

Classification of Ciliary Motion with 3D Convolutional Neural Networks

Charles Lu
University of Georgia
Athens, Georgia
cjl76208@uga.edu

Shannon Quinn
University of Georgia
Athens, Georgia
squinn@cs.uga.edu

ABSTRACT

Cilia are hair-like structures found on nearly every cell in the body. Motile cilia in the nasal and bronchial passages work to propel mucus and foreign matter from the respiratory system. Abnormal ciliary motion results in a wide spectrum of disorders. Currently, defective ciliary motion is diagnosed by visual examination of nasal biopsies for signs of dyskinesia. However, these examinations are highly subjective, lacking a “common language” of standardized quantitative metrics to describe the observed motion.

Previous work relied on autoregressive representations of optical flow features that were fed into a support vector machine classifier. However, this approach was limited by two critical constraints: the strong parametric assumptions imposed by autoregressive models on the underlying structure of the data, and the requirement for experts to manually select regions of interest to be analyzed in the videos.

Deep learning suffers from neither of these limitations. Feature maps are automatically generated from relevant portions of videos. Furthermore, the feed-forward architecture of convolutional neural networks do not impose the strong parametric assumptions of autoregressive models, permitting complex behaviors to be captured. Here, we demonstrate initial results in applying 3D deep convolutional networks to binary classification of ciliary motion. While our preliminary results lag behind those of previous work, there is considerable promise in deep learning for ciliary motion analysis and room for further experimental design.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → *Computational biology*; Imaging;

KEYWORDS

deep learning, convolutional neural networks, cilia, ciliary motion, dynamic textures

ACM Reference format:

Charles Lu and Shannon Quinn. 2017. Classification of Ciliary Motion with 3D Convolutional Neural Networks. In *Proceedings of ACM SE '17, Kennesaw, GA, USA, April 13-15, 2017*, 4 pages.

DOI: <http://dx.doi.org/10.1145/3077286.3077303>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM SE '17, Kennesaw, GA, USA

© 2017 ACM. 978-1-4503-5024-2/17/04...\$15.00

DOI: <http://dx.doi.org/10.1145/3077286.3077303>

1 INTRODUCTION

Cilia are microtubule-based hair-like projections of cells; in humans, they are found on nearly every cell in the body. Disorders known as ciliopathies, where cilia function is disrupted, can result in a wide spectrum of diseases. In primary ciliary dyskinesia (PCD), airway cilia that normally beat in synchrony to mediate mucus clearance can exhibit dyskinetic motion or become immotile, resulting in severe sinopulmonary disease [9]. Cilia defects have been associated with increased respiratory complications and poor post-surgical outcomes [8]. The characterization of cilia abnormalities and specifics of the ciliary motion itself plays a critical role in the identification of specific ciliopathies and the determination of prophylactic respiratory therapies that could improve long-term outcomes in patients. Therefore, an accurate identification of specific ciliary motion patterns is clinically compelling.

An objective, computational method for quantitative assessment of ciliary motion was proposed by Quinn in 2015 [12]. In his work, ciliary motion was considered as an instance of dynamic texture [13]. Dynamic textures are modeled as rhythmic motions of particles subjected to stochastic noise [2]. Examples of dynamic textures include familiar motion patterns such as flickering flames, rippling water, and grass in the wind; each containing a small amount of stochastic behavior altering an otherwise regular visual pattern. Ciliary motion is well described as a dynamic texture. This representation proved robust for normal/abnormal binary ciliary motion classification, achieving an accuracy of over 90% relative to expert ground-truth labeling.

While this study was instrumental in establishing the efficacy of dynamic textures as quantitative representations for ciliary motion, it had some critical limitations, such as the use of autoregressive (AR) models [4] to model ciliary motion. While AR models are the preferred representation of dynamic textures, their primary usage is in distinguishing between *different* types of dynamic textures, rather than between different instances of the *same* dynamic texture. Furthermore, AR models impose strong parametric assumptions (multivariate Gaussian generating distributions that live on a linear manifold) on the underlying structure of the data that severely restrict the representative power of the model, truncating their sensitivity to minor yet important variations in motion dynamics and rendering them incapable of capturing complex, nonlinear interactions.

Also, despite considerable effort in automating much of the quantitative motion analysis, the previous work still ultimately relied on expert clinician collaborators to manually select and extract regions of interest (ROI) from the raw video data in order to focus the analysis of the computational pipeline to portions of videos containing cilia. These ROIs potentially introduce the subjective

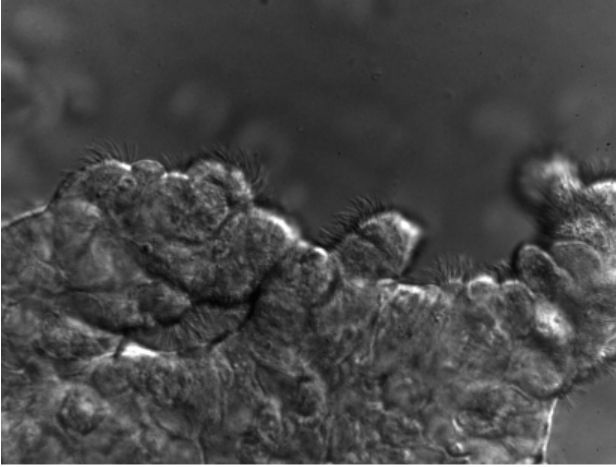


Figure 1: Example frame of a ciliary video labeled as normal; the regions containing cilia are evident along the cell boundaries.

bias of individual preference into the resulting dataset that was used by the AR model.

Here, we examine the application of deep learning, specifically *convolutional neural networks* (CNNs) to perform binary classification on ciliary motion. CNNs do not impose the strong parametric assumptions on the underlying data and can instead capture the complex dynamics and interdependencies through a series of convolutional filters. Furthermore, CNNs do not require manual ROIs to be segmented out and can automatically learn which areas of the videos are relevant to building feature maps. 3D CNNs provide a scalable and automatic solution to ciliary motion classification while still managing to incorporate information about the temporal structure in the video data of cilia biopsies.

2 METHODS

The objective of this experiment was to establish a baseline accuracy with a deep learning approach to cilia classification. To take advantage of the temporal dimension of video, an artificial neural network with 3D convolution filters was implemented to classify abnormality in ciliary motion. Similar computer vision techniques from previous studies were used to preprocess the raw video data [11]; specifically, the derivatives of optical flow were computed from the cilia videos and differential flow invariants (rotation) were extracted and fed into the 3D CNN in batches of frame arrays.

The dataset comprised of grayscale videos from nasal brush biopsies on 149 patients and examined by a blinded panel of clinicians. These videos were recorded at 200 frames per second for several seconds and taken at 640×480 resolution (Fig. 1). The panel of experts assigned labels to each patient to indicate the degree of abnormal cilia in the videos, where 1 constituted completely normal motion, 4 constituted highly abnormal motion, and 2 and 3 expressed gradations of abnormality.

Multiple videos were collected per patient, and each video was given the label of its patient's label. Some video samples were discarded due to extraneous camera movement, poor lighting, or

other recording artifacts. Videos from patients with labels of 2 or 3 were removed from the training set to leave more canonical examples of normal and abnormal motion so that the network could learn better feature maps from more distinctive motion types. Out of the total 325 videos, 266 (81.85%) canonical videos were used to train the network. From this subset, 127 videos (47.77%) were labeled as examples of abnormal ciliary motion, while the rest of the videos were labeled as examples of normal motion.

Optical flow is used in computer vision to predict the probable motion of pixels between consecutive frames and represents the movements of a pixel as a displacement vector field. Taking linear combinations of the derivatives of optical flow yields differential image invariants, also known as the measures: rotation, divergence, and deformation. These quantities are orientation-invariant which allow ciliary motion to be analyzed irrespective of camera orientation, lighting differences, and even the direction in which the cilia beats [12]. The computed curl (rotation) of ciliary motion was used for training since it was the only scalar valued component (divergence and deformation being vector quantities).

The videos were standardized to the absolute maximum value in order to preserve the most information of pixel variance when converting from floating point to 16 bit integer data type. Some videos had extremely high maximum values so scalars for each video were individually computed to prevent pixel overflow while preserving the maximum range of the integer data type. Each video had 250 frames that were reduced from 640×480 to 80×60 pixels and split into blocks of 10 frames, effectively producing 25 arrays of $80 \times 60 \times 10$ inputs per video. These blocks of arrays were fed in batches of 10 into a 3D CNN which was implemented in TensorFlow and trained on a p2 instance with a NVIDIA K80 GPU from Amazon Web Services (AWS).

3 RESULTS

The 3D CNN architecture had 2 stacked layers, each composed of a 3D convolutional layer, a 3D max pooling layer, and a layer of Rectified-Linear Unit [7] activation functions. These two stacked layers fed into a fully connected dense layer that outputted to a binary softmax classifier. Most CNN “best” practices [3] were adapted for three dimensions. The first layer had 32 3D convolutional filters and the second layer had 64 filters. The kernel size of each filter was $3 \times 3 \times 3$ with zero padding used and dropout [15] added for regularization. The fully connected layer had 57600 weights ($80 \times 60 \times 10 = 48,000$ plus 9,600 extra weights from the zero padding). A visualization of the network architecture is shown in Fig. 2.

An adaptive learning rate was used with the ADAM [6] optimizer. A low batch size of 10 was selected to accommodate the memory capacities of our hardware resources. After 10 iterations, the model converged to a local minimum. We tested classification accuracy with a 66-33% holdout train-test split.

The average performance of the 3D CNN on the test set was around 75%, which is below the 90% benchmark accuracy from previous work [12] (Fig. 3). Using larger batch sizes and pixel dimensions slightly improved performance but at the cost of making the model much more computationally expensive and exhausting the memory capabilities of our machine. Trading dimension size

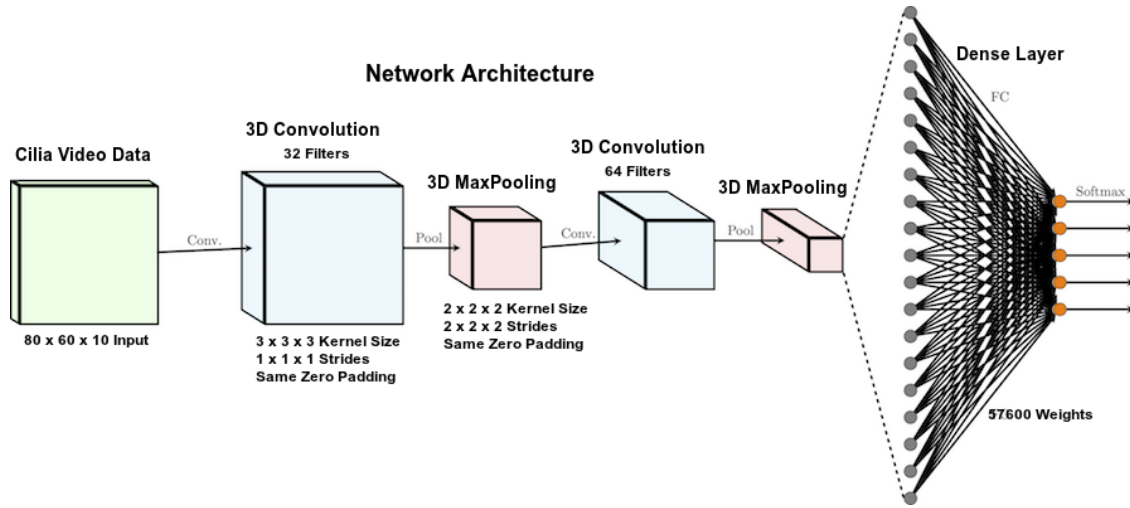


Figure 2: Overview of the 3D CNN architecture as implemented in TensorFlow.

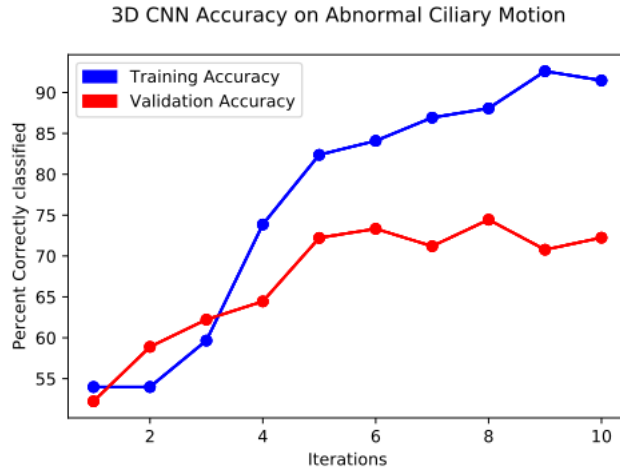


Figure 3: Training and testing accuracy of the 3D CNN as a function of network epochs.

of each frame for a larger block size had little effect in terms of accuracy but also greatly increased computational requirements.

4 DISCUSSION

Clearly, the model is severely overfitting the current dataset. Several regularization techniques could be implemented to combat this: transfer learning from pre-trained networks, data augmentation, and batch normalization. However, each pose their own difficulties. For transfer learning [10], most pre-trained network weights are currently designed for image rather than video data. While data augmentation is feasible, transformations must be done congruently within batches to maintain the temporal structure and would require more processing power for videos than images when done in memory. Also, the underlying structure of the processed rotation data might not allow for the same transformations as the original

grayscale videos. Batch Normalization [5] would be straightforward enough to implement but seems unlikely to improve current classification accuracy to the state of the art benchmark. Insight from already experimenting with the dropout regularization parameter suggests that the learning capacity of the 3D CNN is too complex to extract good feature maps from the down-sampled data. We believe a better solution would be to provide the network with more labeled training examples and enough computing resources to allow for larger input size dimensions.

A limitation of the current dataset is each video takes the label at a patient level instead of the video level itself, which might not contain any abnormal motion despite the patient being labeled as abnormal or vice versa. Also, a training example might contain multiple cells with some cells showing normal motion while others showing abnormal motion in the same video which complicates training. We attempted to mitigate this effect of misleading examples by following Bengio's [1] principle of curriculum learning, opting to train only on patients labeled 1 and 4 (trading quantity of training examples for more canonical and distinct motion examples).

Another limitation was the drastic reduction in cilia video dimensions. The processed videos had to be reduced to $\frac{1}{8}$ their original size to remain within the bounds of our server's memory limits. This downsampling greatly diminished the network's capacity to extract relevant features and to discriminate between normal and abnormal motion. With more computing resources and a more precisely labeled dataset, we expect increased classification accuracy with the 3D CNN model. With additional preprocessing, the deformation vector data could also be incorporated with the rotation data to provide even more contextual information for network to learn better representations about the dynamics of ciliary motion.

5 CONCLUSION

Although deep learning methods usually require an abundant amount of labeled training data, one possible way to circumvent this issue

would be to train another CNN to automatically generate a heatmap of candidate areas of videos where cilia are likely to be present, similar to the deep learning pipeline [16] that won the Camelyon 17 challenge for breast cancer classification. These candidate areas would then be fed into another classifier such as a Gradient Boosted Decision Tree or even another CNN to predict the probability of abnormal cilia in that region.

Other approaches that could leverage deep learning for cilia classification are Recurrent Neural Networks (RNNs) such as the Long-Short Term Memory (LSTM) network, which inherently incorporates the dimension of time to build feature maps. It is even possible to stack CNNs on top of RNNs to leverage aspects of both network architecture types. Other researchers [14] have already applied a similar approach with excellent results on the localization of protein structures. Another interesting area of deep learning that also presents a promising approach to solving the broader problems of cilia segmentation and identification of different cilia motion phenotypes (both normal and abnormal) include the attention-seeking mechanisms of RNNs and General Adversarial Networks (GAN). Attention mechanisms and GANs could be used to segment out cilia and to generate a canonical phenotype of a specific ciliary motion, respectively.

6 ACKNOWLEDGEMENTS

The authors would like to acknowledge the University of Pittsburgh and Children's Hospital of the University of Pittsburgh Medical Center for initial acquisition of video microscopy data. In particular, we would like to acknowledge and thank Dr. Chakra Chennubhotla, Dr. Maliha Zahid, and Dr. Cecilia Lo.

REFERENCES

- [1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. ACM, 41–48.
- [2] Jie Chen, Guoying Zhao, Mikko Salo, Esa Rahtu, and Matti Pietikainen. 2013. Automatic dynamic texture segmentation using local descriptors and optical flow. *IEEE Transactions on Image Processing* 22, 1 (2013), 326–339.
- [3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT Press.
- [4] Midori Hyndman, Allan D Jepson, and David J Fleet. 2007. Higher-order Autoregressive Models for Dynamic Textures.. In *BMVC*. 1–10.
- [5] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
- [6] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [7] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 807–814.
- [8] Nader Nakhleh, Richard Francis, Rachel A Giese, Xin Tian, You Li, Maimoona A Zariwala, Hisato Yagi, Omar Khalifa, Safina Kureshi, Bishwanath Chatterjee, and others. 2012. High prevalence of respiratory ciliary dysfunction in congenital heart disease patients with heterotaxy. *Circulation* (2012), CIRCULATIONAHA-111.
- [9] Christopher O'Callaghan, Mark Chilvers, Claire Hogg, Andrew Bush, and Jane Lucas. 2007. Diagnosing primary ciliary dyskinesia. (2007).
- [10] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.
- [11] Shannon Quinn, Richard Francis, Cecilia Lo, and Chakra Chennubhotla. 2011. Novel use of differential image velocity invariants to categorize ciliary motion defects. In *Biomedical Sciences and Engineering Conference (BSEC), 2011*. IEEE, 1–4.
- [12] Shannon P Quinn, Maliha J Zahid, John R Durkin, Richard J Francis, Cecilia W Lo, and S Chakra Chennubhotla. 2015. Automated identification of abnormal respiratory ciliary motion in nasal biopsies. *Science translational medicine* 7, 299 (2015), 299ra124–299ra124.
- [13] Payam Saisan, Gianfranco Doretto, Ying Nian Wu, and Stefano Soatto. 2001. Dynamic texture recognition. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 2. IEEE, II–II.
- [14] Søren Kaae Sønderby, Casper Kaae Sønderby, Henrik Nielsen, and Ole Winther. 2015. Convolutional LSTM networks for subcellular localization of proteins. In *International Conference on Algorithms for Computational Biology*. Springer, 68–80.
- [15] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.
- [16] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. 2016. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016).