STACKED NEURAL NETWORKS FOR END-TO-END CILIARY MOTION ANALYSIS

Charles Lu[⋆] *M. Marx*[†] *M. Zahid*[§] *C. W. Lo*[§] *C. Chennubhotla*[†] *S. P. Quinn*[⋆‡]

[⋆] Department of Computer Science, University of Georgia, Athens, GA USA
[†] Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA USA
[§] Department of Developmental Biology, University of Pittsburgh Medical Center, Pittsburgh, PA USA
[‡] Corresponding author {squinn@cs.uga.edu}

ABSTRACT

Cilia are hairlike structures protruding from nearly every cell in the body. Diseases known as *ciliopathies*, where cilia function is disrupted, can result in a wide spectrum of disorders. However, most techniques for assessing ciliary motion rely on manual identification and tracking of cilia; this process is laborious and error-prone, and does not scale well. Even where automated ciliary motion analysis tools exist, their applicability is limited. Here, we propose an end-to-end computational machine learning pipeline that automatically identifies regions of cilia from videos, extracts patches of cilia, and classifies patients as exhibiting normal or abnormal ciliary motion. In particular, we demonstrate how convolutional LSTM are able to encode complex features while remaining sensitive enough to differentiate between a variety of motion patterns. Our framework achieves 90% with only a few hundred training epochs. We find that the combination of segmentation and classification networks in a single pipeline yields performance comparable to existing computational pipelines, while providing the additional benefit of an end-to-end, fully-automated analysis toolbox for ciliary motion.

Index Terms— Cilia, Ciliopathies, Semantic Segmentation, Convolutional Neural Networks, Recurrent Neural Networks, Computer Vision

1. INTRODUCTION

Cilia are microtubule based hair-like projections of the cell that can be motile or immotile, and in humans are found on nearly every cell of the body. Ciliopathies, or diseases with disruption of nonmotile or motile cilia function, can result in a wide spectrum of disorders, ranging from sinopulmonary disease such as in primary ciliary dyskinesia (PCD) [1], to mirror symmetric organ placement or randomized left-right organ placement as in heterotaxy [2]. Each of these conditions are associated with increased respiratory complications and poor postsurgical outcomes [3, 4]. Diagnosing patients with ciliary motion (CM) abnormalities prior to surgery may provide the clinician with opportunities to institute prophylactic respiratory therapies to prevent these complications. Together, these findings suggest motile cilia dysfunction may have a very broad clinical impact.

Current methods for assessing CM rely on a combination of techniques often used in concert, including electron microscopy [5], ciliary beat frequency (CBF) [6, 7], and visual assessment of ciliary beat pattern by expert reviewers [8, 9]. However, each of these methods has drawbacks and limitations [10]; in addition, none are amenable to cross-institutional comparisons and collaborations. Some semi-automated methods have been proposed [10, 11], but these are all of limited utility to clinicians, requiring some form of manual annotation.

To overcome these deficiencies, we have developed an end-to-end computational pipeline using deep neural networks. Deep learning approaches have attained state of the art performance on many benchmark datasets in biomedical imaging, and are ideal for spatiotemporal analysis. Our pipeline automatically identifies regions of high-speed digital videos of ciliary biopsies which contain *beating* or *non-beating* cilia, and extracts them for downstream analysis. Once the regions of cilia are extracted, the temporal behavior of the cilia is parameterized and used to train a binary classifier. Finally, the classifier predicts the motion of the cilia for each patient. Once the stacked deep nets are trained, videos of ciliary biopsies can be fed directly into our pipeline and a CM prediction rendered, without requiring any manual intervention from clinicians.

2. DATA

We used data from our previous work [10] that includes: nasal brush biopsies from 75 patients (35 healthy controls, 40 with a diagnosed ciliopathy), totaling 268 videos. Nasal epithelial tissue was collected by curettage of the inferior nasal turbinate under direct visualization with an appropriately sized nasal speculum using Rhino-probe (Arlington Scientific). Three passages were made, and the collected tissue was resuspended in L-15 medium (Invitrogen) for immediate videomicroscopy using a Leica inverted microscope with a 100× oil objective and differential interference contrast optics. Digital high-speed videos were recorded at a sampling frequency of 200 Hz using a Phantom v4.2 camera. To establish ground truth CM, these samples were analyzed by a panel of researchers blinded to the subject's clinical diagnosis and associated pathology reports. After reviewing all videos associated with a patient, a call of normal or abnormal CM was made by consensus. Where differences could not be resolved, the majority vote was accepted.

3. METHODS

3.1. Data Preprocessing and Augmentation

Each video depicted cilia combined with varying levels of recording artifacts such as extraneous camera movement, uneven lighting, or poor focus. Additionally, cilia could be depicted at different angles relative to the camera's perspective, drastically altering appearance. To address this, we created four annotation "classes" in the first of our stacked deep networks: side-view (lateral) cilia, top-down cilia, cell body, and background. These four annotation classes were used for creating ground-truth segmentation masks using ITK-SNAP for a small number of videos. Random crops of 256 × 256 and horizontal vertical flips were used for data augmentation. Some other videos were discarded due to poor quality recording and absence of discernible cilia.

After identifying spatial regions containing cilia, we computed spatial and temporal derivatives of the optical flow [12] to derive differential invariants: instantaneous rotation, divergence, and deformation. In this study, we used only rotation; this quantity is a linear feature transformation that has demonstrated good empirical performance for differentiating CM patterns [10, 13]. We extracted small patches from the regions predicted to contain cilia, and fed the corresponding rotation values at these patches coordinates into our second stacked deep network to build higher-order features for classification.

3.2. DenseNets for Cilia Segmentation

The first step in our end-to-end pipeline is the automated identification of regions in a video containing cilia. We motivate the use of densely-connected convolutional networks, or "DenseNets," to automatically generate semantic segmentations of a given input video.

DenseNets entail the use of fully-connected neural networks where each layer has direct access to the gradients and loss from the original input. While densely-connected layers add more parameters per layer, the overall number of parameters is reduced, as fewer feature maps are needed in each layer. This allows DenseNets to be very deep while remaining parameter-efficient. This particular architecture is highly amenable to image-related tasks, such as semantic segmentation, where the fully-connected layers can propagate high-resolution information regarding where one region "ends" and another "begins"; in our case, this is ideal for automatically differentiating regions containing cilia from those that do not, *regardless of whether the cilia are moving.* We implement a version of fully convolutional dense networks for segmentation similar to the Tiramisu architecture [14].

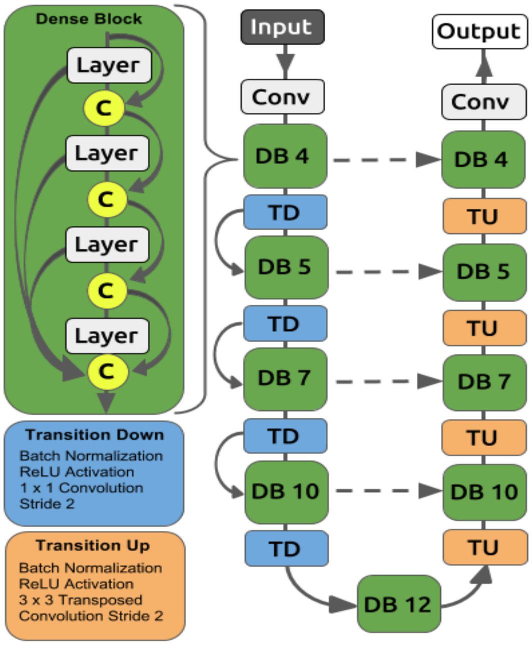

Fig. 1. Diagram of FC-DenseNet 74, so-named for having 74 total layers. Yellow circles represent concatenation operation and dashed lines represent skip connections.

The DenseNet architecture is composed of dense blocks (DB) in both the downsampling and upsampling paths (Fig. 1), with multiple layers stacked in each DB. Ultimately a bottleneck is reached, after which the number of layers decrease in each subsequent DB [15]. Skip connections connect DBs in the downsampling and upsampling paths to facilitate information flow from shallow layers so high-level features can be reused in deep layers. A single convolutional layer is added before the first block and after the last block. Ultimately, the network builds a sophisticated set of feature maps that, given a ground-truth segmentation map, will predict masks for semantic regions of new, unobserved input image.

3.3. Convolutional LSTMs for CM Classification

The second step in our end-to-end pipeline is modeling the CM as time series. While previous work has demonstrated some promise using Markov chains and autoregressive models [10], and even using simple 3D convolutional networks to capture three time points simultaneously in a single deep network [16], we propose to use recurrent neural networks (RNNs) to leverage much longer-term temporal dependencies in the data, theoretically enabling much deeper and richer representations of CM.

We use a variant of RNNs called "long short-term memory" (LSTM) networks, which use a series of gates to intelligently determine what portions of input sequences should be "remembered" and which should be "forgotten." These gates include the *forget* (Eq. 1), *input* (Eq. 2), and *output* (Eq. 3)

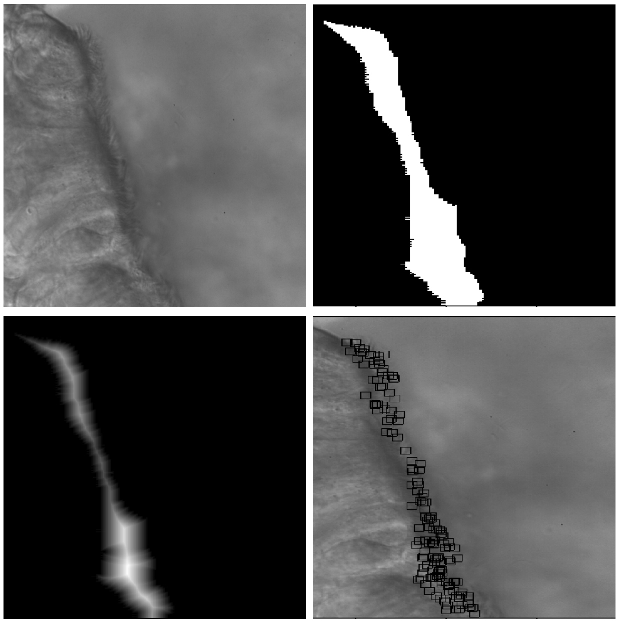

Fig. 2. Inputs and outputs of the pipeline. Input videos (upper left) are segmented, producing a segmentation mask (upper right). From this mask, a probability map is computed (bottom left) to sample patches (bottom right) that are then fed to the convolutional LSTM for analysis and classification.

gates, which determine whether a new input should be incorporated into the neuron's existing state (Eq. 5) or thrown away entirely; this enables the neuron to learn long-term dependencies. For sequences of images, such as videos of cilia biopsies, convolutional LSTM networks are ideal. Convolutional LSTMs [17] are similar to standard LSTM networks, except the inputs to each gate inside an LSTM neuron are convolved through kernel filters to extract spatial features, exactly like in a standard convolutional layer. These convolutional gates also preserve the temporal information inside the LSTM. See eqs. (1) to (5) for the convolutional LSTM transformations at each gate ($\circ$ denotes entry-wise product and $*$ denotes convolution).

$$f_t = \sigma(W_f * x_t + U_f * h_{t-1} + V_f \circ c_{t-1} + b_f) \tag{1}$$

$$i_t = \sigma(W_i * x_t + U_i * h_{t-1} + V_i \circ c_{t-1} + b_i) \tag{2}$$

$$o_t = \sigma(W_o * x_t + U_o * h_{t-1} + V_o \circ c_{t-1} + b_o) \tag{3}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma(W_c * x_t + U_c * h_{t-1} + b_c) \tag{4}$$

$$h_t = o_t \circ \sigma_h(c_t) \tag{5}$$

4. RESULTS

For cilia segmentation, we trained DenseNets with different numbers of layers to study the optimal tradeoff between speed and accuracy. We trained with the Adam optimizer to minimize categorical cross-entropy loss and vary regularization parameters tunings: dropout, l^2 weight decay, and learning rate annealing. Each model was trained for 100 epochs with a batch size of 4 on 2× Titan X GPU cards. All networks were implemented in Tensorflow with Keras. We evaluated resulting models on overall pixel classification accuracy, using the class with the highest probability for each pixel as the predicted pixel class.

Model	Parameters	Dropout	Decay	Accuracy
U-Net [18]	30 M	0.3	0.001	76.9%
FC-DenseNet 55	2.4 M	0.5	$1e^{-3}$	81.1%
FC-DenseNet 79	5 M	0.3	$1e^{-4}$	84.7%
FC-DenseNet 109	9.4 M	0.1	None	86.2%

Table 1. Segmentation performances of models using densely connected layers. Parameters are represented in millions.

We found that DenseNets with 74 layers and 2.4 million parameters struck a good balance between accuracy and training time. The model produced quality segmentation maps of cilia, with a weighted Dice coefficient of 0.437 for cilia mask predictions (Table 1), beating out the seminal U-Net architecture [18] in accuracy, and with an order of magnitude fewer parameters.

From the segmentation masks of the first stack of deep networks (Fig. 2), we extract small patches. We observed from the segmentation results that the cilia were most likely to be found in the middle of the predicted masks. Therefore, we computed a distance map from the mask (Fig. 2, bottom left) and used this as a sampling distribution. We sampled from the pixels within this mask without replacement until a saturation threshold was reached (proportional to the ratio of the area of the mask to the size of the patches). We then used the coordinates of the sampled pixels as the centers of 11×11 patches, and extracted 250 frames for each patch from the rotation data.

We collected a total of 24,577 patches from 75 patients. The label of each patch (normal or abnormal) was inherited from the patient. We performed cross-validation with three splits of 7,898, 8,519, and 8,160 patches (patches from the same patient were retained within the same fold to prevent testing contamination). Random horizontal and vertical flip augmentations of patches were performed during training. We used a convolutional LSTM with binary softmax classifier. We trained for 200 epochs with early stopping, using binary cross-entropy as the loss function.

For each patch in the validation set, the output probabilities from the final softmax layer was rounded based on a threshold of 0.5. All patches from the same video were averaged and rounded to 1 if above 0.5, and 0 if below 0.5. The resulting predicted classes for each video voted with a simple majority to determine the overall label of the patient. The classifier had an overall validation accuracy of 88% and with an F1 score of 0.8965 (Table 2). It performed extremely well on cilia depicting abnormal CM (Table 3).

Epochs	Accuracy	F1	Recall	Precision
100	0.81	0.84	0.93	0.77
200	0.88	0.90	0.98	0.83

Table 2. Convolutional LSTM classification performance as a function of training epochs.

	Normal (predicted)	Abnormal (predicted)
Normal (actual)	27	8
Abnormal (actual)	1	39

Table 3. Patient classification confusion matrix.

5. DISCUSSION

The convolutional LSTM classifies abnormal patients nearly perfectly; it struggles somewhat more on patients with normal CM. We found a similar classifier pathology in [10], despite using a distinct classification pipeline, suggesting the "distributions" of normal and abnormal CM overlap in that abnormal is a highly restricted subset of normal. Also, we observed mask predictions with in videos with unambiguous cell bodies had much better classification results. Upon inspection, we found the masks in these cases were of much higher quality and more likely to contain cilia, and therefore the extracted patches as well, directly impacting classification.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrate the efficacy of a pipeline of stacked deep nets for fully-automated, end-to-end analysis of CM. While achieving a high level of accuracy, future work will entail deeper training of the segmentation model to incorporate temporal information when determining the masks. Additionally, given the high variability in normal CM, we eventually aim to conduct fully unsupervised CM analysis using unmixing to determine motion subtypes and their pathological implications.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

8. REFERENCES

[1] O'Callaghan C *et al*, "Diagnosing primary ciliary dyskinesia," *Thorax*, vol. 62, no. 8, pp. 656–657, 2007.

[2] Garrod AS *et al*, "Airway ciliary dysfunction and sinopulmonary symptoms in patients with congenital heart disease," *Annals of the American Thoracic Society*, vol. 11, no. 9, pp. 1426–1432, 2014.

[3] Nakhleh N *et al*, "High prevalence of respiratory ciliary dysfunction in congenital heart disease patients with heterotaxy," *Circulation*, vol. 125, no. 18, pp. 2232–2242, 2012.

[4] Harden B *et al*, "Increased postoperative respiratory complications in heterotaxy congenital heart disease patients with respiratory ciliary dysfunction," *The Journal of Thoracic and Cardiovascular Surgery*, vol. Available online 22 July 2013, 2013.

[5] Stannard WA *et al*, "Diagnostic testing of patients suspected of primary ciliary dyskinesia," *American Journal of Respiratory and Critical Care Medicine*, vol. 181, no. 4, pp. 307–314, 2010.

[6] Olm MAK *et al*, "Primary ciliary dyskinesia: evaluation using cilia beat frequency assessment via spectral analysis of digital microscopy images," *Journal of Applied Physiology*, vol. 111, no. 1, pp. 295–302, 2011.

[7] Mantovani G *et al*, "Automated software for analysis of ciliary beat frequency and metachronal wave orientation in primary ciliary dyskinesia," *European Archives of Oto-Rhino-Laryngology*, vol. 267, no. 6, pp. 897–902, 2010.

[8] O'Callaghan C *et al*, "Analysis of ependymal ciliary beat pattern and beat frequency using high speed imaging: comparison with the photomultiplier and photodiode methods," *Cilia*, vol. 1, no. 1, pp. 8, 2012.

[9] Raidt J *et al*, "Ciliary beat pattern and frequency in genetic variants of primary ciliary dyskinesia," *European Respiratory Journal*, pp. erj00520–2014, 2014.

[10] Quinn SP *et al*, "Automated identification of abnormal respiratory ciliary motion in nasal biopsies," *Science Translational Medicine*, vol. 7, no. 299, pp. 299ra124–299ra124, 2015.

[11] Margaret W Leigh and Michael R Knowles, "Assessment of ciliary beat pattern: Variability in healthy control subjects has implications for use as test for primary ciliary dyskinesia," *CHEST Journal*, vol. 151, no. 5, pp. 958–959, 2017.

[12] Sun D *et al*, "Secrets of optical flow estimation and their principles," in *CVPR*, 2010, pp. 2432–2439.

[13] Quinn SP *et al*, "Novel use of differential image velocity invariants to categorize ciliary motion defects," in *Biomedical Sciences and Engineering Conference (BSEC)*. 2011, pp. 1–4, IEEE.

[14] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1175–1183.

[15] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[16] Charles Lu and Shannon Quinn, "Classification of ciliary motion with 3d convolutional neural networks," in *Proceedings of the SouthEast Conference*. ACM, 2017, pp. 235–238.

[17] Shi Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.