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, Stanford University

Abstract

Much of our understanding of early sensory systems comes from studies using artificial stimuli, such as white noise and structured stimuli (e.g. bars, gratings, and flashes).

Stimulus → Sensory system → Response

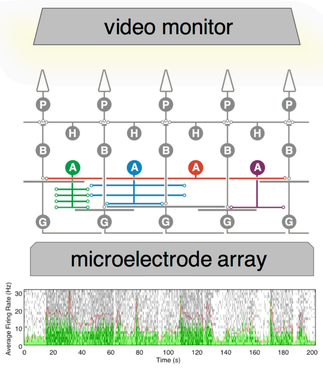
Previous models of the retinal response, such as the linear-nonlinear (LN) model, capture responses to these stimuli but fail to generalize to ethologically relevant, natural stimuli.

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

We demonstrate that deep neural network models are considerably more accurate than pre-existing models at modeling retinal responses to artificial and natural stimuli and generalize significantly better across stimulus types.

Furthermore, probing the models using structured stimuli reveals nonlinear computations important for biological vision.

Methods



Retinal recordings

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

Visual stimuli

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

The spatiotemporal binary white noise stimulus consisted of 55 µm checkers at 30 Hz.

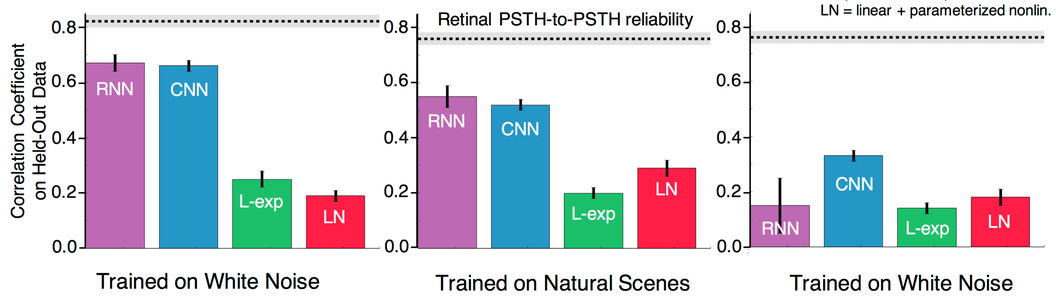
Models were never trained on structured stimuli.

Model training

Models were trained with a loss function corresponding to the negative log-likelihood under Poisson spike generation, $\text{Loss} = \frac{1}{T}\sum_{t=0}^{T} \hat{y}_t - y_t \log \hat{y}_t$. 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 fully connected), and the size of the convolutional filters.

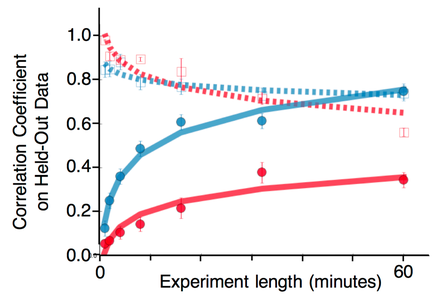
Performance

RNN = Recurrent neural network
CNN = Convolutional neural network
L-exp = Linear + exponential nonlin.
LN = linear + parameterized nonlin.

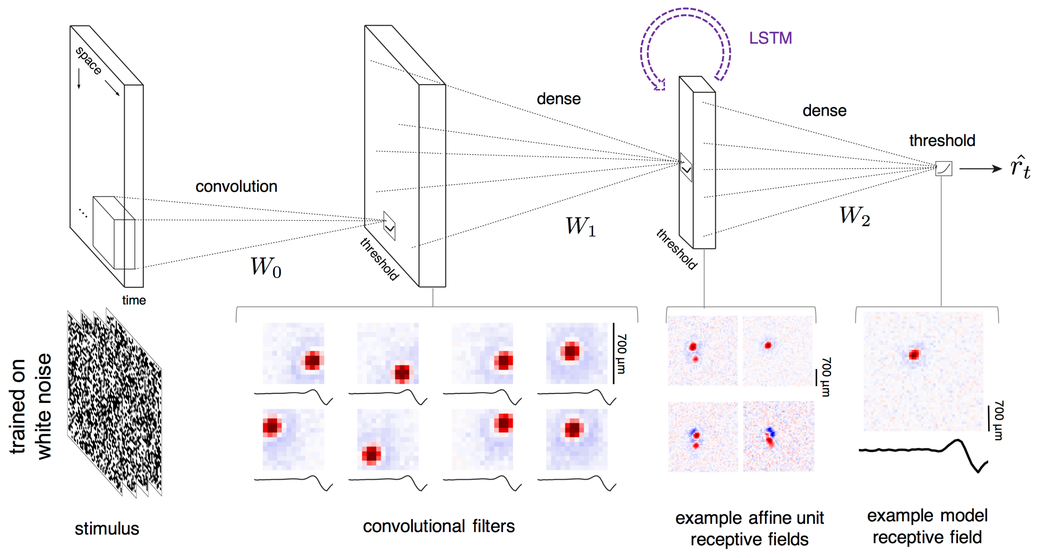


LN models are more susceptible to overfitting than CNNs

(At right) We find that the CNN model trained using just 8 minutes of data has better held out performance than an LN model fit using a 60 minute recording.



Architecture and learned features



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

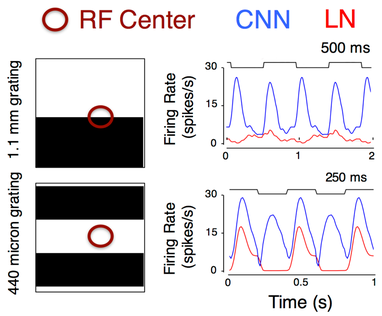
Acknowledgements

LM: NSF, NVIDIA Titan X Award, NM: NSF, AN and SB: NEI grants, SG: Burroughs Wellcome, Sloan, McKnight, Simons, James S. McDonnell Foundations and the ONR

Structured stimuli responses

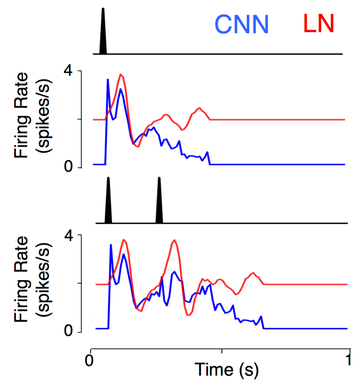
Frequency doubling with reversing gratings

CNNs respond to both phases of the reversing grating, unlike LN models, indicative of nonlinear spatial integration.

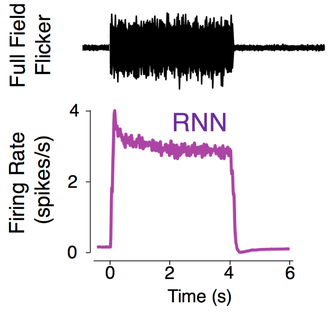


Paired flash response

CNNs exhibit luminance adaptation, by suppressing the response to a paired flash.



RNNs trained on natural scene stimuli exhibit contrast adaptation



(At left) Average RNN response to changes in contrast during full-field flicker sequences show overshoot and undershoot during the low-to-high and high-to-low contrast transitions, respectively. LN and CNN models do not capture this phenomenon.

Conclusions

- Convolutional and recurrent neural networks are substantially better at predicting retinal responses than LN models on white noise and natural scene stimuli.
- These networks trained on natural scenes recapitulate nonlinear retinal response properties to held-out structured stimuli.
- Distilling computational insights from deep networks trained to model responses of biological sensory systems can unlock the mystery of how these systems encode natural stimuli.

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).

Contact information

Lane McIntosh, lmcintosh@stanford.edu