Classification of Ciliary Motion with 3D Convolutional Neural Networks

Short Paper

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, therefore, 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 does 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 nets to binary ciliary motion classification. 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, Video Classification

1 INTRODUCTION

Cilia are microtubule-based hair-like projections of cells; in humans, they are found on nearly every cell of the body. Diseases 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 postsurgical outcomes [8]. The characterization of ciliary motion abnormalities and specifics of the ciliary motion properties itself plays a critical role in the identification of specific ciliopathy and the determination of prophylactic respiratory therapies that could improve long-term outcomes in patients. Therefore, accurate identification of specific ciliary motion patterns is clinically compelling.

An objective, computational method for quantitative assessment of ciliary motion was proposed in [Quinn 2015]. In this 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 with a small amount of stochastic behavior altering an otherwise regular visual pattern. ciliary motion is well described as a dynamic texture. In [Quinn 2015], the authors found this representation 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 was critically limited. First, the authors used autoregressive (AR) models [4] to model the dynamics of ciliary motion. While AR models are the preferred representation of dynamic textures, their primary use is in distinguishing between different types of dynamic textures, rather than between different instances of the same dynamic texture. Furthermore, the strong parametric assumptions imposed by the AR models on the underlying structure of the data – multivariate Gaussian generating distributions that live on a linear manifold–severely restrict the representative power of AR models, truncating their sensitivity to minor but important variations in motion dynamics and rendering them incapable of capturing complex, nonlinear interactions.

Second, despite the considerable effort in automating much of the quantitative motion analysis, the authors still ultimately relied on expert clinician collaborators to manually select and extract regions of interest (ROI) from the raw video data to focus the analysis of the computational pipeline. These ROIs potentially introduce the same subjective bias of individual preference and training into the resulting data.

Here, we examine the application of deep learning, specifically convolutional neural networks (CNNs) to perform binary ciliary motion classification. CNNs do not impose strong parametric assumptions on the underlying data, instead capturing complex dynamics and interdependencies through a series of convolutional filters. Furthermore, CNNs do not require manual ROI selection in order to function, and can instead learn for themselves what feature maps to focus on.

3D convolutional neural networks provide a more scalable solution to ciliary motion classification while still incorporating information about the temporal structure inherent in video data of these cilia biopsies.

2 METHODS

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

The dataset comprised of grayscale videos of nasal brush biopsies from 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 640x480 resolution. The panel of expert clinicians assigned labels to the patients to indicate degree of ciliary motion abnormality, where 1 constituted completely normal motion, 4 constituted highly abnormal motion, and 2 and 3 expressed gradations of abnormality.

	Patient	class
Abnormality		
1	82	82
2	14	14
3	13	13
4	40	40

Figure 1: A sample black and white graphic that has been resized with the `includegraphics` command.

Multiple videos per patient were collected and each video was given the abnormality label of its patient. Some video samples were removed due to extraneous camera movement, poor lighting, bad focus, or other recording artifacts. Videos of patients with labels of 2 and 3 were removed from the training set, hopefully leaving more canonical examples of normal and abnormal motion for the network to learn from. This non-canonical subset of videos comprised 18.15% of the total 325 videos, leaving 266 videos left for training the network. Out of the canonical data, 127 videos or 47.77% were labeled as examples of abnormal cilia motion, while the other half were labeled as normal.

Optical flow is used in computer vision to predict the probable motion of pixels between consecutive frames and represents movement of a pixel as a displacement vector field. Taking linear combinations of the derivatives of the optical flow yielded differential image invariants, colloquially known as measures of rotation, divergence, and deformation. These quantities are orientation-invariant, allowing ciliary motion to be analyzed irrespective of the camera orientation, lighting differences, or even the direction in which the cilia is beating[12]. The computed curl (rotation) of ciliary motion was used for training because it was the only component with scalar valued quantities.
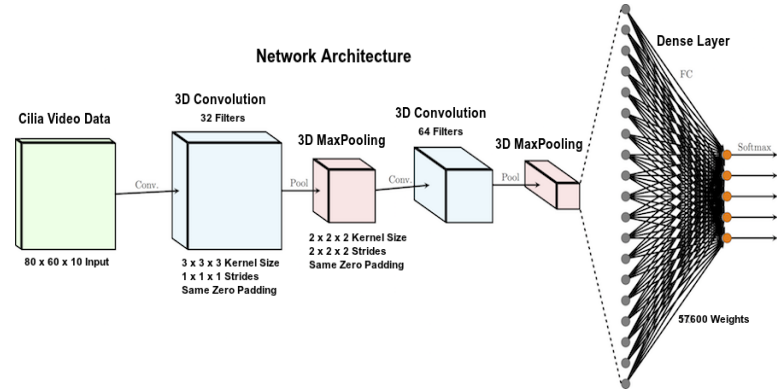
Figure 2: A sample black and white graphic that has been resized with the `includegraphics` command.

The videos were standardized to the absolute maximum value to preserve the most information when converting from floating point to 16 bit integer values. Some videos had extremely high maximum values so a scalar for each video was individually computed to prevent pixel value overflow and within range of data type. Each video had 250 frames, which were each reduced from 640x480 to 80x60 pixels and split into blocks of 10, effectively producing 25 arrays of 80 x 60 x 10 inputs per video. These blocks were fed in batches of 10 per iteration into a 3D CNN implemented in Tensorflow and trained with a p2 instance on Amazon Web Services (AWS).

3 RESULTS

The 3D CNN architecture had 2 stacked layers, each composed of a 3D convolution layer and a 3D max pooling layer with Rectified-Linear Unit [7] activation functions, and a fully connected layer outputting 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. The kernel size of each filter was 3 x 3 x 3 and zero padding used and dropout [15] added for regularization. The fully connected layer had 57600 weights due to 80 * 60 * 10 = 48,000 plus 9,600 extra weights from zero padding. The varying learning rates were tried with ADAM[6] as the optimizer. A low batch size of 10 was selected to accommodate the memory capacities of our hardware resources. After 10 iterations, the model appeared to find a local minimum. We tested classification accuracy with a 66-33% train-test split. The average performance of the 3D CNN on the test set was around 75%, still below the 90% the current state-of-the art benchmark. Using larger batch sizes and pixel dimensions slightly improved performance but made training much more computationally expensive and exhausted all memory capabilities of our machine. Trading dimension size of each frame for a larger block size had little difference in terms of accuracy but greatly increased computational memory requirements.
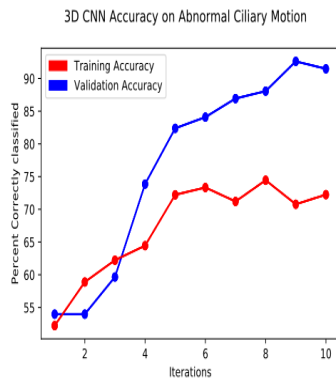
Figure 3: Overview of the 3D CNN structure.

We tested classification accuracy with a 66-33% train-test split. The average performance of the 3D CNN of the test set is around 75%, well below the 90% benchmark of the current state-of-the art.

4   DISCUSSION

Clearly, the model is severely overfitting to the dataset. Several regularization techniques could be implemented to combat this: adding batch normalization, using data augmentation, and fine-tuning the weights. However, each pose their own difficulties. Most pre-trained network weights are designed for image input instead of video data and transfer learning [10] might be too generalized to differentiate the subtleties of ciliary motion. Data augmentation is feasible but must be done congruently with batches which would require more memory to be done online for video data. Also, the underlying structure of the rotation data might not allow for some of the same transformations as the raw grayscale videos of ciliary motion. Batch Normalization [5] would be rather straightforward to implement but would be unlikely to improve accuracy enough to current benchmarks. From experimenting with the dropout parameter, the 3D CNN model?s learning capacity is too complex for its current input data. A better solution would be to provide the model with more labeled training examples and enough computing resources to allow for larger input dimensions.

A limitation of this dataset was that each video was labeled at a patient level instead of per video, which might not contain any abnormal motion at all despite the patient being labeled abnormal or vise versa. We attempted to mitigate training on difficult and misleading examples by following [1] idea of curriculum learning by opting to only train on patients labeled 1 and 4, and this decreased the number of training examples.

Another limitation for the model was the drastic reduction in the dimensions of cilia data. The processed videos had to be reduced to 1/8 of their original size to avoid running into memory issues. This downsampling greatly diminished the network capacity to extract the most relevant features in distinct ciliary motion. With more computing resources and a more precisely labeled dataset, we would expect to see an increase in 3D CNN classification accuracy. With additional preprocessing, the deformation vector data could also be incorporated in addition to the rotation data to provide more relevant feature information for network to learn.

5   CONCLUSION

Although deep learning methods usually requires an abundant amount of labeled training data, one possible way to circumvent this issue would be to train a CNN to generate a heatmap of patches of the video where Cilia are likely present, similar to the deep learning pipeline that won the Camelyon 17 challenge for Breast Cancer Classification[16]. Then feed this into another classifier such as a Gradient Boosted Decision Tree or another CNN to predict abnormality of that cilia patch.

Other approaches that can leverage deep learning for cilia classification are Recurrent Neural Networks (RNNs) such as a Long-Short Term Memory (LSTM) network to actually incorporate a times-series element instead of simulating time with a third dimension in CNNs to build feature maps. It is even possible to stack CNNs on top of RNNs to get the best of both worlds. Others in computational have already applied this approach with good results [14]. Other rapidly growing areas of deep learning might be used to solve the broader problem of cilia segmentation and identification of ciliary motion phenotypes include attention-seeking mechanisms of RNNs and Reinforcement Learning in segmentation and unsupervised deep learning techniques like General Adversarial Networks to generate a perfect canonical phenotype of ciliary motion.

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 OâĂŹcallaghan, 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).