Deep Convolutional Neural Network Models of the Retina

Lane McIntosh*[1], Niru Maheswaranathan*[1], Aran Nayebi[2], Surya Ganguli[3], Stephen Baccus[4]

[1]Neurosciences PhD Program, [2]Computer Science Department, [3]Department of Applied Physics, [4]Department of Neurobiology

SCIEN
The Stanford Center for Image Systems Engineering

Abstract

The retina consists of three layers of cells that form the first stage of visual processing, however current models of the retina poorly predict responses to natural scenes.

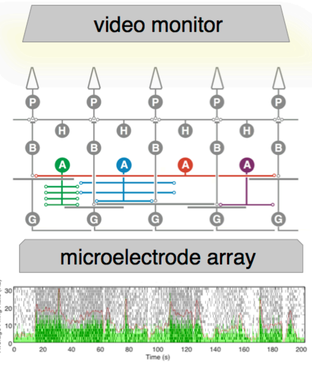
Deep convolutional neural networks demonstrate success at many image and pattern recognition tasks [1], but can these models capture the computations used in biological visual pathways when viewing natural movies?

We demonstrate that these deep learning models are considerably more accurate than pre-existing models and generalize significantly better across stimuli classes. We furthermore probe the model using structured stimuli to reveal nonlinear computations important for biological vision.

Methods

Retinal recordings

Models were trained on multielectrode array recordings of salamander retinal ganglion cells.



Natural scenes stimulus

The natural stimulus consisted of a sequence of jittered natural images, replaced every second, sampled from the Tkacik natural image database [2].

White noise stimulus

The spatiotemporal binary white noise stimulus consisted of 55 µm checkers resampled at 60 Hz.

Model training

Models were trained using Keras and Theano libraries [3] on NVIDIA Titan GPUs, using a loss function corresponding to the negative log-likelihood under Poisson spike generation,

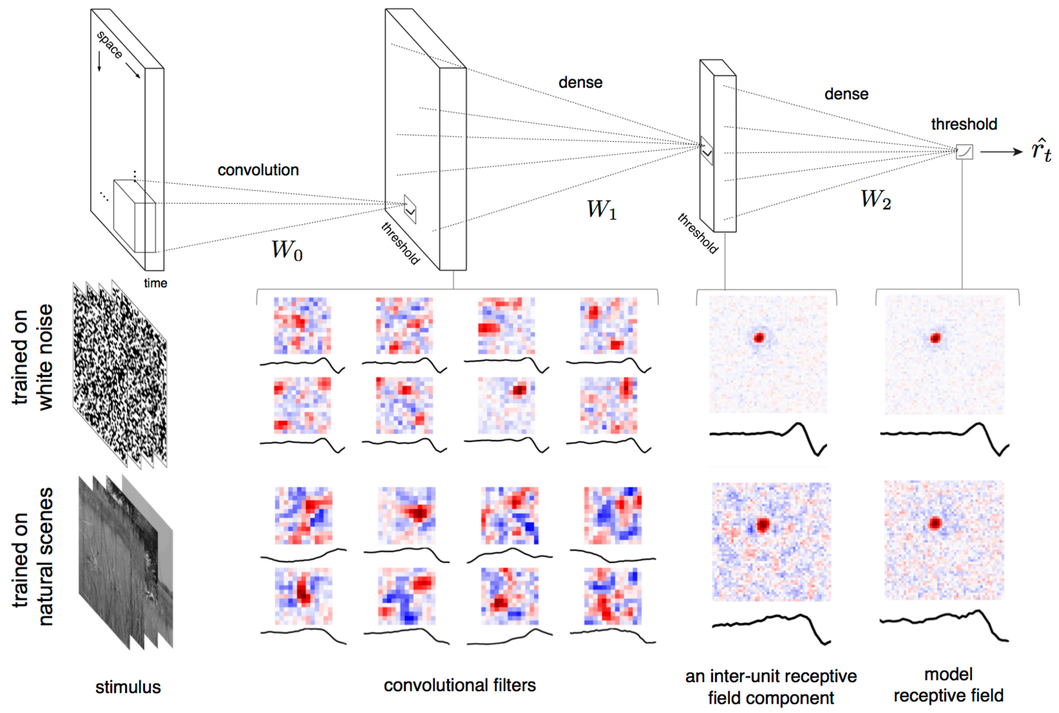
$$\text{Loss} = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t - y_t \log \hat{y}_t \text{, and}$$

Adam, a first-order gradient-based optimization algorithm [4]. We explored a variety of architectures for the convolutional network, varying the number of layers, number of filters per layer, the type of layer (convolutional or dense), and the size of the convolutional filters.

Benchmarks

We fit GLMs using stimulus filters, spike history filters, and cross-coupling terms amongst simultaneously recorded neurons. LN models were fit using spatiotemporal filters and a parameterized soft rectifying nonlinearity.

Architecture and learned features

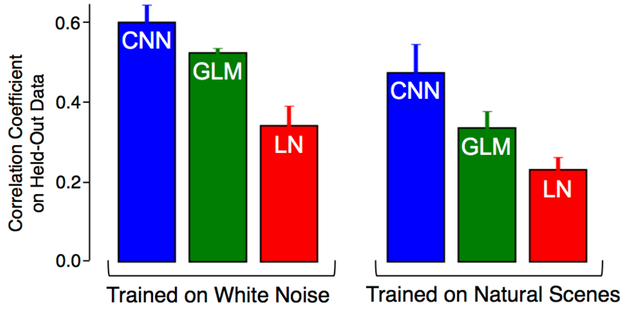


Convolutional filters learn features that have the same spatial bandwidth as bipolar cells; these are pooled to yield inter-units with overlapping center surround receptive fields. The overall receptive field of the model reproduces the spike-triggered average of the ganglion cell.

Performance

Within-class performance

Convolutional neural network models better predict retinal responses to both white noise (> 15% improvement) and natural scenes (> 40% improvement) stimuli as compared to GLM and LN models.



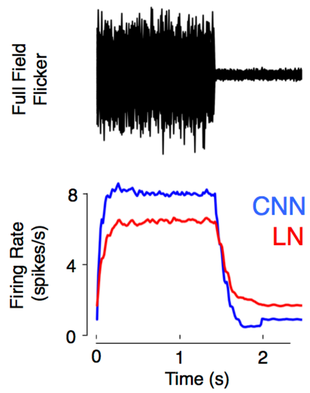
Generalization on different stimuli classes

When models were tested on the stimuli class different from its training stimuli, convolutional neural network models also generalized the best.

	Test White Noise	Test Natural Scenes
Train White Noise	CNN: 0.6 LN: 0.3	CNN: 0.4 LN: 0.2
Train Natural Scenes	CNN: 0.2 LN: 0.1	CNN: 0.5 LN: 0.2

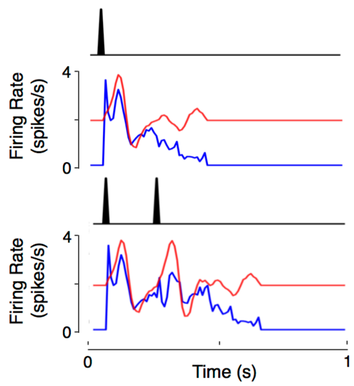
Responses to structured stimuli

Contrast adaptation



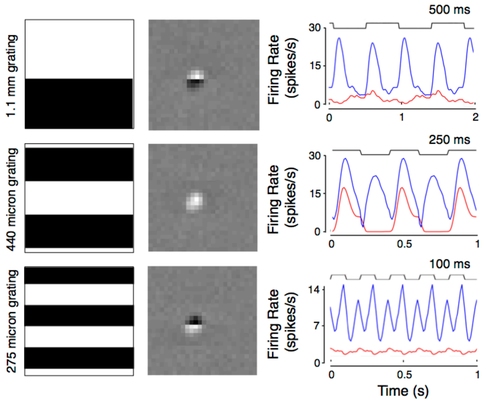
Averaged traces of LN and CNN responses to high-low contrast steps. The CNN overshoots its response to low contrast stimuli, capturing the increase in retinal sensitivity to low contrast stimuli after adaptation to high contrast stimuli.

Paired pulse response



LN and CNN responses to single and paired flashes. While the LN model responds with equal strength to the second pulse, the CNN has an attenuated response to the second flash.

Frequency doubling with reversing gratings



Reversing gratings probe the presence of nonlinear structure. In all three conditions, the CNN responds to the onset of a reversal, regardless of polarity. However, the LN model either fails to encode reversals robustly or only encodes alternate reversals.

Conclusions

Convolutional neural network models are substantially better at predicting retinal responses than previous models.

Convolutional neural network models trained on natural scenes recapitulate nonlinear retinal response properties to structured stimuli.

Visualizing the computations in these deep models trained on biological visual pathways can yield insights for ideal computer vision systems.

References

[1] LeCun et al. Deep learning. *Nature* 521, 436–444 (2015).
[2] Tkacik et al. Natural images from the birthplace of the human eye. *PLoS ONE* 6: e20409 (2011).
[3] Bastien et al. Theano: new features and speed improvements. *NIPS deep learning workshop* (2012).
[4] Kingma and Ba. Adam: A Method for Stochastic Optimization. *International Conference for Learning Representations* (2015).

Contact information

Lane McIntosh, lmcintosh@stanford.edu

PDF