Open Access | https://doi.org/10.25080/HXCJ6205

SciPy 2024

July 8 - July 14, 2024

Proceedings of the 23nd Python in Science Conference
ISSN: 2575-9752

Training a Supervised Cilia Segmentation Model from Self-Supervision

Seyed Alireza Vaezi[1], and **Shannon Quinn**[1]

[1]University of Georgia

Abstract

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 analyzing 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.

Keywords Cilia, Unsupervised biomedical Image Segmentation, Optical Flow, Autoregressive, Deep Learning

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 [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 [2]. Motile cilia beat in a coordinated manner with a specific frequency and pattern [3]. 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 [4].

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 [5]. 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 [6].

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

Published Jul 10, 2024

Correspondence to
Seyed Alireza Vaezi
sv22900@uga.edu

altogether [7]. 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.

The rest of this article is structured as follows: The Section 2 enumerates the studies relevant to our methodology, followed by a detailed description of our approach in the Section 3. Finally, the Section 4 delineates our experiment and provides a discussion of the results obtained.

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 [8].

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. [5] 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. [9] 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, [4] 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 [10], which utilizes both labeled and unlabeled data to enhance learning accuracy by leveraging the data's underlying distribution. Active learning [11] 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 [12], [13], [14], [15], [16], [17], [18], [19], such as image transformations and synthetic data generation through Generative Adversarial Networks [20], [21], increase the diversity and volume of training data, enhancing model robustness and reducing overfitting. Transfer learning [16], [22], [23], [24] transfers knowledge from one task to another, minimizing the need for extensive labeled data in new tasks. Self-supervised learning [25], [26], [27] 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 [28], [29] 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 [30]. 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. Methodology

Dynamic textures, such as sea waves, smoke, and foliage, are sequences of images of moving scenes that exhibit certain stationarity properties in time [31]. 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.

3.1. 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.

$$I(x, y, t) = I(x + u\delta t, y + v\delta t, t + \delta t) \tag{1}$$

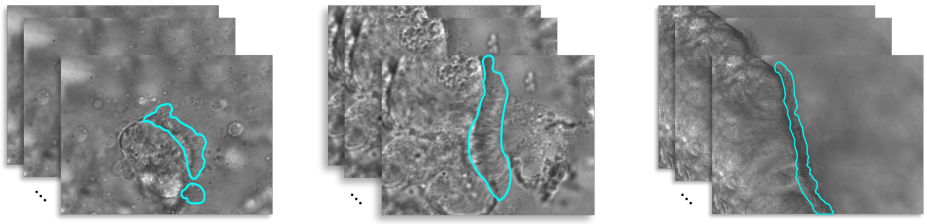

Figure 1. *A sample of three videos in our cilia dataset with their manually annotated ground truth masks.*

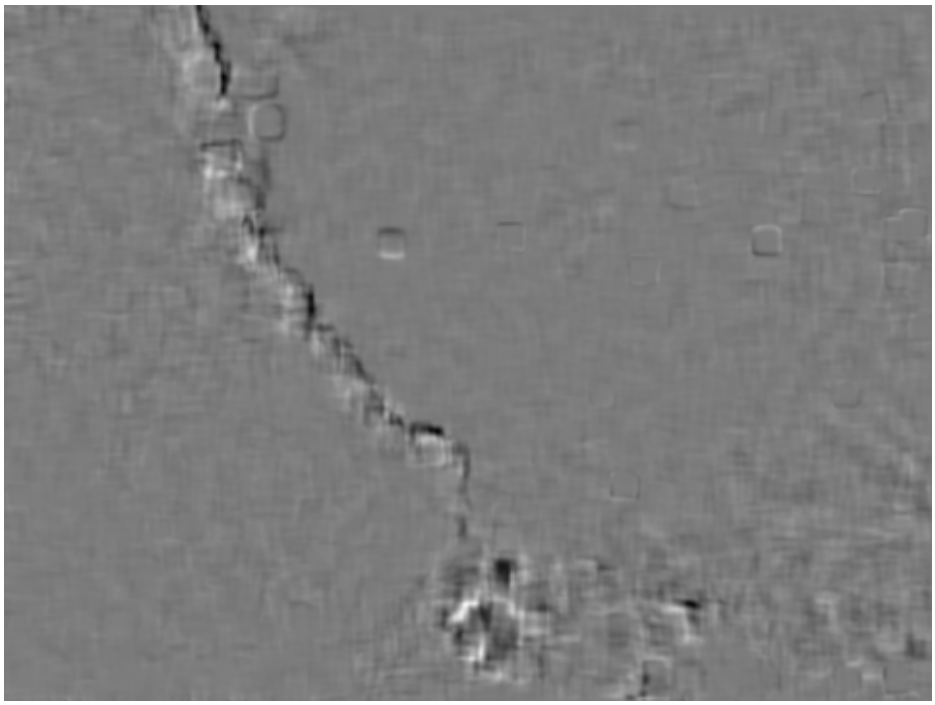

Figure 2. *Representation of rotation (curl) component of OF at a random time*

Where $I(x, y, t)$ is the pixel intensity at position (x, y) a time t. Here, $(u\delta t, v\delta t)$ are small changes in the next frame taken after δt 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).

3.2. Autoregressive Modeling

Figure 2 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 (AR) model [32]. 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.

$$y_t = C\vec{x_t} + \vec{u} \tag{2}$$

$$\vec{x}_t = A_1\vec{x}_{t-1} + A_2\vec{x}_{t-2} + ... + A_d\vec{x}_{t-d} + \vec{v_t} \tag{3}$$

In equation Equation 2, $\vec{y}_t$ represents the appearance of cilia at time t influenced by noise $\vec{u}$. Equation Equation 3 represents the state $\vec{x}$ of the ciliary motion in a low-dimensional subspace defined by an orthogonal basis C at time t, plus a noise term $\vec{v}_t$ and how the state changes from t to $t + 1$.

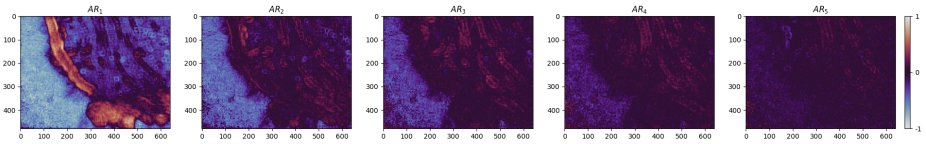

Figure 3. *The pixel representation of the 5-order AR model of the OF component of a sample video. The x and y axes correspond to the width and height of the video.*

Equation Equation 3 is a decomposition of each frame of a ciliary motion video $\vec{y}_t$ into a low-dimensional state vector $\vec{x}_t$ using an orthogonal basis C. This equation at position x_t is a function of the sum of d of its previous positions $\vec{x}_{t-1}, \vec{x}_{t-2}, \vec{x}_{t-d}$, each multiplied by its corresponding coefficients $A = A_1, A_2, ..., A_d$. The noise terms $\vec{u}$ and $\vec{v}$ 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 Figure 3, 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 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 4 illustrates the steps of the process.

3.3. 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 1 shows annotated samples of our dataset.

In our study, we employed a Feature Pyramid Network (FPN) [33] architecture with a ResNet-34 encoder. The model was configured to handle grayscale images with a single

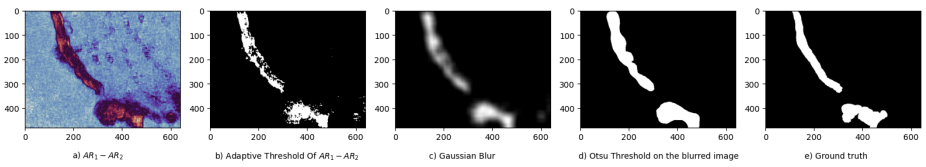

Figure 4. *The process of computing the masks. a) Subtracting the second-order AR parameter from the first-order, followed by b) Adaptive thresholding, which suffers from under/over-segmentation. c) A Gaussian blur filter, followed by d) An Otsu thresholding eliminates the under/over-segmentation.*

Aspect	Details
Architecture	FPN with ResNet-34 encoder
Input	Grayscale images with a single input channel
Batch Size	2
Training Samples	28,869
Validation Samples	5,095
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, validation, and testing
Data Augmentation Techniques	Resizing, random cropping, and rotation
Implementation	Using a Python library with Neural Networks for Image Segmentation based on PyTorch [34]

Table 1. *Summary of model architecture, training setup, and dataset distribution*

input channel and produce binary segmentation masks. For the training input, one mask is generated per video using our methodology, and we use all of the frames from each video in the control group making a total of 33,964 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 [34] based on PyTorch Lightning to facilitate efficient training and evaluation. Table 1 contains a summary of the model parameters and specifications.

The next section discusses the results of the experiment and the performance of the model in detail.

4. Results and Discussion

The model's performance metrics, including IoU, Dice score, sensitivity, and specificity, are summarized in Table 2. The validation phase achieved an IoU of 0.398 and a Dice score of 0.569, which indicates a moderate overlap between the predicted and ground truth masks. The high sensitivity (0.997) observed during validation suggests that the model is proficient in identifying ciliary regions, albeit with a specificity of 0.882, indicating some degree of false positives. In the testing phase, the IoU and Dice scores decreased to 0.132 and 0.233, 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 sensitivity of 0.479 and specificity of 0.806.

Figure 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 15 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.

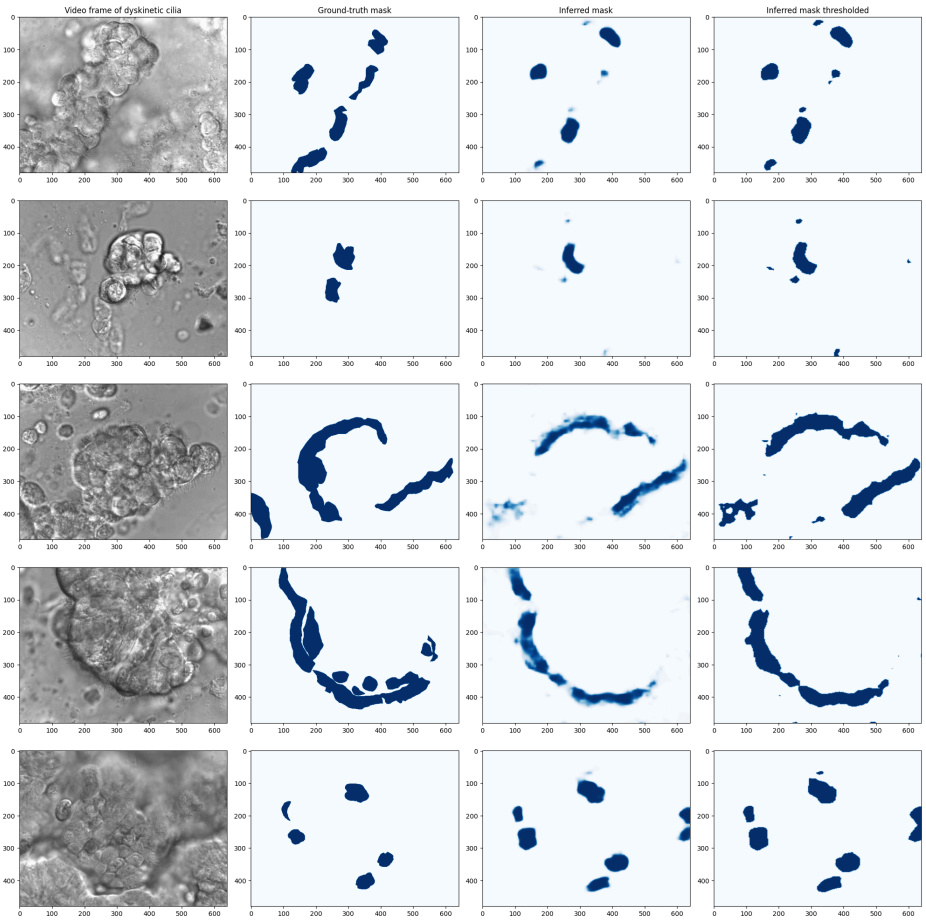

Figure 5. *The model predictions on 5 dyskinetic cilia samples. The first column shows a frame of the video, the second column shows the manually labeled ground truth, the third column is the model's prediction, and the last column is a thresholded version of the prediction.*

4.1. Training the model using control and dyskinetic data

Since dyskinetic videos contain cilia that show some degree of movement we generated pseudo-labels for 283 dyskinetic videos and used them along with the 76 control videos from the previous experiment in another experiment. Training the model again for 15 epochs over healthy and dyskinetic videos exhibited a loss of performance in the validation phase, however, in the testing phase all of the metrics improved except for the specificity. Since in this experiment the model was trained on an additional subset of the dyskinetic videos, improved performance in detecting and masking dyskinetic ciliary regions is expected. The results are depicted in Table 3.

After using the model to infer more samples we detected a pattern for how the model performs in regions with specific visual properties. We observed that the model can more successfully and more accurately detect ciliary regions in images where they appear sharp

Phases	**Metrics**			
	IoU over dataset	Dice Score	Sensitivity	Specificity
Validation	0.398	0.569	0.997	0.882
Testing	0.132	0.233	0.479	0.806

Table 2. *The performance of the model in validation and testing phases after 15 epochs of training.*

Phases	Metrics			
	IoU over dataset	Dice Score	Sensitivity	Specificity
Validation	0.202	0.337	0.999	0.765
Testing	0.139	0.245	0.732	0.696

Table 3. *The performance of the model after retraining with an addition of 283 videos of dyskinetic cilia to the training dataset.*

and in focus, and do not overlap other cellular structures. On the other hand, as shown in all samples in Figure 5, the most number of false negatives occur where the ciliary regions are in close proximity to other cellular structures or overlapping them. Furthermore, the most false positives occur along sharp cellular borders. Cell borders are most where cilia can be found the most in the videos, and the model may have learnt to look for or prioritize sharp cell borders and boundaries as ciliary regions. More investigation is required to further examine whether the model's attention mechanism or feature extraction layers are overly biased towards sharp edges and boundaries, potentially leading to incorrect predictions. This investigation could involve analyzing the model's learned features, adjusting training strategies, or incorporating additional data augmentation techniques to improve its performance in complex regions.

The 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.

5. Conclusions

In this paper, 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.

References

[1] S. Hoyer-Fender, "Primary and Motile Cilia: Their Ultrastructure and Ciliogenesis," in *Cilia and Nervous System Development and Function*, K. L. Tucker and T. Caspary, Eds., Dordrecht: Springer Netherlands, 2013, pp. 1–53. doi: 10.1007/978-94-007-5808-7_1.

[2] J. N. Hansen, S. Rassmann, B. Stüven, N. Jurisch-Yaksi, and D. Wachten, "CiliaQ: a simple, open-source software for automated quantification of ciliary morphology and fluorescence in 2D, 3D, and 4D images," *The European Physical Journal E*, vol. 44, no. 2, p. 18, 2021, doi: https://doi.org/10.1140/epje/s10189-021-00031-y.

[3] W. Lee, P. Jayathilake, Z. Tan, D. Le, H. Lee, and B. Khoo, "Muco-ciliary transport: effect of mucus viscosity, cilia beat frequency and cilia density," *Computers & Fluids*, vol. 49, no. 1, pp. 214–221, 2011, doi: https://doi.org/10.1016/j.compfluid.2011.05.016.

[4] M. Zain, E. Miller, S. Quinn, and C. Lo, "Low Level Feature Extraction for Cilia Segmentation," in *Proceedings of the Python in Science Conference*, 2022. doi: https://doi.org/10.25080/majora-212e5952-026.

[5] M. Zain *et al.*, "Towards an unsupervised spatiotemporal representation of cilia video using a modular generative pipeline," in *Proceedings of the Python in Science Conference*, 2020. doi: http://dx.doi.org/10.25080/Majora-342d178e-017.

[6] C. Lu, M. Marx, M. Zahid, C. W. Lo, C. Chennubhotla, and S. P. Quinn, "Stacked Neural Networks for end-to-end ciliary motion analysis," *arXiv preprint arXiv:1803.07534*, 2018, doi: https://doi.org/10.48550/arXiv.1803.07534.

[7] C. Kempeneers and M. A. Chilvers, "To beat, or not to beat, that is question! The spectrum of ciliopathies," *Pediatric Pulmonology*, vol. 53, no. 8, pp. 1122–1129, 2018, doi: https://doi.org/10.1002/ppul.24078.

[8] S. A. Vaezi, G. Orlando, M. S. Fazli, G. E. Ward, S. N. Moreno, and S. Quinn, "A Novel Pipeline for Cell Instance Segmentation, Tracking and Motility Classification of Toxoplasma Gondii in 3D Space.," in *SciPy*, 2022, pp. 60–63. doi: https://doi.org/10.25080/majora-212e5952-009.

[9] S. P. Quinn, M. J. Zahid, J. R. Durkin, R. J. Francis, C. W. Lo, and S. C. Chennubhotla, "Automated identification of abnormal respiratory ciliary motion in nasal biopsies," *Science translational medicine*, vol. 7, no. 299, p. 299, 2015, doi: http://dx.doi.org/10.1126/scitranslmed.aaa1233.

[10] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine learning*, vol. 109, no. 2, pp. 373–440, 2020, doi: http://dx.doi.org/10.1007/s10994-019-05855-6.

[11] B. Settles, "Active learning literature survey," 2009.

[12] C. Chen *et al.*, "Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images," *Frontiers in Cardiovascular Medicine*, vol. 7, 2020, doi: https://doi.org/10.3389/fcvm.2020.00105.

[13] J. Krois *et al.*, "Generalizability of deep learning models for dental image analysis," *Scientific Reports*, vol. 11, no. 1, p. 6102, 2021, doi: http://dx.doi.org/10.1038/s41598-021-85454-5.

[14] W. Yan *et al.*, "MRI Manufacturer Shift and Adaptation: Increasing the Generalizability of Deep Learning Segmentation for MR Images Acquired with Different Scanners," *Radiology: Artificial Intelligence*, vol. 2, no. 4, p. e190195, 2020, doi: 10.1148/ryai.2020190195.

[15] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Scientific Reports*, vol. 9, no. 1, p. 16884, 2019, doi: https://doi.org/10.1016/j.imu.2021.100779.

[16] A. Yakimovich, A. Beaugnon, Y. Huang, and E. Ozkirimli, "Labels in a haystack: Approaches beyond supervised learning in biomedical applications," *Patterns*, vol. 2, no. 12, p. 100383, 2021, doi: https://doi.org/10.1016/j.patter.2021.100383.

[17] D. A. Van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001, doi: http://dx.doi.org/10.1198/10618600152418584.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012, doi: 10.1145/3065386.

[19] O. Ronneberger, P. Fischer, and T. 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*, 2015, pp. 234–241. doi: 10.48550/arXiv.1505.04597.

[20] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014, doi: 10.48550/arXiv.1406.2661.

[21] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical image analysis*, vol. 58, p. 101552, 2019, doi: https://doi.org/10.48550/arXiv.1809.07294.

[22] T. H. Sanford *et al.*, "Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model," *AJR Am J Roentgenol*, vol. 215, no. 6, pp. 1403–1410, 2020, doi: http://dx.doi.org/10.2214/AJR.19.22347.

[23] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds., 2019. doi: https://doi.org/10.48550/arXiv.1902.07208.

[24] M. L. Hutchinson, E. Antono, B. M. Gibbons, S. Paradiso, J. Ling, and B. Meredig, "Overcoming data scarcity with transfer learning." 2017. doi: https://doi.org/10.48550/arXiv.1711.05099.

[25] D. Kim, D. Cho, and I. S. Kweon, "Self-supervised video representation learning with space-time cubic puzzles," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, pp. 8545–8552. doi: https://doi.org/10.48550/arXiv.1811.09795.

[26] A. Kolesnikov, X. Zhai, and L. Beyer, "Revisiting self-supervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1920–1929. doi: https://doi.org/10.48550/arXiv.1901.09005.

[27] A. Mahendran, J. Thewlis, and A. Vedaldi, "Cross pixel optical-flow similarity for self-supervised learning," in *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part V 14*, 2019, pp. 99–116. doi: https://doi.org/10.48550/arXiv.1807.05636.

[28] F.-F. Li, R. Fergus, P. Perona, and others, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 28, no. 4, pp. 594–611, 2006, doi: http://dx.doi.org/10.1109/TPAMI.2006.79.

[29] E. G. Miller, N. E. Matsakis, and P. A. Viola, "Learning from one example through shared densities on transforms," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, 2000, pp. 464–471. doi: https://doi.org/10.1109/CVPR.2000.855856.

[30] T. Khatibi, N. Rezaei, L. Ataei Fashtami, and M. Totonchi, "Proposing a novel unsupervised stack ensemble of deep and conventional image segmentation (SEDCIS) method for localizing vitiligo lesions in skin images," *Skin Research and Technology*, vol. 27, no. 2, pp. 126–137, 2021, doi: http://dx.doi.org/10.1111/srt.12920.

[31] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," *International journal of computer vision*, vol. 51, pp. 91–109, 2003, doi: https://doi.org/10.1023/A:1021669406132.

[32] M. Hyndman, A. D. Jepson, and D. J. Fleet, "Higher-order Autoregressive Models for Dynamic Textures," in *British Machine Vision Conference*, 2007. doi: http://dx.doi.org/10.5244/C.21.76.

[33] A. Kirillov, K. He, R. Girshick, and P. Dollár, "A unified architecture for instance and semantic segmentation," in *Computer Vision and Pattern Recognition Conference*, 2017. doi: https://doi.org/10.48550/arXiv.2112.04603.

[34] P. Iakubovskii, "Segmentation Models Pytorch." [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch