From Specific to Universal: One Biomedical Image Segmentation Model To Rule Them All

by

Seyed Alireza Vaezi

(Under the Direction of Shannon Quinn)

Abstract

This section of the page is where your abstract will go. Fill in this spot in the dissertation.tex file with your abstract. Doctoral 350 words max, 150 max for masters.

Index words: [PUT YOUR LIST OF INDEX WORDS HERE SEPERATED BY COMMAS]

From Specific to Universal: One Biomedical Image Segmentation Model To Rule Them All

by

Seyed Alireza Vaezi

M.S., Iran University of Science and Technology (IUST), Iran, 2016

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy of in Computer Science

Athens, Georgia

2025

From Specific to Universal: One Biomedical Image Segmentation Model To Rule Them All

by

Seyed Alireza Vaezi

Major Professor: Shannon Quinn

Committee: Hamid R. Arabnia
Tianming Liu
Kyle Johnsen

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
Month Year

Dedication

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

THIS PAGE IS OPTIONAL

Acknowledgments

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

THIS PAGE IS OPTIONAL

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Image segmentation started with traditional methods and algorithms such as thresholding, watershed, and optical flow, which rely on pixel intensity. Traditional rule-based methods like Thresholding and Watershed analyze pixel values to identify borders and boundaries within areas of interest. Machine learning-based methods - such as Support Vector Machines (SVM), random forests, and contrastive learning - involve the use of statistical machine learning models and increased the popularity and applicability of segmentation. Finally, deep learning-based methods leverage neural networks to learn hierarchical feature representation from raw images without requiring manual feature engineering. This process saw significant improvement with the introduction of Convolutional Neural Networks (CNN). CNNs are trained to detect features in regions of interest, enabling them to perform similarly on new images. Segmentation techniques can be divided into three categories: supervised, semi-supervised, and unsupervised.

U-Net and its variants are extensively adopted in biomedical image segmentation for their capability to automatically extract features from images without manual intervention or preprocessing. They can learn high-level semantic information and low-level spatial information from large-scale data. U-Net architectures are categorized based on their design and functionality. Basic U-Nets, like the original U-Net and 3D U-Net, are foundational, with the latter extending to 3D data for volumetric segmentation useful in CT and MRI scans. Advanced U-Nets include Attention U-Net, which uses attention mechanisms for precision; Inception U-Net, capturing multi-scale information through varied kernel sizes; Residual U-Net, which incorporates residual connections to aid deep network training; and Dense U-Net, promoting feature reuse via dense connections. Currently, CNNs represent the state-of-the-art in image segmentation, with U-Net being the predominant architecture, especially in the field of biomedical image segmentation. These advancements have greatly improved the accuracy and efficiency of biomedical image analysis, making it an essential tool in various fields of research.

Biomedical images come in a vast variety of formats, types, and modalities. The modalities in medical imaging include computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound, while microscopy modalities include fluorescent microscopy, bright-field, lens-free microscopy, light microscopy, volume electron microscopy, and phase contrast microscopy, just to name a few. Similarly, due to the variety of biological structures, segmentation targets can vary

from nuclei and cell membranes to organelles such as mitochondria, cilia, tumors, and lesions, as well as blood vessels, bone, and brain structures. This diversity in imaging techniques and segmentation targets highlights the need for specialized and customizable deep learning models in biomedical applications.

The Segment Anything Model (SAM) can segment an object within an image using user inputs, including a single point, multiple points, an entire mask, a bounding box, or textual descriptions. This functionality is based on the model's inherent ability to recognize objects, which enables it to segment unfamiliar object types without further training, effectively supporting zero-shot learning. Furthermore, the effectiveness of SAM is enhanced by its specialized architecture and the use of a significantly large dataset.

1.1 Challenges

Despite their success, CNN methods face challenges including poor generalizability, limited transferability, and the complexity of model development as well as fine-tuning pre-trained models in biomedical applications. This is due to the fact that manual labeling of data in biomedicine requires expert knowledge and is a costly and time-consuming task, making large and quality annotated datasets scarce. As a result, there exists a vast variety of deep learning models, each tailored to a specific modality and target structure. Unsupervised methods, on the other hand, do not require pre-training or an existing dataset and rely on domain-specific rules and heuristics. Although these methods exhibit less accuracy than CNN methods, they excel in reproducibility and generalizability as they do not depend on prior data knowledge. These different approaches to image segmentation provide a range of options for researchers to choose from, depending on their specific needs and resources.

In the biomedical field, where labeled data is often scarce and costly to obtain, several solutions have been proposed to augment and utilize available data effectively. These include semi-supervised learning, which utilizes both labeled and unlabeled data to enhance learning accuracy by leveraging the data's underlying distribution. Active learning focuses on selectively querying the most informative data points for expert labeling, optimizing the training process by using the most valuable examples. Data augmentation techniques, such as image transformations and synthetic data generation through Generative Adversarial Networks, increase the diversity and volume of training data, enhancing model robustness and reducing overfitting. Transfer learning transfers knowledge from one task to another, minimizing the need for extensive labeled data in new tasks. Self-supervised learning creates its labels by defining a pretext task, like predicting the position of a randomly cropped image patch, aiding in the learning of useful data representations. Additionally, few-shot, one-shot, and zero-shot learning techniques are designed to operate with minimal or no labeled examples, relying on generalization capabilities or metadata for making predictions about unseen classes.

Generalizability refers to the trained model's ability to perform well on unseen data outside of the training set. It is a crucial aspect of machine learning, particularly in biomedical applications, where variations in image acquisition conditions, tissue types, and other factors can be substantial. Data augmentation techniques and transfer learning are two excellent methods to overcome overfitting and improve general-

izability where the training data is small. While transfer learning is a powerful technique for leveraging pre-trained models to boost performance, especially in scenarios with limited data, it does come with its own set of challenges and limitations such as domain mismatch, risk of overfitting, computational demands, and potential biases from the source dataset.

Reproducibility refers to obtaining consistent results using the same input data, computational steps, methods, and conditions of analysis. This concept is key in scientific research to ensure that outcomes can be reliably replicated under the same conditions, fostering trust and confidence in the findings. Reproducibility is influenced by various factors including dataset variability, model architecture specifics, optimization procedures, and computational infrastructure. Apart from the loss of validity of a scientific method, non-reproducibility can lead to wasted resources, stalled scientific progress, erroneous conclusions, and significant ethical concerns. To ensure reproducibility in deep learning for medical image segmentation, Renard et al. advocate for comprehensive documentation, standardized practices, fixed random seeds, cross-validation, multiple evaluation metrics, and sharing of source code and dependencies.

Deep learning models are effective across various applications but their usability depends on several factors such as the complexity of the task at hand, data availability, and the extent of necessary model customization. For users who prefer straightforward applications, ease of use is crucial. They benefit from methods that do not require extensive modifications or tuning to achieve optimal results. Incorporating an intuitive graphical user interface (GUI) and ensuring interactivity can enhance the usability of these tools, making them more accessible to non-expert users, such as biologists, who need practical, ready-to-use solutions without the intricacies of model adjustments.

Chapter 2

Toxo and TSeg

Chapter 3

Training a Supervised Cilia Segmentation Model from Self-Supervision

Cilia are organelles found on the surface of some cells in the human body that sweep rhythmically to transport substances. Dysfunctional cilia are indicative of diseases that can disrupt organs such as the lungs and kidneys. Understanding cilia behavior is essential in diagnosing and treating such diseases. But, the tasks of automatically analysing cilia are often a labor and time-intensive since there is a lack of automated segmentation. In this work we overcome this bottleneck by developing a robust, self-supervised framework exploiting the visual similarity of normal and dysfunctional cilia. This framework generates pseudolabels from optical flow motion vectors, which serve as training data for a semi-supervised neural network. Our approach eliminates the need for manual annotations, enabling accurate and efficient segmentation of both motile and immotile cilia.

3.1 Introduction

Cilia are hair-like membranes that extend out from the surface of the cells and are present on a variety of cell types such as lungs and brain ventricles and can be found in the majority of vertebrate cells. Categorized into motile and primary, motile cilia can help the cell to propel, move the flow of fluid, or fulfill sensory functions, while primary cilia act as signal receivers, translating extracellular signals into cellular responses $[@doi : 10.1007/978 - 94 - 007 - 5808 - 7_1]$. Ciliopathies is the term commonly used to describe diseases caused by ciliary dysfunction. These disorders can result in serious issues such as blindness, neurodevelopmental defects, or obesity $[@Hansen2021 - fd]$. Motile cilia beat in a coordinated manner with a specific frequency and pattern $[@doi : 10.1016/j.compfluid.2011.05.016]$. Stationary, dyskinetic, or slow ciliary beating indicates ciliary defects. Ciliary beating is a fundamental biological process that is essential for the proper functioning of various organs, which makes understanding the

ciliary phenotypes a crucial step towards understanding ciliopathies and the conditions stemming from it [@$zain2022low$].

Identifying and categorizing the motion of cilia is an essential step towards understanding ciliopathies. However, this is generally an expert-intensive process. Studies have proposed methods that automate the ciliary motion assessment [@$zain2020towards$]. These methods rely on large amounts of labeled data that are annotated manually which is a costly, time-consuming, and error-prone task. Consequently, a significant bottleneck to automating cilia analysis is a lack of automated segmentation. Segmentation has remained a bottleneck of the pipeline due to the poor performance of even state-of-the-art models on some datasets. These datasets tend to exhibit significant spatial artifacts (light diffraction, out-of-focus cells, etc.) which confuse traditional image segmentation models [@$doi : 10.48550/arXiv.1803.07534$].

Video segmentation techniques tend to be more robust to such noise, but still struggle due to the wild inconsistencies in cilia behavior: while healthy cilia have regular and predictable movements, unhealthy cilia display a wide range of motion, including a lack of motion altogether [@$doi : 10.1002/ppul.24078$]. This lack of motion especially confounds movement-based methods which otherwise have no way of discerning the cilia from other non-cilia parts of the video. Both image and video segmentation techniques tend to require expert-labeled ground truth segmentation masks. Image segmentation requires the masks in order to effectively train neural segmentation models to recognize cilia, rather than other spurious textures. Video segmentation, by contrast, requires these masks in order to properly recognize both healthy and diseased cilia as a single cilia category, especially when the cilia show no movement.

To address this challenge, we propose a two-stage image segmentation model designed to obviate the need for expert-drawn masks. We first build a corpus of segmentation masks based on optical flow (OF) thresholding over a subset of healthy training data with guaranteed motility. We then train a semi-supervised neural segmentation model to identify both motile and immotile data as a single segmentation category, using the flow-generated masks as "pseudolabels". These pseudolabels operate as "ground truth" for the model while acknowledging the intrinsic uncertainty of the labels. The fact that motile and immotile cilia tend to be visually similar in snapshot allows us to generalize the domain of the model from motile cilia to all cilia. Combining these stages results in a semi-supervised framework that does not rely on any expert-drawn ground-truth segmentation masks, paving the way for full automation of a general cilia analysis pipeline.

3.2 Background

Dysfunction in ciliary motion indicates diseases known as ciliopathies, which can disrupt the functionality of critical organs like the lungs and kidneys. Understanding ciliary motion is crucial for diagnosing and understanding these conditions. The development of diagnosis and treatment requires the measurement of different cell properties including size, shape, and motility [@$vaezi2022novel$].

Accurate analysis of ciliary motion is essential but challenging due to the limitations of manual analysis, which is /labor-intensive, subjective, and prone to error. [@$zain2020towards$] proposed a modular generative pipeline that automates ciliary motion analysis by segmenting, representing, and modeling

the dynamic behavior of cilia, thereby reducing the need for expert intervention and improving diagnostic consistency. [@$quinn2015automated$] developed a computational pipeline using dynamic texture analysis and machine learning to objectively and quantitatively assess ciliary motion, achieving over 90% classification accuracy in identifying abnormal ciliary motion associated with diseases like primary ciliary dyskinesia (PCD). Additionally, [@$zain2022low$] explored advanced feature extraction techniques like Zero-phase PCA Sphering (ZCA) and Sparse Autoencoders (SAE) to enhance cilia segmentation accuracy. These methods address challenges posed by noisy, partially occluded, and out-of-phase imagery, ultimately improving the overall performance of ciliary motion analysis pipelines. Collectively, these approaches aim to enhance diagnostic accuracy and efficiency, making ciliary motion analysis more accessible and reliable, thereby improving patient outcomes through early and accurate detection of ciliopathies. However, these studies rely on manually labeled data. The segmentation masks and ground-truth annotations, which are essential for training the models and validating their performance, are generated by expert reviewers. This dependence on manually labeled data is a significant limitation making automated cilia segmentation the bottleneck to automating cilia analysis.

In the biomedical field, where labeled data is often scarce and costly to obtain, several solutions have been proposed to augment and utilize available data effectively. These include semi-supervised learning [@$YAKIMOVICH2021100383$], [@$van2020survey$], which utilizes both labeled and unlabeled data to enhance learning accuracy by leveraging the data's underlying distribution. Active learning [@$settles2009active$] focuses on selectively querying the most informative data points for expert labeling, optimizing the training process by using the most valuable examples. Data augmentation techniques [@10.3389/$fcvm$.2020.00105], [@$Krois2021$], [@10.1148/$ryai$.2020190195], [@$Sandfort2019$], [@$YAKIMOVICH2021100383$], [@$van2001art$], [@$krizhevsky2012imagenet$], [@$ronneberger2015u$], such as image transformations and synthetic data generation through Generative Adversarial Networks [@$goodfellow2014generative$], [@$yi2019generative$], increase the diversity and volume of training data, enhancing model robustness and reducing overfitting. Transfer learning [@$YAKIMOVICH2021100383$], [@$Sanford2020-yg$], [@$NEURIPS2019_e$b1e7832], [@$hutchinson2017overcoming$] transfers knowledge from one task to another, minimizing the need for extensive labeled data in new tasks. Self-supervised learning [@$kim2019self$], [@$kolesnikov2019revisiting$], [@$mahendran2019cross$] creates its labels by defining a pretext task, like predicting the position of a randomly cropped image patch, aiding in the learning of useful data representations. Additionally, few-shot, one-shot, and zero-shot learning techniques [@$li2006one$], [@$miller2000learning$] are designed to operate with minimal or no labeled examples, relying on generalization capabilities or metadata for making predictions about unseen classes.

A promising approach to overcome the dependency on manually labeled data is the use of unsupervised methods to generate ground truth masks. Unsupervised methods do not require prior knowledge of the data [@$khatibi2021proposing$]. Using domain-specific cues unsupervised learning techniques can automatically discover patterns and structures in the data without the need for labeled examples, potentially simplifying the process of generating accurate segmentation masks for cilia. Inspired by advances in unsupervised methods for image segmentation, in this work, we firstly compute the motion vectors

using optical flow of the ciliary regions and then apply autoregressive modelling to capture their temporal dynamics. Autoregressive modelling is advantageous since the labels are features themselves. By analyzing the OF vectors, we can identify the characteristic motion of cilia, which allows us to generate pseudolabels as ground truth segmentation masks. These pseudolabels are then used to train a robust semi-supervised neural network, enabling accurate and automated segmentation of both motile and immotile cilia.

3.3 Methodology

Dynamic textures, such as sea waves, smoke, and foliage, are sequences of images of moving scenes that exhibit certain stationarity properties in time $[@doretto2003dynamic]$. Similarly, ciliary motion can be considered as dynamic textures for their orderly rhythmic beating. Taking advantage of this temporal regularity in ciliary motion, OF can be used to compute the flow vectors of each pixel of high-speed videos of cilia. In conjunction with OF, autoregressive (AR) parameterization of the OF property of the video yields a manifold that quantifies the characteristic motion in the cilia. The low dimension of this manifold contains the majority of variations within the data, which can then be used to segment the motile ciliary regions.

3.4 Optical Flow Properties

Taking advantage of this temporal regularity in ciliary motion, we use OF to capture the motion vectors of ciliary regions in high-speed videos. OF provides the horizontal u and vertical v components of the motion for each pixel. From these motion vectors, several components can be derived such as the magnitude, direction, divergence, and importantly, the curl (rotation). The curl, in this context, represents the rotational motion of the cilia, which is indicative of their rhythmic beating patterns. We extract flow vectors of the video recording of cilia, under the assumption that pixel intensity remains constant throughout the video.

$$I(x, y, t) = I(x + u\delta t, y + v\delta t, t + \delta t) \tag{3.4.1}$$

(3.4.1) Where $I_{x,y,t}$ is the pixel intensity at position x, y a time t. Here, u_t, v_t are small changes in the next frame taken after t time, and u, v, respectively, are the OF components that represent the displacement in pixel positions between consecutive frames in the horizontal and vertical directions at pixel location x, y.

3.5 Autoregressive Modeling

Figure 3.1 shows a sample of the OF component at a random time. From OF vectors, elemental components such as rotation are derived, which highlights the ciliary motion by capturing twisting and turning movements. To model the temporal evolution of these motion vectors, we employ an autoregressive

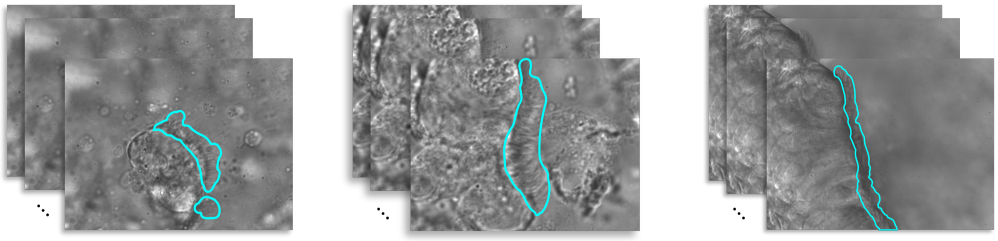

Figure 3.1: A sample of three videos in our cilia dataset with their manually annotated ground truth masks.

(AR) model [$@doi : 10.5244/C.21.76$]. This model captures the dynamics of the flow vectors over time, allowing us to understand how the motion evolves frame by frame. The AR model helps in decomposing the motion into a low-dimensional subspace, which simplifies the complex ciliary motion into more manageable analyses.

$$y_t = C\vec{x_t} + \vec{u} \tag{3.5.1}$$

$$\vec{x_t} = A_1\vec{x}_{t-1} + A_2\vec{x}_{t-2} + ... + A_d\vec{x}_{t-d} + \vec{v_t} \tag{3.5.2}$$

In equation (3.5.1), y_t represents the appearance of cilia at time t influenced by noise u. Equation (3.5.2) represents the state x of the ciliary motion in a low-dimensional subspace defined by an orthogonal basis C at time t, plus a noise term v_t and how the state changes from t to $t + 1$.

Equation (3.5.2) is a decomposition of each frame of a ciliary motion video yt into a low-dimensional state vector x_t using an orthogonal basis C. This equation at position x_t is a function of the sum of d of its previous positions $x_{t-1}, x_{t-2}, x_{t-d}$ each multiplied by its corresponding coefficients $A = A_1, A_2, ..., A_d$. The noise terms u and v are used to represent the residual difference between the observed data and the solutions to the linear equations. The variance in the data is predominantly captured by a few dimensions of C, simplifying the complex motion into manageable analyses.

Each order of the autoregressive model roughly aligns with different frequencies within the data, therefore, in our experiments, we chose d=5 as the order of our autoregressive model. This choice allows us to capture a broader temporal context, providing a more comprehensive understanding of the system's dynamics. We then created raw masks from this lower-dimensional subspace, and further enhanced them with adaptive thresholding to remove the remaining noise.

In 3.2 , the first-order AR parameter is showing the most variance in the video, which corresponds to the frequency of motion that cilia exhibit. The remaining orders have correspondence with other different frequencies in the data caused by, for instance, camera shaking. Evidently, simply thresholding the first-order AR parameter is adequate to produce an accurate mask, however, in order to get a more refined result we subtracted the second order from the first one, followed by a Min-Max normalization of pixel intensities and scaling to an 8-bit unsigned integer range. We used adaptive thresholding to extract the

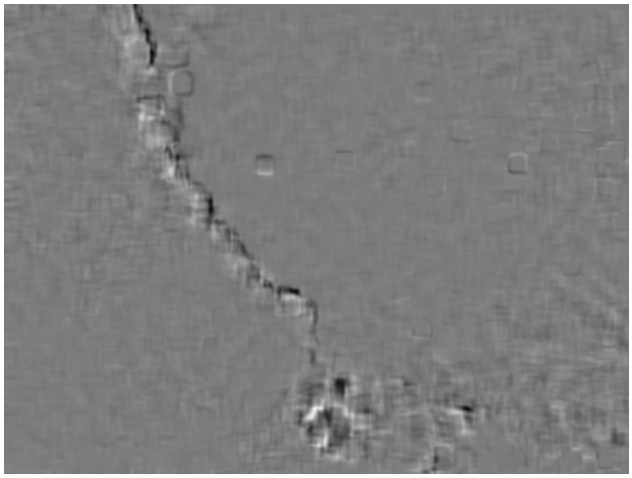

Figure 3.2: Representation of rotation (curl) component of OF at a random time.

mask on all videos of our dataset. The generated masks exhibited under-segmentation in the ciliary region, and sparse over-segmentation in other regions of the image. To overcome this, we adapted a Gaussian blur filter followed by an Otsu thresholding to restore the under-segmentation and remove the sparse over-segmentation. Figure 3.4 illustrates the steps of the process.

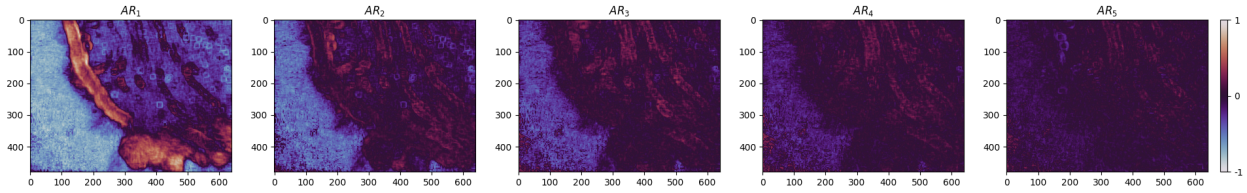

Figure 3.3: The pixel representation of the 5-order AR model of the OF component of a sample video. The x and y axes correspond to the width and height of the video.

3.6 Training the model

Our dataset includes 512 videos, with 437 videos of dyskinetic cilia and 75 videos of healthy motile cilia, referred to as the control group. The control group is split into %85 and %15 for training and validation respectively. 108 videos in the dyskinetic group are manually annotated which are used in the testing step. Figure 3.1 shows annotated samples of our dataset.

In our study, we employed a Feature Pyramid Network (FPN) [@$kirillov2017unified$] architecture with a ResNet-34 encoder. The model was configured to handle grayscale images with a single input

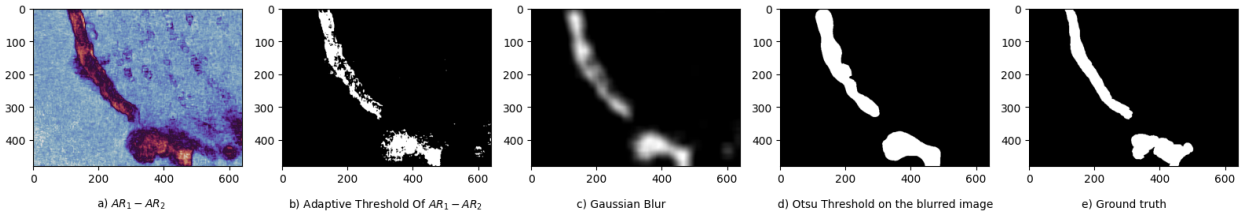

Figure 3.4: The process of computing the masks. **a)** Subtracting the second-order AR parameter from the first-order, followed by **b)** Adaptive thresholding, which suffers from under/over-segmentation. **c)** A Gaussian blur filter, followed by **d)** An Otsu thresholding eliminates the under/over-segmentation.

channel and produce binary segmentation masks. For the training input, one mask is generated per video using our methodology, and we use the first 250 frames from each video in the control group making a total of 18,750 input images. We utilized Binary Cross-Entropy Loss for training and the Adam optimizer with a learning rate of 10^{-3}. To evaluate the model's performance, we calculated the Dice score during training and validation. Data augmentation techniques, including resizing, random cropping, and rotation, were applied to enhance the model's generalization capability. The implementation was done using a library [$@Iakubovskii : 2019$] based on PyTorch Lightning to facilitate efficient training and evaluation. Table 3.1 contains a summary of the model parameters and specifications.

Table 3.1: Summary of model architecture, training setup, and dataset distribution

Aspect	**Details**
Architecture	FPN with ResNet-34 encoder
Input	Grayscale images with a single input channel
Number of Epochs	20
Batch Size	4
Training Samples	15,662
Validation Samples	2,763
Test Samples	108
Loss Function	Binary Cross-Entropy Loss
Optimizer	Adam optimizer with a learning rate of 10^{-3}
Evaluation Metric	Dice score during training and validation
Data Augmentation Techniques	Resizing, random cropping, and rotation
Implementation	Using a Python library with Neural Networks for Image Segmentation based on PyTorch $Iakubovskii : 2019$

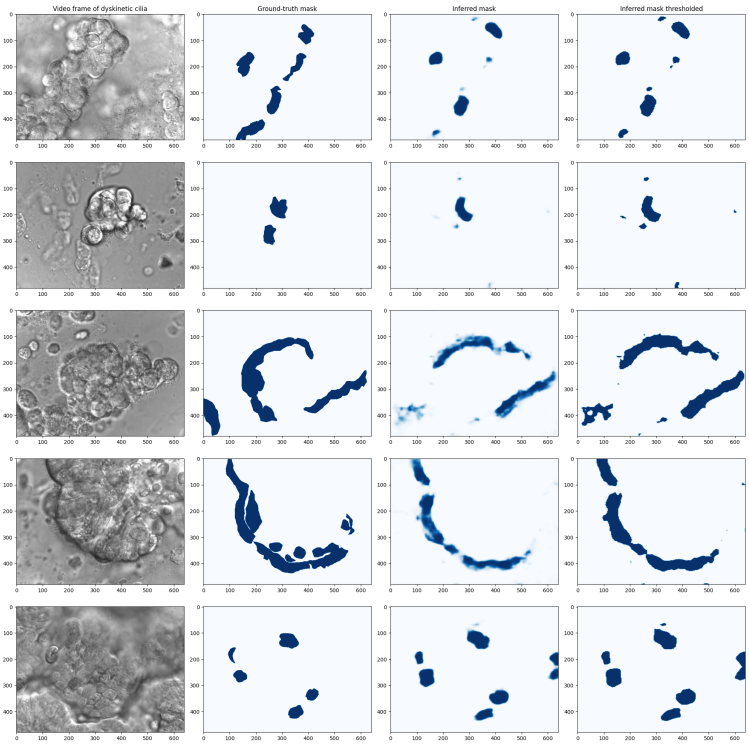

Figure 3.5: The model predictions on 5 dyskinetic cilia samples. The first column shows a frame of the video, the second column shows the manually labeled ground truth, the third column is the model's prediction, and the last column is a thresholded version of the prediction.

3.7 Results and Discussion

The model's performance metrics, including IoU, Dice score, sensitivity, and specificity, are summarized in @tbl:metrics. The validation phase achieved an IoU of 0.312 and a Dice score of 0.476, which indicates a moderate overlap between the predicted and ground truth masks. The high sensitivity (0.999) observed during validation suggests that the model is proficient in identifying ciliary regions, albeit with a specificity of 0.813, indicating some degree of false positives. In the testing phase, the IoU and Dice scores decreased to 0.230 and 0.374, respectively, reflecting the challenges posed by the dyskinetic cilia data, which were not included in the training or validation sets. Despite this, the model maintained a reasonable sensitivity of 0.631 and specificity of 0.787.

Figure 3.5 provides visual examples of the model's predictions on dyskinetic cilia samples, alongside the manually labeled ground truth and thresholded predictions. The dyskinetic samples were not used in the training or validation phases. These predictions were generated after only 20 epochs of training with a small training data. The visual comparison reveals that, while the model captures the general structure of ciliary regions, there are instances of under-segmentation and over-segmentation, which are

more pronounced in the dyskinetic samples. This observation is consistent with the quantitative metrics, suggesting that further refinement of the pseudolabel generation process or model architecture could enhance segmentation accuracy.

Table 3.2: The performance of the model in validation and testing phases.

Phases	**IoU over dataset**	**Dice Score**	**Sensitivity**	**Specificity**
Validation	0.312	0.476	0.999	0.813
Testing	0.230	0.374	0.631	0.787

These results show the potential of our approach to reduce the reliance on manually labeled data for cilia segmentation. The use of this unsupervised learning framework allows the model to generalize from the motile cilia domain to the more variable dyskinetic cilia, although with some limitations in accuracy. Future work could focus on expanding the dataset and improving the process of generating pseudolabels to enhance the model's accuracy.

3.8 Conclusions and Final Remarks

In this study, we introduced a self-supervised framework for cilia segmentation that eliminates the need for expert-labeled ground truth segmentation masks. Our approach takes advantage of the inherent visual similarities between healthy and unhealthy cilia to generate pseudolabels from optical flow-based motion segmentation of motile cilia. These pseudolabels are then used as ground truth for training a semi-supervised neural network capable of identifying regions containing dyskinetic cilia. Our results indicate that the self-supervised framework is a promising step towards automated cilia analysis. The model's ability to generalize from motile to dyskinetic cilia demonstrates its potential applicability in clinical settings. Although there are areas for improvement, such as enhancing segmentation accuracy and expanding the dataset, the framework sets the foundation for more efficient and reliable cilia analysis pipelines.

Chapter 4

Enter Title Here

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Chapter 5

Conclusion

This is where your Conclusion will go

Appendix A

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.