it does not require prior knowledge of the number of clusters, offering flexibility in the clustering process. Its iterative approach converges to local optima, providing reliable results. Additionally, k-means is robust to outliers and can reasonably handle datasets with non-linear boundaries.

To find the ideal number of clusters, silhouette analysis is employed [168, 85]. This method effectively assesses clustering quality, especially when the number of clusters is unknown. Unlike other techniques, silhouette analysis offers a quantitative measure of cluster cohesion and separation. The silhouette coefficient ranges from -1 to 1, with higher values indicating better clustering. It is robust across various data distributions and easy to understand. Values close to +1 suggest that the sample is distant from neighboring clusters, while a value of 0 indicates proximity to the boundary between adjacent clusters. Negative values indicate the misallocation of samples to clusters. The optimal number of clusters is determined by selecting the highest silhouette values across all samples.

Following the determination of the optimal number of clusters using the silhouette coefficient (Fig. 4.7A), the framework identifies samples that are either close to or on the borderlines (Fig. 4.7B) as a separate cluster.

4.3.7. Tuning the segmentation in each active learning iteration

In this section, the last step (step 7) in Fig. 4.2 is explained. In every active learning iteration, the framework first presents some objects (Fig. 4.8) to the user. Then, the settings associated with the queried objects are adjusted according to the user's feedback. The settings are all automatically updated by updating the mathematical formulation.

To implement this mathematical formulation, the framework uses a random forest regressor [72], mapping every object to the setting associated with it. When an object is chosen and presented to the user, the setting of the given object is updated. The regressor with the updated label is then re-trained, where the settings associated with others are also updated.

Once a sample is chosen, it is marked as visited and will not be presented to the user again. The process of selecting samples, presenting them to the user, adjusting the settings based on their feedback, and updating the regressor is repeated iteratively until convergence. The framework is considered converged when the user no longer requests any adjustments or

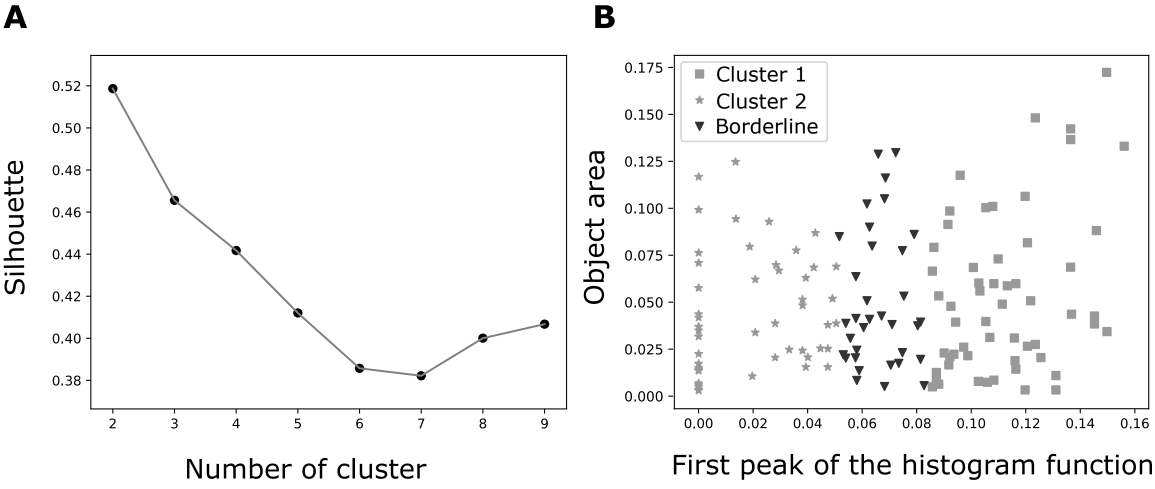


Figure 4.7.: A) Silhouette coefficient for a different number of clusters. The maximum value indicates the optimal number of clusters, which, in this case, is 2. B) Representative 2D visualized data samples, where only 2 features (Area and local maxima) are selected. Three different clusters (cluster 1, cluster 2, and borderline) with different marks based on the optimal number of clusters (2)

when all samples have been reviewed and marked as visited. In practice, convergence often occurs before all samples are marked, as the regressor becomes sufficiently accurate to generalize settings for the remaining samples.

4.4. Evaluation

In this section, we present a biological case study, which allowed us to evaluate our framework based on the following three research questions:

- RQ1: Is the framework, from the end-user perspective, converging to the desired results?
- RQ2: How satisfying is the quality of the segmentation results from the end-user perspective?
- RQ3: How effective is our framework in helping end-users with their segmentation tasks?

We conduct all procedures on a laptop with Intel(R) Core(TM) i7-8665U CPU 1.90GHz 2.11 GHz processor and 32.0 GB RAM, using the Microsoft Windows 11 operating system. The code is written in Python 3, uses the “Scikit-learn” package, and is publicly released on [62].

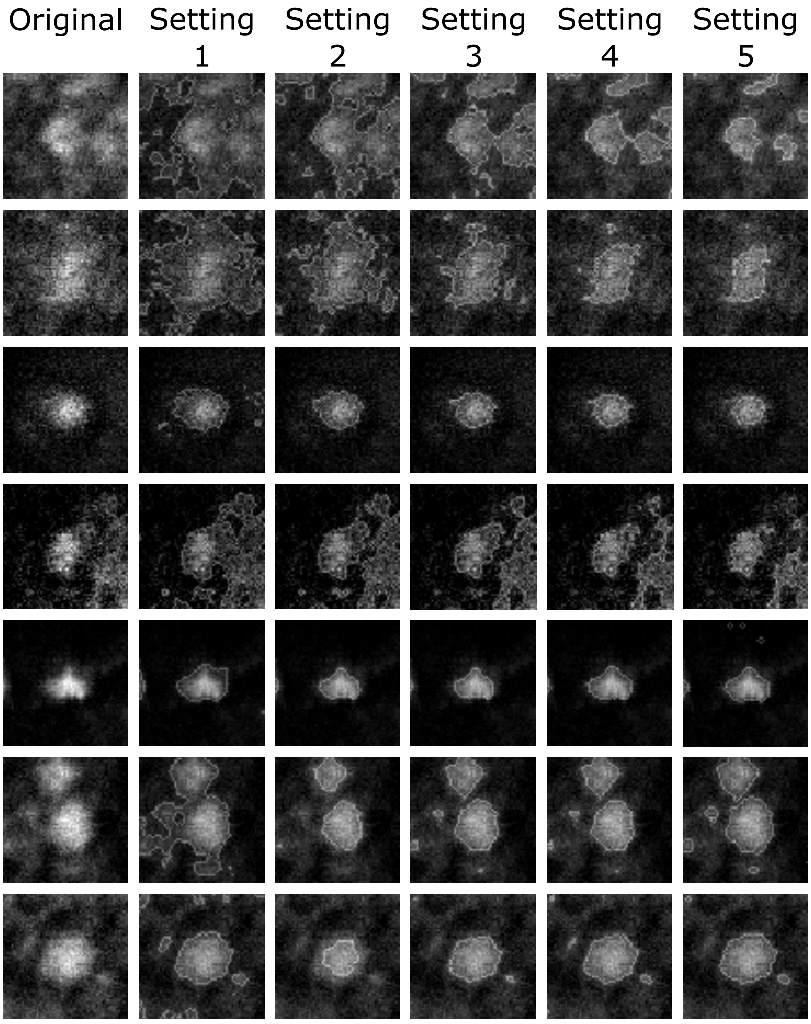


Figure 4.8.: Different items with varying configurations shown to the users. These images show Pol II Ser5P clusters in the mid-plane, captured using STEDD super-resolution microscopy from a zebrafish embryo fixed on a sphere stage. Each image is 60 by 60 pixels. The user's task is to choose the best segmentation option from the five suggested settings for each object, based on how clear and accurate the segmentation looks. The selection of these images is guided by the optimization steps explained in Section 4.3.5

4.4.1. Experimental setup

We used input images recorded by STEDD microscopy in a previous study [134, 69]. The images are maximum-intensity projections of cell nuclei in fixed zebrafish embryos, where Pol II Serine 5 phosphorylation was labeled by indirect immunofluorescence (Fig. 4.1). The dataset contained images of 60 nuclei. We extract an object set based on cropped ROI images of 60×60 pixel size, which are centered on prominent groups of Pol II. We obtained a set of 148 objects that were used for the further application of our framework.

To address the research questions, we asked five biologists from two separate laboratories to use our script for adjusting object segmentation (second scenario in Fig. 4.10). These biologists are part of the intended users of the framework, ensuring that the evaluation

Table 4.2.: Evaluation results provided by expert biologists on the convergence of the segmentation framework. The table summarizes how often the final iterations produced the best results and how often the segmented objects appeared stable across multiple iterations, indicating early convergence.

Choice	Choice
The last iteration contains the best results	90%
Segmented objects are unchanged in the last two iterations	40%
Segmented objects are unchanged in the last three iterations	10%

reflects real-world applications in different research settings. In line with the variety in object morphologies mentioned in [134], a feature space (as introduced in Section 4.3.2) is built, and objects are categorized into three clusters (Section 4.3.2, Section 4.3.6). Initially, through an unsupervised iteration, segmentation algorithms listed in Table 4.1 are applied to the objects (Section 4.3.3), and the preliminary segmentation settings are obtained by optimizing the unsupervised metric introduced in Section 4.3.4 (Fig. 4.5). This unsupervised iteration is performed without any queries or input from the biologists.

We then mathematically formulate the mapping function between objects and their associated settings by training a random forest regressor on the samples. The input to the regressor consists of feature vectors representing the objects, and the output corresponds to the associated settings. During each active learning iteration, six objects (selected according to the criteria explained in Section 4.3.5) are presented to the user. The user provides feedback by adjusting the settings for the presented objects and their neighbors. Following this, the regressor is re-trained using the updated feature vectors as input and the new settings as output. The updated regressor is then used to refine the settings for all remaining objects (Section 4.3.7).

The active learning iteration continues until convergence. The framework is considered converged when the user no longer requests adjustments or when all objects are marked as visited. In practice, the framework typically converged before all objects were visited, as the regressor became sufficiently accurate to generalize the settings for unvisited objects.

To address RQ1, we randomly selected 10 objects and shuffled the results of the last five iterations to assign choices [1..5] randomly. This was done to avoid introducing bias in the order of the segmented images. Fig. 4.9 demonstrates the segmented results for some randomly selected objects in the last five iterations. These segmented objects were presented to expert biologists, who were asked to visually assess and rank the segmented images from good to bad. An additional option was provided to indicate if the segmentation appeared almost unchanged across some iterations.

Table 4.2 summarizes the evaluation results. The experts identified that in 90% of the cases, the last iteration produced the best segmentation based on visual quality. Furthermore, in 40% of cases, the segmented objects were almost unchanged over the last two iterations, and in 10% of cases, they remained unchanged across the last three iterations. These results demonstrate that the learning procedure is effective and that the framework converges quickly, with stable results often achieved before all iterations are completed.

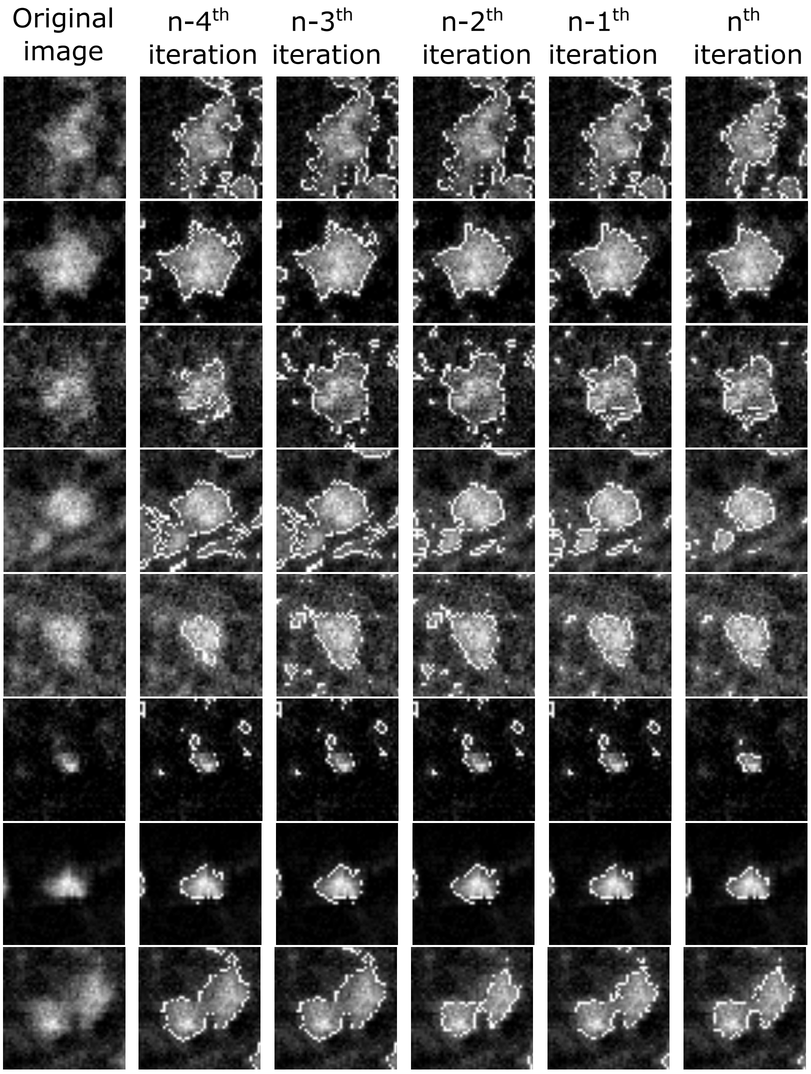


Figure 4.9.: Different objects annotated with the segmentation line obtained by the framework in the last 5 iterations while n_{th} iteration indicates the last iteration. Each image is of size 60 × 60 pixels, The images are randomly selected from a dataset of size 148.

To address RQ2, we needed a reference to compare the results of our framework. To establish this reference, we asked one expert biologist with extensive knowledge of ML to select a segmentation algorithm based on the object set and manually tune the algorithm's settings (first scenario in Fig. 4.10). This expert was highly familiar with segmentation algorithms and the application data. The manual adjustment process for this expert took approximately 130 minutes. The results obtained through this process are considered the best that can be achieved using the existing resources.

To evaluate the performance of our framework, we invited a group of biologists (five) to interact with the framework and provide feedback. The framework automatically selected objects based on its active learning strategy and presented them to the biologists, who then provided feedback to adjust the segmentation settings. To assess the quality of the segmentation results, we presented 10 randomly selected pairs of reference objects (manually

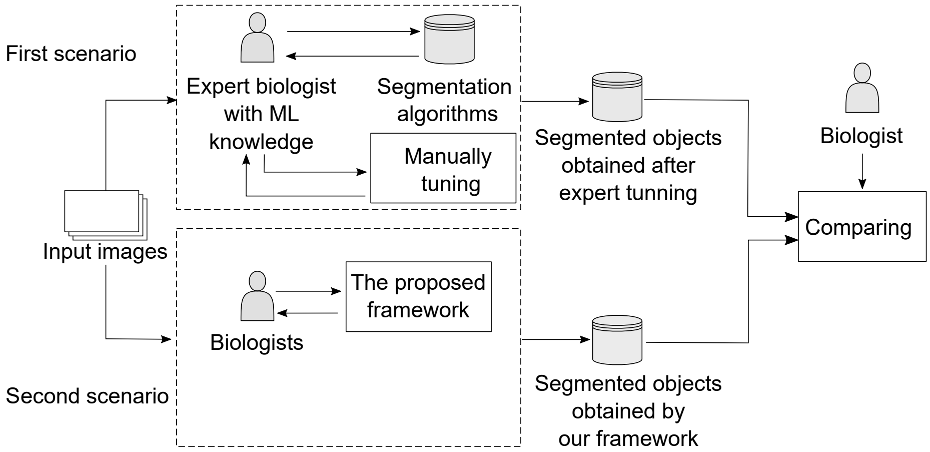


Figure 4.10.: A subjective metric is used for evaluation. The results obtained by our framework are compared to the results obtained after tuning the algorithm by an expert. This evaluation is based on a human visual assessment.

Table 4.3.: Evaluation results performed by the experts

Choice	**Percentage**
Results obtained by our framework outperform	80%
Results obtained after expert tuning outperform	0%
Both results look almost similar	10%
It depends on ...	10%

segmented) and results obtained by our framework to the expert biologist for a visual evaluation. A sample questionnaire presented to the biologists is available at [63].

Using a human to visually assess the result is a kind of subjective metric. We provided four choices for each comparison pair: (i) segmented image A is better, (ii) segmented image B is better, (iii) both segmentation results are almost equal, and (iv) the evaluation of segmentation depends on an additional consideration formulated by the expert. To avoid bias, we randomly alternated the assignment to choices A and B. Additionally, the expert was not informed about which image was manually segmented and which was generated by our framework. This ensured that the evaluation was blind and based purely on the quality of the segmentation, without prior assumptions influencing the expert's judgment.

The evaluation results indicate that, in the eyes of the expert, 90% of the segmentations obtained by our framework are better than or almost equal to the segmentations obtained after manual tuning by an expert (Table 4.3).

To address the RQ3, we ask five biologists to use our script and segment objects with our framework assistance (second scenario of Fig. 4.10). These five biologists are among the final target groups whom the framework eventually assists. We then record the time spent and the number of interactions required for tuning the script. Every interaction is one iteration of active learning. In Table 4.4, we refer to the experiment associated with every biologist as a case study. Table 4.4 reports the obtained results, indicating that the framework on average, converges within 5.4 user interactions. The convergence takes on

Table 4.4.: Number of interactions and the dedicated time required for convergence

	Total number of interactions	**dedicated time**	**Simplicity** **1: not simple** **7: maximally simple**
Participant 1	5	235 seconds	7
Participant 2	7	365 seconds	5
Participant 3	6	250 seconds	5
Participant 4	4	165 seconds	7
Participant 5	5	250 seconds	7
Average	5.4	253 seconds	6.2

average 253 seconds. This time includes only the time when the framework is working. The time that biologists take to decide on which result is better is not calculated. We then ask the biologists to rate the simplicity of our framework from 1 (not simple) to 7 (maximally simple) and the simplicity of our proposed framework was rated high 6.2 out of 7.

4.5. Threats to validity

The example application and evaluation setup give a strong base for our conclusions, though there are some ways we could improve generalizability. Our study uses data from one type of biological sample and one specific microscope, selected to ensure consistency. We have also used different samples to address daily variations in sample quality. While other microscopes may have differences, such as intensity and resolution, our approach is well-suited for the data we used. Adding data from other sample types and microscopes in future work could help make our findings more widely applicable.

The biologists from two separate laboratories tested the framework by adjusting segmentation on the object set. This process could introduce some bias, as the segmented objects may be influenced by the individual users interacting with the framework. Therefore, user interaction represents a potential consideration for external validity, as it remains uncertain whether different users would achieve varying segmentation outcomes. This potential bias could be reduced by involving additional users from a broader range of research groups.

Another factor in our experimental design is the limited pool of expert biologists assessing segmentation quality. With only a small group visually evaluating the segmentation outcomes, there may be some variation in alignment with end-user expectations, as different experts might assess quality differently. Expanding the assessment to include a wider range of experts would contribute to a more balanced perspective, enhancing the construct validity of our study.

To promote the reliability of our results, we enable the reproduction or replication of our study by making the underlying artifacts publicly available. These artifacts include the framework implementation [62], the questionnaires [63], and the dataset reused from a previous study [69].

4.6. Related work

Our work primarily relates to two categories: 1) Interactive approaches with user-guided parameter selection and 2) Weakly supervised segmentation methods. While these two categories can overlap, we review them separately for clarity. Additionally, we highlight the differences between these approaches and our work to clarify the necessity and novelty of our contribution.

4.6.1. Interactive approaches with user-guided parameter selection

Interactive object segmentation has been an active area of research for almost a decade. Early approaches formulated the problem as energy minimization tasks on a graph. An image is represented as an undirected graph defined by a joint probability p, and the goal is to assign a label to each pixel (0 for background and 1 for objects) such that p is minimized [7, 17, 60, 23].

More recent interactive segmentation methods require users to annotate the edges (extreme left, extreme right, extreme top, and extreme bottom) of objects. The connection lines between these edges represent the boundary line, and training on these annotated objects allows the method to detect the boundaries of unseen objects later [94, 135]. The region inside the boundaries is considered the foreground (object), while the outside area is labeled as the background. However, applying these methods to certain biological segmentation tasks, such as subcellular segmentation, presents challenges. Simple edge annotations cannot inherently represent some biological components, and the assumption that the entire region inside the boundaries constitutes the object does not always hold true. For example, doughnut-shaped structures have inner regions that might not be part of the object. Unlike these methods, which often rely on intensive user input and specific annotations, our approach minimizes manual effort by using weak supervision and active learning, allowing users to select between segmentation results to refine the model dynamically for irregular biological shapes.

4.6.2. Weakly supervised segmentation methods

Deep Learning-Based Interactive Segmentation

Recently, DL methods have gained popularity in interactive segmentation tasks [181, 175]. These methods allow users to annotate edges or bounding boxes, enabling the network to learn segmentation boundaries for unseen inputs [114, 27, 107]. Reinforcement learning-based methods further improve segmentation by using user inputs to refine predictions, which has been successful in CT/MRI image segmentation [26, 1].

Other weakly supervised methods focus on using limited annotations. For instance, image-level labels combined with self-attention mechanisms generate pseudo-instance-level annotations for segmentation [125]. Similarly, Box2Mask [34] utilizes bounding boxes for 3D instance segmentation. Snorkel [142] uses heuristics and noisy labels to train models without requiring manual annotations. However, these approaches often depend on predefined structures or labels, which may not generalize to the variability in biological images. Our approach, instead, dynamically adapts to unseen data with weak supervision and active learning, making it more flexible for segmenting biological structures without explicit or structured annotations.

4.6.3. Foundation models for segmentation

Recent advancements include foundation models like the Segment Anything Model (SAM) by Meta AI [32]. SAM, trained on 1.1 billion segmentation labels, generalizes across multiple domains and offers interactive segmentation with minimal annotations. However, its performance on microscopy images remains uncertain, as it was primarily trained on natural scenes. SAM requires powerful hardware for fine-tuning, which limits accessibility for many biologists. While SAM demonstrates strong generalization capabilities, it is less suited for tasks requiring dynamic adaptation to highly variable biological shapes, which our framework achieves by integrating active learning and weak supervision for refinement.

Extensions like the Segment Anything for Microscopy plugin for Napari[1] [4] improve SAM's applicability to bioimaging but still rely on manual annotations for high accuracy. Specialized models tailored for specific bioimaging tasks often outperform SAM in segmentation accuracy and efficiency.

4.6.4. Recent cell segmentation benchmarks

In a recent study [112], a cell segmentation benchmark was introduced, consisting of over 1,500 labeled images from more than 50 biological experiments. The goal of this benchmark is to create universal cell segmentation solutions that can be applied to various microscopy images without the need for manual parameter adjustments. During the development phase of the challenge, participants were given a dataset of 1,000 labeled and 1,725 unlabeled microscopy images. The participants were evaluated based on their ability to apply segmentation algorithms to diverse and previously unseen images during the testing phase.

The winning algorithm demonstrated robust generalization across diverse microscopy images, surpassing previous leading algorithms in cell segmentation challenges. The

---

[1] `https://napari.org`

study highlights the success of Transformer-based models in developing universal cell segmentation algorithms capable of handling a wide range of microscopy images. The results suggest that these advanced algorithms, now integrated into user-friendly interfaces like Napari, have the potential to enhance biological image analysis and provide discoveries in quantitative biological research. Additionally, effective strategies for using unlabeled data and mitigating catastrophic forgetting are crucial for improving segmentation performance. Future efforts should focus on developing biologist-in-the-loop systems and expanding benchmarks to include 3D microscopy images and classification tasks [112]. Unlike these benchmarks that depend on fully labeled datasets, our method eliminates the reliance on labeled data, enabling segmentation performance even without annotations through weak supervision and active learning.

4.6.5. Limitations of DL-based segmentation

A major limitation of DL-based segmentation methods is their reliance on annotated images for training, which can be time-consuming and require significant domain expertise [68]. This challenge is amplified in biological data, where object shapes vary widely, and annotating large datasets, such as those with hundreds of entities, is often impractical. Subcellular components further complicate this task due to their complex shapes and unclear boundaries [77].

Integration of pretrained DL models in the framework Our framework is designed to be adaptable, allowing for the integration of pretrained DL models as part of the initial segmentation step. These models can be included as segmentation options within the framework, enabling users to benefit from their robust feature extraction capabilities.

Part III.

Epilogue: Conclusion

5. Summary

5.1. Contribution 1 - reuse protocol for DL denoising with quality assurance in fluorescence microscopy

In this chapter, a reuse protocol is presented to bridge the gap between expert biologists and the technical requirements of DL tools. This protocol provides a quality-assured method for applying DL denoising, specifically N2V, in light microscopy, enabling biologists without programming skills to utilize advanced image processing with confidence in data reliability.

Motivation and research questions: Biological microscopy often requires careful balancing of image clarity with reduced photodamage. Additionally, DL image processing can introduce artificial structures, which may compromise data trustworthiness. A gap exists between the capabilities of DL tools and the technical expertise of many biologists who may not be familiar with programming but require reliable data for analysis. This protocol addresses these challenges by exploring three main questions: (1) how DL denoising can be reused to enhance SNR in microscopy images, (2) how quality and accuracy of DL-processed images can be assured without ground-truth data, and (3) how to determine optimal exposure settings for high-speed imaging to maintain data quality.

Approach: The protocol is structured to be accessible to non-programmers, with instructions for DL denoising accompanied by metrics for quality assessment, such as structural similarity (SSIM, local SSIM) and FRC, to detect and avoid artificial structures. A two-phase acquisition process is introduced, allowing for efficient data collection with subsequent quality-assured post-processing. This framework facilitates DL model reuse by biologists, enhancing research capabilities without requiring programming knowledge.

From a software engineering perspective: The protocol also contributes to software engineering (SE) by addressing the challenges of applying DL in domain-specific workflows. It provides a structured approach to quality assurance in time-lapse image analysis, a task complicated by variable noise levels and dynamic sample properties. Metrics such as FRC and SSIM are used to validate the accuracy and reliability of reconstructed images. Additionally, this study highlights the importance of tailoring quality assurance methods to specific domains. Unlike general-purpose imaging, biological microscopy involves unique conditions, such as fluorescence variability and sample dynamics, which require customized evaluation strategies. By combining self-supervised DL techniques with robust quality assurance metrics, the protocol ensures reproducibility and reliability in specialized applications. This approach also demonstrates how DL tools can be adapted for use in diverse imaging tasks without requiring paired noisy and high-quality datasets.

Findings and impact: Through imaging studies on zebrafish embryos, the protocol enabled fivefold faster imaging speeds while preserving data accuracy and structural reliability. This process facilitated clear visualization of RNA polymerase II dynamics, yielding new biological insights. Comparisons with traditional deconvolution methods demonstrated that N2V consistently improved effective resolution while minimizing artifact risk, thus enhancing data reliability for biologists.

5.2. Contribution 2 - framework for reusing ML segmentation with active learning and quality assurance

In this chapter, a data-driven framework is presented to simplify the selection, application, and reuse of ML segmentation algorithms in biological imaging. This framework addresses key challenges in ML model selection and reuse, particularly for researchers without programming skills, by incorporating active learning to minimize user input and streamline segmentation. The framework combines the ease of traditional segmentation methods with the flexibility to handle complex objects, making it adaptable to a variety of scientific tasks and fields.

Motivation and research questions: Image segmentation is a critical step in biomedical analysis, but choosing and reusing the most suitable ML model is often challenging and relies heavily on trial and error. Traditional segmentation methods perform well with clear object boundaries but face difficulties with ambiguous images. On the other hand, DL models are more adaptable but require significant labeled data and expertise. The framework presented in this chapter addresses these challenges by answering three main questions: (1) how to simplify the reuse of ML segmentation for non-technical users, (2) how to improve initial segmentation algorithm selection without requiring labeled data, and (3) how to minimize user interactions while refining segmentation accuracy through active learning.

Approach: The framework begins with an unsupervised method for selecting initial segmentation algorithms, guided by an unsupervised metric that evaluates segmentation quality. This approach avoids the need for labeled training data. Active learning is then applied to iteratively refine segmentation settings, significantly reducing the number of user interactions required. The active learning process leverages three query strategies: selecting diverse samples from different clusters, targeting the most informative samples, and identifying the most uncertain samples. Feedback from these interactions is generalized to similar objects using a regression model, ensuring consistent and reusable segmentation knowledge across the dataset.

From a software engineering perspective: This framework addresses key software engineering challenges related to reusing and refining ML models. First, it introduces a structured approach to integrate active learning into the segmentation workflow, reducing the need for manual tuning and minimizing user interactions. Second, it incorporates an unsupervised quality metric to guide the initial segmentation, allowing researchers to begin with reasonable settings even in the absence of labeled data. Additionally, the use of active learning query strategies ensures efficient convergence by focusing user input where it is most impactful. From a broader perspective, the framework demonstrates how domain-specific workflows can be simplified and enhanced by combining ML models with intuitive quality assurance and feedback mechanisms. This is particularly relevant for scientific imaging, where labeled data are often scarce and segmentation requirements vary across experiments.

Findings and impact: The framework was validated through experiments on biological imaging, specifically segmenting RNA Polymerase II clusters in zebrafish embryos. The results showed that the framework effectively minimized user input while maintaining high segmentation accuracy. The active learning process demonstrated rapid convergence, requiring only a small number of user interactions to achieve reliable results. Moreover, the framework's flexibility in handling diverse object shapes and sizes suggests it can be applied broadly in other domains where labeled data are limited or expensive to generate.

6. Future Work

In this chapter, I first explore general future directions and advancements in DL and ML for biological research. This helps to show how my research fits into the broader progress in these fields. After that, I focus on specific aspects of my project and suggest ways it could be extended or improved. By starting with the broader context and then narrowing down to my own work, this chapter provides a clear roadmap for future researchers to build on both the field's developments and the specific contributions of this thesis.

6.1. Future work: advancements and challenges in next-generation DL models

In this thesis, we have explored both traditional ML methods and more recent DL approaches. DL, as an emerging field, is developing rapidly and shows great potential for future research. It is advancing faster than traditional methods in many areas, indicating its likely importance in future developments. This section highlights areas within DL that are well-suited for further exploration, building on the core machine learning methods we have already covered.

6.1.1. Development of robust and generalist models

Research in DL is continuously advancing in terms of model robustness and generalization. Significant progress has been made in competitions like the Multi-Modality Cell Segmentation Challenge (MMCSC) [112]. For example, the KIT-GE model [161], which ranked among the top performers in the Cell Segmentation Benchmark (CSB), was used as a baseline in this challenge, along with models such as Cellpose, Cellpose 2.0, and OmniPose [112]. Analysis of these top-performing models highlights the importance of backbone network selection for next-generation models. While many models are based on U-Net-like architectures, leading models in multi-modality cell segmentation challenges have utilized advanced backbones such as SegFormer [179], ConvNeXt [103], and ResNeXt [180]. Techniques like self-configured neural networks, such as nnU-Net [76], and neural architecture search further improve segmentation performance.

6.1.2. Addressing fine-tuning limitations and enhancing generalization

A critical observation from the multi-modality cell segmentation challenge [112] is the limitation of fine-tuning for models like Cellpose 2.0, which may suffer significant performance loss on tasks outside their fine-tuning dataset. The top-performing MMCSC model [112] outperformed pre-trained generalist models such as Cellpose and Omnipose and a fine-tuned Cellpose 2.0 model. This success was achieved with a testing set that included images distinct from the training data from new biological experiments, demonstrating these leading models' robustness and improved generalization. However, this does not imply that the generalizability problem is fully solved. Instead, it highlights a promising trend toward developing models with strong generalization capabilities, reducing reliance on additional fine-tuning. Continued research and innovation in this direction are essential to further enhance the robustness and applicability of bioimage analysis models across diverse datasets and experimental conditions.

6.1.3. Future directions and architectural innovations

The availability and usability of top-ranking models remain critical for future development. Although the winning solution and its code are accessible on GitHub, it is yet to be seen if these models will offer documentation and interface tools comparable to established models like StarDist, Cellpose 2.0, or Omnipose. Achieving a truly general, user-friendly, one-click segmentation model without additional tuning would mark a significant advancement.

The trajectory of future enhancements—whether driven primarily by transformer architectures, advancements in convolutional networks, or hybrid approaches—remains an open question. The competition between these architectural innovations will shape the next generation of DL models.

6.1.4. Development of foundation models

Creating foundation models that work well as starting points for different tasks is a key area of interest in both research and industry. An example is Meta AI Research's Segment Anything Model (SAM), which is trained on a large dataset with 1.1 billion segmentation labels across 11 million high-quality images [32]. Although SAM mainly learned from images of natural scenes, it also includes some microscopy images, showing its flexibility. Extensions of SAM, such as Segment Anything for Microscopy, are designed to work across various types of images, using interactive segmentation to create full image masks with only minimal input from the user [4].

However, the current fine-tuning methodology in SAM for microscopy, which means updating the full model weights, is resource-intensive and requires a GPU. This can be time-consuming and limits the ability to perform finetuning for users with access to powerful hardware. Also, training models specialized for specific bioimaging applications. These specialized models could provide more accurate and relevant results for particular types of microscopy data.

6.2. Promoting FAIR principles in both ML and DL model development

In parallel with advancements in model architecture, there is an increasing emphasis on making models Findable, Accessible, Interoperable, and Reusable (FAIR). Public accessibility of models, comprehensive documentation, and well-packaged code for data processing and model training is crucial. Container platforms like Docker facilitate this by providing isolated environments and pre-configured installation processes, reducing the barriers to reuse. Initiatives like the MMCSC [112] encourage reproducible workflows by making source code available in public repositories and employing containers for software distribution.

A good example is the BioImage Model Zoo [130], a community-driven repository for pre-trained DL models in bioimage analysis. The BioImage Model Zoo promotes a standard model description format, ensuring that metadata is well-documented and models are easily accessible. This standardization is critical for making models Findable, Accessible, Interoperable, and Reusable (FAIR). The repository supports the execution of models through various popular bioimage tools, such as Ilastik [14], Fiji, CellProfiler [118], and QuPath [74], making it potentially widely applicable in the bioimage analysis community. It also facilitates inference on the BioImage Model Zoo web application and other web applications via the ImJoy plugin framework [131], potentially enabling seamless integration and use of models across different platforms.

Moreover, the BioImage Model Zoo [130] allows models to be extracted and imported into other tools, providing interoperability and reuse. This flexibility is important for researchers using models in diverse computational environments. However, currently, there are some compatibility issues with BioImage Model Zoo, such as a model requiring a specific package that the software does not support, which could decrease usability.

6.2.1. Enhancing model interoperability

Despite progress, a significant challenge remains in achieving full interoperability of models. Differences in model outputs necessitate distinct post-processing steps, complicating integration with existing tools. Establishing standardized inputs and outputs for models would streamline this process, enhancing the overall utility and adaptability of DL models in various applications.

6.2.2. Emphasizing model efficiency

Efficiency in algorithmic performance is another important factor influencing model adoption, particularly in low-resource settings. Competitions like MMCSC [112] evaluate models based on their trade-offs between accuracy and efficiency in runtime and memory usage. Future efforts to optimize bioimage analysis pipelines and develop efficient models are important for training and inference tasks on local desktops and web-based tools, especially where cloud data transfer is not feasible.

6.2.3. Improving access to tools and resources

The balance between local and cloud-based tools is evolving. Cloud-based solutions offer advantages in storage capacity, high-performance hardware, and accessibility. Notebooks, such as those provided by ZeroCostDL4Mic [29], offer a versatile platform for running bioimage analysis workflows, either locally or on cloud platforms like Google Colab. Additionally, tools like the BioContainers project and BIAFLOWS [147] provide frameworks for deploying and accessing various reproducible image analysis workflows.

For users comfortable with containers, tools like DeepCell Kiosk [8] and Distributed-Something [177] offer scalable solutions for image analysis workflows, advancing cloud infrastructure efficiently. These advancements reflect the necessity to make bioimage analysis tools more accessible, customizable, and user-friendly, driving the broader adoption of state-of-the-art DL models in the field.

This comprehensive approach to the future of DL models in bioimage analysis highlights the continuous evolution and integration of advanced techniques, emphasizing the importance of robustness, generalization, efficiency, and accessibility in developing next-generation models.

6.3. Extension of this thesis

In the previous sections, we explored the potential future directions and advancements in deep learning and machine learning for biology at large. Now, we will focus on the specific aspects of my own project. I will outline possible extensions and improvements that could be made to the research we have conducted in this thesis. This will provide a roadmap for how this work can be continued and expanded upon by future researchers or by further studies that we might undertake.

6.3.1. Extension to incorporate a broader array of segmentation algorithms

A way to extend the research presented in this thesis is the integration of additional segmentation algorithms into the existing framework. While the current system utilizes diverse algorithms, exploring further segmentation methods could offer improved adaptability and performance across various types of imaging data. For instance, incorporating advanced DL-based segmentation techniques might enhance the framework's ability to handle complex image structures and ambiguous boundaries, which are common in bioimaging. This expansion would broaden the framework's applicability and potentially increase its performance in real-world applications.

6.3.2. Expansion through a comprehensive user study

Conducting an extensive user study would be beneficial to validate further and refine the segmentation framework developed in this thesis. This study would mean a diverse group of researchers and practitioners who routinely work with bioimaging data. By assessing their interactions with the framework, feedback can be gathered on its usability, efficiency, and overall performance in real-world settings. Additionally, this study could identify common challenges users face, leading to targeted improvements in the framework's design and functionality. Ultimately, the insights gained from a thorough user study would demonstrate the framework's effectiveness and guide future enhancements to better meet the needs of the scientific community. For example, one can conduct these case studies to find room for improvement.

Usability testing: Conduct a series of usability tests in which participants from different scientific backgrounds use the segmentation framework to process their image data. The objectives would assess ease of use, interface intuitiveness, and user satisfaction. Tasks could include importing data, selecting segmentation algorithms, adjusting parameters, and interpreting output. Observations and timed task completions could help identify usability bottlenecks.

Comparative study: Design a comparative study where users perform segmentation tasks using the proposed framework and one or more established tools (e.g., Ilastik, CellProfiler). This study would measure each tool's effectiveness, accuracy, and time efficiency. Users would provide feedback through structured interviews or questionnaires, offering insights into the proposed framework's strengths and weaknesses compared to existing solutions.

Longitudinal study: Implement a longitudinal study where a smaller group of users engages with the framework over an extended period (e.g., several months). The study would track how users adapt to the framework and how well it integrates into their regular workflows. Periodic interviews and performance metrics could be used to assess learning curves, long-term usability, and sustained benefits or issues that arise with continued use.

Field study: Organize a field study where the framework is deployed in a real-world laboratory setting. Researchers would use the framework as part of their regular experimental routine, and data would be gathered on how the tool fits into complex workflows, its reliability under varied data conditions, and its scalability. Feedback sessions would help identify specific environmental or contextual challenges that might not be evident in controlled settings.

Task completion and error rate analysis Conduct a task completion study focusing on specific segmentation tasks critical for user workflows, such as segmenting noisy images or images with high object density. Measure the error rates and the type of errors commonly made versus not using the framework. This data can help improve algorithms and training materials to reduce common errors.

6.3.3. Integration of active learning with foundation deep learning models

A forward-thinking extension of the proposed segmentation framework would be integrating active learning mechanisms with generalized foundation DL models. This enhancement would mean developing a system where the active learning component assists users in selecting the most appropriate DL model tailored to their specific dataset characteristics. Here are some detailed steps and benefits of this approach:

Model selection engine: Develop an engine within the framework that houses multiple pre-trained foundation models, each specialized for different types of data (e.g., varied image resolutions, object densities, or imaging modalities). The active learning component could then analyze initial user data samples and suggest the most suitable model for their needs.

Adaptive learning and feedback loop: As users interact with the framework, an active learning component continually refines its model recommendations based on user feedback and segmentation success rates. This feedback loop uses user inputs on model performance to train the active learning algorithm, enhancing its ability to make more accurate predictions about model suitability for given data types. This process personalizes the user experience and improves the framework's accuracy.

User-guided customization: Allow users to override model suggestions, providing insights into user preferences and perceptions of model effectiveness, which can be invaluable for further refining the active learning algorithms.

6.3.4. Incorporation of a chatbot interface using language models for enhanced data specification

An intriguing extension to the framework would be the implementation of a chatbot interface powered by a large language model (LLM) to facilitate a more intuitive interaction between the user and the segmentation system. This chatbot would allow users to describe their data or provide domain-specific information in natural language, which the LLM would then interpret to adjust the segmentation process accordingly. Here's a detailed outline of how this could be implemented:

Chatbot interface: Develop a user-friendly chat interface that prompts users to describe their data and any specific segmentation challenges they face. This could include details like object size, density, background noise, or particular image anomalies.

Language model integration: Utilize a state-of-the-art language model to parse and understand the user's descriptions and queries. The model would extract meaningful information and translate it into technical specifications that the segmentation model can understand.

Dynamic model adjustment: Based on the information processed by the LLM, dynamically adjust the segmentation model's parameters or choose a different foundation model better suited to the described data characteristics. This could mean modifying threshold settings, choosing between supervised or unsupervised learning approaches, or altering the feature extraction processes.

Feedback mechanism: Allow the chatbot to ask follow-up questions based on initial user inputs, refining the data description and ensuring the model adjustments are as accurate as possible. This iterative dialogue can help pinpoint the most effective segmentation strategy.

Domain-specific tailoring: Train the LLM on specific data types relevant to the users' fields, such as biomedical imaging, environmental monitoring, or industrial quality control, enhancing the model's ability to effectively understand and process domain-specific information.

By embedding a chatbot within the segmentation framework, users can directly influence model behavior through conversational input, making the tool more accessible and tailored to individual needs. This extension enhances user engagement and potentially improves segmentation accuracy by incorporating user expertise and domain-specific knowledge directly into the model configuration process.

Acknowledgement

OpenAI's ChatGPT (GPT-4) was used to assist in drafting and rephrasing the Future Work sections (6.3.3 and 6.3.4). The tool supported the formulation of text based on my own ideas. All generated content was critically reviewed and edited by me.

Bibliography

[1] David Acuna, Amlan Kar, and Sanja Fidler. “Devil is in the edges: Learning semantic boundaries from noisy annotations”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 11075–11083.

[2] Navid Alemi Koohbanani et al. “NuClick: A deep learning framework for interactive segmentation of microscopic images”. In: *Medical Image Analysis* 65 (2020), p. 101771. ISSN: 1361-8415. DOI: `https://doi.org/10.1016/j.media.2020.101771`. URL: `https://www.sciencedirect.com/science/article/pii/S1361841520301353`.

[3] Noah Apthorpe et al. “Automatic neuron detection in calcium imaging data using convolutional networks”. In: *Advances in neural information processing systems* 29 (2016), pp. 3270–3278.

[4] Anwai Archit et al. “Segment anything for microscopy”. In: *bioRxiv* (2023), pp. 2023–08.

[5] Ignacio Arganda-Carreras et al. “Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification”. In: *Bioinformatics* 33.15 (2017), pp. 2424–2426.

[6] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.12 (2017), pp. 2481–2495.

[7] Xue Bai and Guillermo Sapiro. “Geodesic matting: A framework for fast interactive image and video segmentation and matting”. In: *International journal of computer vision* 82.2 (2009), pp. 113–132.

[8] Dylan Bannon et al. “DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes”. In: *Nature methods* 18.1 (2021), pp. 43–45.

[9] Niccolo Banterle et al. “Fourier ring correlation as a resolution criterion for super-resolution microscopy”. In: *Journal of Structural Biology* 183.3 (2013), pp. 363–367.

[10] Joshua Batson and Loic Royer. “Noise2Self: Blind Denoising by Self-supervision”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 524–533.

[11] Thorsten Beier et al. “Multicut brings automated neurite segmentation closer to human performance”. In: *Nature Methods* 14.2 (2017), pp. 101–102.

[12] Chinmay Belthangady and Loic A Royer. “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction”. In: *Nature methods* 16.12 (2019), pp. 1215–1225.

[13] Chinmay Belthangady and Loic A. Royer. “Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction”. In: *Nature Methods* 16.12 (2019), pp. 1215–1225.

[14] Stuart Berg et al. “Ilastik: interactive machine learning for (bio) image analysis”. In: *Nature Methods* 16.12 (2019), pp. 1226–1232.

[15] Udo J Birk. *Super-resolution microscopy: a practical guide*. John Wiley & Sons, 2017.

[16] Ajay Boyat and Brijendra Kumar Joshi. “Image denoising using wavelet transform and median filtering”. In: *2013 Nirma University International Conference on Engineering (NUiCONE)*. IEEE. 2013, pp. 1–6.

[17] Yuri Y Boykov and M-P Jolly. “Interactive graph cuts for optimal boundary & region segmentation of objects in ND images”. In: *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*. Vol. 1. IEEE. 2001, pp. 105–112.

[18] Dominique Brunet, Edward R. Vrscay, and Zhou Wang. “On the mathematical properties of the structural similarity index”. In: *IEEE Transactions on Image Processing* 21.4 (2011), pp. 1488–1499.

[19] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. “A non-local algorithm for image denoising”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Vol. 2. IEEE. 2005, pp. 60–65.

[20] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. “A review of image denoising algorithms, with a new one”. In: *Multiscale modeling & simulation* 4.2 (2005), pp. 490–530.

[21] Tim-Oliver Buchholz et al. “Content-aware image restoration for electron microscopy”. In: *Methods in cell biology* 152 (2019), pp. 277–289.

[22] Tim-Oliver Buchholz et al. “Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 502–506.

[23] Stefano Cagnoni et al. “Genetic algorithm-based interactive segmentation of 3D medical images”. In: *Image and Vision Computing* 17.12 (1999), pp. 881–895.

[24] Anne E Carpenter et al. “CellProfiler: image analysis software for identifying and quantifying cell phenotypes”. In: *Genome biology* 7.10 (2006), pp. 1–11.

[25] Filomena A Carvalho and Nuno C Santos. “Atomic force microscopy-based force spectroscopy—biological and biomedical applications”. In: *IUBMB life* 64.6 (2012), pp. 465–472.

[26] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. “Geodesic active contours”. In: *International journal of computer vision* 22.1 (1997), pp. 61–79.

[27] Lluis Castrejon et al. “Annotating object instances with a polygon-rnn”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5230–5238.

[28] Lucas von Chamier, Romain F Laine, and Ricardo Henriques. “Artificial intelligence for microscopy: what you should know”. In: *Biochemical Society Transactions* 47.4 (2019), pp. 1029–1040.

[29] Lucas von Chamier et al. “ZeroCostDL4Mic: an open platform to use Deep-Learning in Microscopy”. In: *BioRxiv* (2020), pp. 2020–03.

[30] Ge Chen et al. “Fluorescence microscopy images denoising via deep convolutional sparse coding”. In: *Signal Processing: Image Communication* 117 (2023), p. 117003. ISSN: 0923-5965. DOI: `https://doi.org/10.1016/j.image.2023.117003`. URL: `https://www.sciencedirect.com/science/article/pii/S0923596523000851`.

[31] Jianxu Chen et al. “The Allen Cell and Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images”. In: *bioRxiv* (2020), p. 491035.

[32] Tianle Chen et al. “Segment Anything Model (SAM) Enhanced Pseudo Labels for Weakly Supervised Semantic Segmentation”. In: *arXiv preprint arXiv:2305.05803* (2023).

[33] Yangyang Chen et al. “Review on high spatial resolution remote sensing image segmentation evaluation”. In: *Photogrammetric Engineering & Remote Sensing* 84.10 (2018), pp. 629–646.

[34] Julian Chibane et al. “Box2Mask: Weakly Supervised 3D Semantic Instance Segmentation using Bounding Boxes”. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*. Springer. 2022, pp. 681–699.

[35] Won-Ki Cho et al. “RNA Polymerase II cluster dynamics predict mRNA output in living cells”. In: *eLife* 5 (2016), e13617.

[36] Imre Csiszár, Paul C Shields, et al. “Information theory and statistics: A tutorial”. In: *Foundations and Trends® in Communications and Information Theory* 1.4 (2004), pp. 417–528.

[37] Kostadin Dabov et al. “Image denoising by sparse 3-D transform-domain collaborative filtering”. In: *IEEE Transactions on image processing* 16.8 (2007), pp. 2080–2095.

[38] James C Davis et al. “Reusing deep learning models: Challenges and directions in software engineering”. In: *2023 IEEE John Vincent Atanasoff International Symposium on Modern Computing (JVA)*. IEEE. 2023, pp. 17–30.

[39] Bhupendra S Deshmukh and Vijay H Mankar. “Segmentation of microscopic images: A survey”. In: *2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies*. IEEE. 2014, pp. 362–364.

[40] Nilanjan Dey et al. “Social group optimization supported segmentation and evaluation of skin melanoma images”. In: *Symmetry* 10.2 (2018), p. 51.

[41] Michael J Dykstra and Laura E Reuss. *Biological electron microscopy: theory, techniques, and troubleshooting*. Springer Science & Business Media, 2011.

[42] Amicia D Elliott. “Confocal microscopy: principles and modern practices”. In: *Current protocols in cytometry* 92.1 (2020), e68.

[43] Dennis Eschweiler et al. “Towards annotation-free segmentation of fluorescently labeled cell membranes in confocal microscopy images”. In: *Simulation and Synthesis in Medical Imaging: 4th International Workshop, SASHIMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 4*. Springer. 2019, pp. 81–89.

[44] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *Nature* 542.7639 (2017), pp. 115–118.

[45] Andreas Ettinger and Torsten Wittmann. “Fluorescence live cell imaging”. In: *Methods in cell biology* 123 (2014), pp. 77–94.

[46] Orestis Faklaris et al. “Long-term quality assessment and monitoring of light microscope performance through accessible and reliable protocols, tools and metrics”. In: *bioRxiv* (2021), pp. 2021–06.

[47] Orestis Faklaris et al. “Quality assessment in light microscopy for routine use through simple tools and robust metrics”. In: *Journal of Cell Biology* 221.11 (2022), e202107093.

[48] Michael Felderer et al. *Toward Research Software Engineering Research*. 2023. DOI: `10.5281/zenodo.8020525`. URL: `https://doi.org/10.5281/zenodo.8020525`.

[49] Alexander Fillbrunn et al. “KNIME for reproducible cross-domain analysis of life science data”. In: *Journal of biotechnology* 261 (2017), pp. 149–156.

[50] Linda S. Forero-Quintero et al. “Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene”. In: *Nature Communications* 12.1 (2021), pp. 1–16.

[51] Melanie M Frigault et al. “Live-cell microscopy–tips and tools”. In: *Journal of cell science* 122.6 (2009), pp. 753–767.

[52] Felix Fuentes-Hurtado, Jean-Baptiste Sibarita, and Virgile Viasnoff. “Generalizable Denoising of Microscopy Images using Generative Adversarial Networks and Contrastive Learning”. In: *arXiv preprint arXiv:2303.15214* (2023).

[53] Tobias Gebäck and Petros Koumoutsakos. “Edge detection in microscopy images using curvelets”. In: *BMC bioinformatics* 10.1 (2009), pp. 1–14.

[54] Nodar Gogoberidze and Beth A Cimini. “Defining the boundaries: challenges and advances in identifying cells in microscopy images”. In: *Current Opinion in Biotechnology* 85 (2024), p. 103055.

[55] Mona Golmohammadzadeh et al. “Advanced imaging techniques: microscopy”. In: *Advances in Applied Microbiology* 122 (2023), pp. 1–25.

[56] Bhawna Goyal et al. “Image denoising review: From classical to state-of-the-art approaches”. In: *Information Fusion* 55 (2020), pp. 220–244.

[57] Joe G Greener et al. “A guide to machine learning for biologists”. In: *Nature reviews Molecular cell biology* 23.1 (2022), pp. 40–55.

[58] Noah F Greenwald et al. “Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning”. In: *Nature biotechnology* 40.4 (2022), pp. 555–565.

[59] Fidel A Guerrero-Pena et al. “Multiclass weighted loss for instance segmentation of cluttered cells”. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 2018, pp. 2451–2455.

[60] Varun Gulshan et al. “Geodesic star convexity for interactive image segmentation”. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 2010, pp. 3129–3136.

[61] SA Haider et al. “Fluorescence microscopy image noise reduction using a stochastically-connected random field model”. In: *Scientific reports* 6.1 (2016), p. 20640.

[62] Hamideh Hajiabadi. *Interactive segmentation*. Nov. 2024. DOI: `10.5281/zenodo.5886947`. URL: `https://doi.org/10.5281/zenodo.5886947`.

[63] Hamideh Hajiabadi. *Questionaries*. Apr. 2022. DOI: `10.5281/zenodo.7682591`. URL: `https://doi.org/10.5281/zenodo.7682591`.

[64] Hamideh Hajiabadi, Lennart Hilbert, and Anne Koziolek. “Easing the Reuse of ML Solutions by Interactive Clustering-based Autotuning in Scientific Applications”. In: *2022 48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. 2022, pp. 5–12. DOI: `10.1109/SEAA56994.2022.00011`.

[65] Hamideh Hajiabadi et al. “A data-driven active learning approach to reusing ML solutions in scientific applications”. In: *Journal of Systems and Software* 211 (2024), p. 111986.

[66] Hamideh Hajiabadi et al. “Deep-learning microscopy image reconstruction with quality control reveals second-scale rearrangements in RNA polymerase II clusters”. In: *PNAS Nexus* 1.3 (May 2022), pgac065. ISSN: 2752-6542. DOI: `10.1093/pnasnexus/pgac065`. eprint: `https://academic.oup.com/pnasnexus/article-pdf/1/3/pgac065/48939536/pgac065.pdf`. URL: `https://doi.org/10.1093/pnasnexus/pgac065`.

[67] Wilhelm Hasselbring. “Research Directions for Multi-Disciplinary RSE Research to Impact RSE Practise”. In: *RSE Conference 2024 (RSECon24)*. Newcastle, UK: Zenodo, 2024. DOI: `10.5281/zenodo.14290294`. URL: `https://doi.org/10.5281/zenodo.14290294`.

[68] Mohammad Hesam Hesamian et al. “Deep learning techniques for medical image segmentation: achievements and challenges”. In: *Journal of digital imaging* 32.4 (2019), pp. 582–596.

[69] Lennart Hilbert. *Analysis of RNA polymerase II phosphorylation in STimulated Emission Double Depletion (STEDD) microscopy images*. Steps needed to carry out the analysis: 1. Extract the image data by running the script STEDDistributionsExtraction.m 2. Analyze the extracted data by running the script STEDD_AggregateAnalysis.m 3. Generate example images by running the scriptExampleAggregateZooms.m. Zen-

odo, June 2021. DOI: `10.5281/zenodo.4973062`. URL: `https://doi.org/10.5281/zenodo.4973062`.

[70] Lennart Hilbert et al. “Transcription organizes euchromatin via microphase separation”. In: *Nature Communications* 12.1 (2021), pp. 1–12.

[71] Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: *science* 313.5786 (2006), pp. 504–507.

[72] Tin Kam Ho. “Random decision forests”. In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE. 1995, pp. 278–282.

[73] Reka Hollandi et al. “nucleAIzer: a parameter-free deep learning framework for nucleus segmentation using image style transfer”. In: *Cell systems* 10.5 (2020), pp. 453–458.

[74] MP Humphries, P Maxwell, and M Salto-Tellez. “QuPath: The global impact of an open source digital pathology system”. In: *Computational and Structural Biotechnology Journal* 19 (2021), pp. 852–859.

[75] Jaroslav Icha et al. “Phototoxicity in live fluorescence microscopy, and how to avoid it”. In: *BioEssays* 39.8 (2017), p. 1700003.

[76] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pp. 203–211.

[77] Mostafa Jahanifar, Navid Alemi Koohbanani, and Nasir Rajpoot. “Nuclick: From clicks in the nuclei to nuclear boundaries”. In: *arXiv preprint arXiv:1909.03253* (2019).

[78] Deepack Jakhar and Ishmeet Kaur. “Artificial intelligence, machine learning and deep learning: definitions and differences”. In: *Clinical and experimental dermatology* 45.1 (2020), pp. 131–132.

[79] Ellen C Jensen. “Overview of live-cell imaging: requirements and methods used”. In: *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 296.1 (2013), pp. 1–8.

[80] Ian T Jolliffe and Jorge Cadima. “Principal component analysis: a review and recent developments”. In: *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150202.

[81] William Jones et al. “Computational biology: deep learning”. In: *Emerging Topics in Life Sciences* 1.3 (2017), pp. 257–274.

[82] Iris Jonkers, Hojoong Kwak, and John T. Lis. “Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons”. In: *eLife* 3 (2014), e02407.

[83] Nicole Kilian et al. “Assessing photodamage in live-cell STED microscopy”. In: *Nature Methods* 15.10 (2018), pp. 755–756.

[84] Alexander Kirillov et al. “Segment anything”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 4015–4026.

[85] Trupti M Kodinariya, Prashant R Makwana, et al. “Review on determining number of Cluster in K-Means Clustering”. In: *International Journal* 1.6 (2013), pp. 90–95.

[86] Yousef Al-Kofahi et al. “A deep learning-based algorithm for 2-D cell segmentation in microscopy images”. In: *BMC bioinformatics* 19.1 (2018), pp. 1–11.

[87] Sami Koho et al. “Fourier ring correlation simplifies image restoration in fluorescence microscopy”. In: *Nature communications* 10.1 (2019), p. 3103.

[88] Jeroen Kole et al. “Live-Cell Imaging: A Balancing Act Between Speed, Sensitivity, and Resolution”. In: *Principles of Light Microscopy: From Basic to Advanced*. Springer, 2022, pp. 139–172.

[89] Navid Alemi Koohbanani et al. “NuClick: a deep learning framework for interactive segmentation of microscopic images”. In: *Medical Image Analysis* 65 (2020), p. 101771.

[90] Benjamin G Kopek et al. “Diverse protocols for correlative super-resolution fluorescence imaging and electron microscopy of chemically fixed samples”. In: *Nature protocols* 12.5 (2017), pp. 916–946.

[91] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. “Noise2Void-learning denoising from single noisy images”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2129–2137.

[92] Alexander Krull et al. “Probabilistic Noise2Void: unsupervised content-aware denoising”. In: *Frontiers in Computer Science* 2 (2020), p. 5.

[93] Dai Kusumoto and Shinsuke Yuasa. “The application of convolutional neural network to stem cell biology”. In: *Inflammation and Regeneration* 39.1 (2019), pp. 1–7.

[94] Vivek Kwatra et al. “Graphcut textures: Image and video synthesis using graph cuts”. In: *Acm transactions on graphics (tog)* 22.3 (2003), pp. 277–286.

[95] Alan J Lacey. *Light microscopy in biology: a practical approach*. Vol. 195. OUP Oxford, 1999.

[96] P. Philippe Laissue et al. “Assessing phototoxicity in live fluorescence imaging”. In: *Nature Methods* 14.7 (2017), pp. 657–661.

[97] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444.

[98] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.

[99] Michael Y Lee et al. “CellSeg: a robust, pre-trained nucleus segmentation and pixel quantification software for highly multiplexed fluorescence images”. In: *BMC bioinformatics* 23.1 (2022), p. 46.

[100] Joël Lefebvre et al. “Single-Molecule Localization Microscopy Reconstruction Using Noise2Noise for Super-Resolution Imaging of Actin Filaments”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 1596–1599.

[101] Jaakko Lehtinen et al. “Noise2Noise: Learning Image Restoration without Clean Data”. In: *arXiv* (2018).

[102] Chun Hung Li and CK Lee. “Minimum cross entropy thresholding”. In: *Pattern recognition* 26.4 (1993), pp. 617–625.

[103] Jiachen Li et al. “ConvNeXt-backbone HoVerNet for nuclei segmentation and classification”. In: *arXiv preprint arXiv:2202.13560* (2022).

[104] Jieru Li et al. “Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells”. In: *Cell* 178.2 (2019), 491–506.e28.

[105] Hao Liang et al. “A region-based convolutional network for nuclei detection and segmentation in microscopy images”. In: *Biomedical Signal Processing and Control* 71 (2022), p. 103276.

[106] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. “The global k-means clustering algorithm”. In: *Pattern recognition* 36.2 (2003), pp. 451–461.

[107] Huan Ling et al. “Fast interactive object annotation with curve-gcn”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 5257–5266.

[108] Geert Litjens et al. “A survey on deep learning in medical image analysis”. In: *Medical image analysis* 42 (2017), pp. 60–88.

[109] Zhichao Liu et al. “A survey on applications of deep learning in microscopy image analysis”. In: *Computers in Biology and Medicine* 134 (2021), p. 104523.

[110] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.

[111] Florian Luisier, Thierry Blu, and Michael Unser. “Image denoising in mixed Poisson–Gaussian noise”. In: *IEEE Transactions on image processing* 20.3 (2010), pp. 696–708.

[112] Jun Ma et al. “The multimodality cell segmentation challenge: toward universal solutions”. In: *Nature methods* (2024), pp. 1–11.

[113] Soham Mandal and Virginie Uhlmann. “Splinedist: Automated cell segmentation with spline curves”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 1082–1086.

[114] Kevis-Kokitsi Maninis et al. “Deep extreme cut: From extreme points to object segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 616–625.

[115] Youssef Mansour and Reinhard Heckel. “Zero-Shot Noise2Noise: Efficient Image Denoising Without Any Data”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2023, pp. 14018–14027.

[116] Petr Matula et al. “Quantification of fluorescent spots in time series of 3D confocal microscopy images of endoplasmic reticulum exit sites based on the HMAX transform”. In: *Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol. 7626. International Society for Optics and Photonics. 2010, 76261H.

[117] Maciej A Mazurowski et al. “Segment anything model for medical image analysis: an experimental study”. In: *Medical Image Analysis* 89 (2023), p. 102918.

[118] Claire McQuin et al. “CellProfiler 3.0: Next-generation image processing for biology”. In: *PLoS biology* 16.7 (2018), e2005970.

[119] Erik Meijering. “Cell segmentation: 50 years down the road [life sciences]”. In: *IEEE signal processing magazine* 29.5 (2012), pp. 140–145.

[120] Erick Moen et al. “Deep learning for cellular image analysis”. In: *Nature Methods* 16.12 (2019), pp. 1233–1246.

[121] Erick Moen et al. “Deep learning for cellular image analysis”. In: *Nature methods* 16.12 (2019), pp. 1233–1246.

[122] Volodymyr Nechyporuk-Zloy. *Principles of light microscopy: from basic to advanced*. Springer Nature, 2022.

[123] Elias Nehme et al. “Deep-STORM: super-resolution single-molecule microscopy by deep learning”. In: *Optica* 5.4 (2018), pp. 458–464.

[124] Glyn Nelson et al. “QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy”. In: *Journal of microscopy* 284.1 (2021), pp. 56–73.

[125] Kazuya Nishimura et al. “Weakly supervised cell instance segmentation under various conditions”. In: *Medical Image Analysis* 73 (2021), p. 102182.

[126] Niall O’Mahony et al. “Deep learning vs. traditional computer vision”. In: *Science and Information Conference*. Springer. 2019, pp. 128–144.

[127] Sarah Osterwald et al. “PML induces compaction, TRF2 depletion and DNA damage signaling at telomeres and promotes their alternative lengthening”. In: *Journal of cell science* 128.10 (2015), pp. 1887–1900.

[128] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.

[129] Chawin Ounkomol et al. “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy”. In: *Nature Methods* 15.11 (2018), pp. 917–920.

[130] Wei Ouyang et al. “Bioimage model zoo: a community-driven resource for accessible deep learning in bioimage analysis”. In: *BioRxiv* (2022), pp. 2022–06.

[131] Wei Ouyang et al. “ImJoy: an open-source computational platform for the deep learning era”. In: *Nature methods* 16.12 (2019), pp. 1199–1200.

[132] Marius Pachitariu and Carsen Stringer. “Cellpose 2.0: how to train your own model”. In: *Nature methods* 19.12 (2022), pp. 1634–1641.

[133] Agnieszka Pancholi et al. “RNA polymerase II clusters form in line with surface condensation on regulatory chromatin”. In: *Molecular Systems Biology* 17.9 (2021), pp. 1–26.

[134] Agnieszka Pancholi et al. “RNA polymerase II clusters form in line with surface condensation on regulatory chromatin”. In: *Molecular systems biology* 17.9 (2021), e10272.

[135] Dim P Papadopoulos et al. “Extreme clicking for efficient object annotation”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 4930–4939.

[136] James Pawley. *Handbook of biological confocal microscopy*. Vol. 236. Springer Science & Business Media, 2006.

[137] Jordi Pont-Tuset and Ferran Marques. “Supervised evaluation of image segmentation and object proposal techniques”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.7 (2015), pp. 1465–1478.

[138] Stephan Preibisch et al. “Efficient Bayesian-based multiview deconvolution”. In: *Nature Methods* 11.6 (2014), pp. 645–648.

[139] Judith MS Prewitt and Mortimer L Mendelsohn. “The analysis of cell images”. In: *Annals of the New York Academy of Sciences* 128.3 (1966), pp. 1035–1053.

[140] Yuhui Quan et al. “Self2self with dropout: Learning self-supervised denoising from single image”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 1890–1898.

[141] Valerica Raicu and William F Schmidt. “Advanced microscopy techniques”. In: *G-Protein-Coupled Receptor Dimers* (2017), pp. 39–75.

[142] Alexander Ratner et al. “Snorkel: Rapid training data creation with weak supervision”. In: *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*. Vol. 11. 3. NIH Public Access. 2017, p. 269.

[143] William Hadley Richardson. “Bayesian-based iterative method of image restoration”. In: *Journal of the Optical Society of America* 62.1 (1972), pp. 55–59.

[144] Yair Rivenson et al. “Deep learning microscopy”. In: *Optica* 4.11 (2017), pp. 1437–1443.

[145] Benjamin Rombaut, Joris Roels, and Yvan Saeys. “BioSegment: Active Learning segmentation for 3D electron microscopy imaging.” In: *IAL@ PKDD/ECML*. 2022, pp. 7–26.

[146] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[147] Ulysse Rubens et al. “BIAFLOWS: A collaborative framework to reproducibly deploy and benchmark bioimage analysis workflows”. In: *Patterns* 1.3 (2020).

[148] Prasanna K Sahoo, SAKC Soltani, and Andrew KC Wong. “A survey of thresholding techniques”. In: *Computer vision, graphics, and image processing* 41.2 (1988), pp. 233–260.

[149] Nancy Salem, Noorhan M Sobhy, and Mohamed El Dosoky. “A comparative study of white blood cells segmentation using otsu threshold and watershed transformation”. In: *Journal of Biomedical Engineering and Medical Imaging* 3.3 (2016), p. 15.

[150] Sebastian Salwig, Jakob Drefs, and Jörg Lücke. “Zero-shot denoising of microscopy images recorded at high-resolution limits”. In: *PLOS Computational Biology* 20.6 (June 2024), pp. 1–22. DOI: `10.1371/journal.pcbi.1012192`. URL: `https://doi.org/10.1371/journal.pcbi.1012192`.

[151] Yuko Sato et al. “Histone H3K27 acetylation precedes active transcription during zebrafish zygotic genome activation as revealed by live-cell analysis”. In: *Development* 146.19 (2019), dev179127.

[152] Nico Scherf and Jan Huisken. “The smart and gentle microscope”. In: *Nature Biotechnology* 33.8 (2015), pp. 815–818.

[153] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. In: *Nature methods* 9.7 (2012), pp. 671–675.

[154] Mehmet Sezgin and Bülent Sankur. “Survey over image thresholding techniques and quantitative performance evaluation”. In: *Journal of Electronic imaging* 13.1 (2004), pp. 146–165.

[155] Colin JR Sheppard et al. “Signal-to-noise ratio in confocal microscopes”. In: *Handbook of biological confocal microscopy* (2006), pp. 442–452.

[156] Asm Shihavuddin et al. “Smooth 2D manifold extraction from 3D image stack”. In: *Nature Communications* 8.1 (2017), pp. 1–8.

[157] Jack Sklansky. “Image Segmentation and Feature Extraction”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 8.4 (1978), pp. 237–247. DOI: `10.1109/TSMC.1978.4309944`.

[158] Christoph Sommer et al. “Ilastik: Interactive learning and segmentation toolkit”. In: *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE. 2011, pp. 230–233.

[159] Justin Sonneck, Shuo Zhao, and Jianxu Chen. “On the risk of manual annotations in 3D confocal microscopy image segmentation”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 3894–3902.

[160] Timothy J. Stasevich et al. “Regulation of RNA polymerase II activation by histone acetylation in single living cells”. In: *Nature* 516.7530 (2014), pp. 272–275.

[161] Johannes Stegmaier and Ralf Mikut. “KIT-GE”. In: ().

[162] Stelzer. “Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy”. In: *Journal of Microscopy* 189.1 (1998), pp. 15–24.

[163] David J. Stephens and Victoria J. Allan. “Light microscopy techniques for live cell imaging”. In: *Science* 300.5616 (2003), pp. 82–86.

[164] Barbara Steurer et al. “Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II”. In: *Proceedings of the National Academy of Sciences of the United States of America* 115.19 (2018), E4368–E4376.

[165] Devin P. Sullivan et al. “Deep learning is combined with massive-scale citizen science to improve large-scale image classification”. In: *Nature Biotechnology* 36.9 (2018), pp. 820–828.

[166] Abdel Aziz Taha and Allan Hanbury. “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool”. In: *BMC medical imaging* 15.1 (2015), pp. 1–28.

[167] Nima Tajbakhsh et al. “Convolutional neural networks for medical image analysis: Full training or fine tuning?” In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1299–1312.

[168] Tippaya Thinsungnoena et al. “The clustering validity with silhouette and sum of squared errors”. In: *learning* 3.7 (2015).

[169] Rintu Maria Thomas and Jisha John. “A review on cell detection and segmentation in microscopic images”. In: *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. IEEE. 2017, pp. 1–5.

[170] Giorgio Tortarolo et al. “Evaluating image resolution in stimulated emission depletion microscopy”. In: *Optica* 5.1 (2018), pp. 32–35.

[171] Feng Wang et al. “Noise2Atom: unsupervised denoising for scanning transmission electron microscopy images”. In: *Applied Microscopy* 50.1 (2020), pp. 1–9.

[172] Hongda Wang et al. “Deep learning enables cross-modality super-resolution in fluorescence microscopy”. In: *Nature Methods* 16.1 (2019), pp. 103–110.

[173] Zhaobin Wang, E Wang, and Ying Zhu. “Image segmentation evaluation: a survey of methods”. In: *Artificial Intelligence Review* 53.8 (2020), pp. 5637–5674.

[174] Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

[175] Zian Wang et al. “Object instance annotation with deep extreme level set evolution”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7500–7508.

[176] Martin Weigert et al. “Content-aware image restoration: pushing the limits of fluorescence microscopy”. In: *Nature Methods* 15.12 (2018), pp. 1090–1097.

[177] Erin Weisbart and Beth A Cimini. “Distributed-Something: scripts to leverage AWS storage and computing for distributed workflows at scale”. In: *Nature Methods* 20.8 (2023), pp. 1120–1121.

[178] HS Wu, J Barba, and J Gil. “Iterative thresholding for segmentation of cells from noisy images.” In: *Journal of microscopy* 197.Pt 3 (2000), pp. 296–304.

[179] Enze Xie et al. “SegFormer: Simple and efficient design for semantic segmentation with transformers”. In: *Advances in neural information processing systems* 34 (2021), pp. 12077–12090.

[180] Saining Xie et al. “Aggregated residual transformations for deep neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1492–1500.

[181] Ning Xu et al. “Deep interactive object selection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 373–381.

[182] Rui Xu and Donald Wunsch. “Survey of clustering algorithms”. In: *IEEE Transactions on neural networks* 16.3 (2005), pp. 645–678.

[183] Jingru Yi et al. “Object-guided instance segmentation for biological images”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07. 2020, pp. 12677–12684.

[184] Andrew G. York et al. “Instant super-resolution imaging in live cells and embryos via analog image processing”. In: *Nature Methods* 10.11 (2013), pp. 1122–1126.

[185] Gregory W Zack, William E Rogers, and Samuel A Latt. “Automatic measurement of sister chromatid exchange frequency.” In: *Journal of Histochemistry & Cytochemistry* 25.7 (1977), pp. 741–753.

[186] Hui Zhang, Jason E Fritts, and Sally A Goldman. “Image segmentation evaluation: A survey of unsupervised methods”. In: *computer vision and image understanding* 110.2 (2008), pp. 260–280.

[187] Yu Jin Zhang. “A survey on evaluation methods for image segmentation”. In: *Pattern recognition* 29.8 (1996), pp. 1335–1346.

[188] Jingchun Zhou et al. “Fusion PSPnet image segmentation based method for multi-focus image fusion”. In: *IEEE Photonics Journal* 11.6 (2019), pp. 1–12.

A. Appendix