

Convolutional Neural Network Models of the Retina

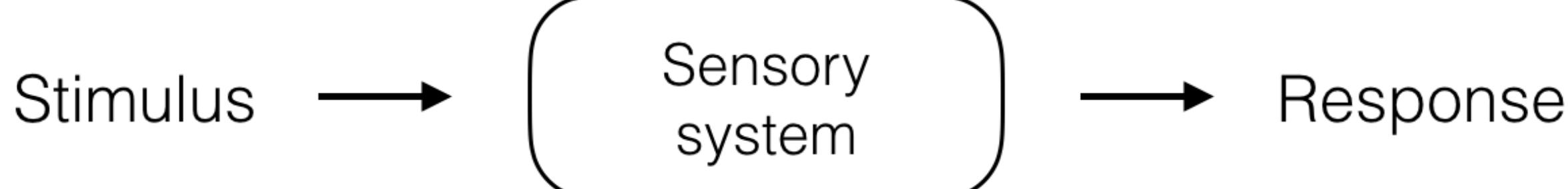
Lane McIntosh^{*1}, Niru Maheswaranathan^{*1}, Aran Nayebi², Surya Ganguli³, Stephen Baccus⁴

¹Neurosciences PhD Program, ²Computer Science Department, ³Department of Applied Physics, ⁴Department 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).



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?

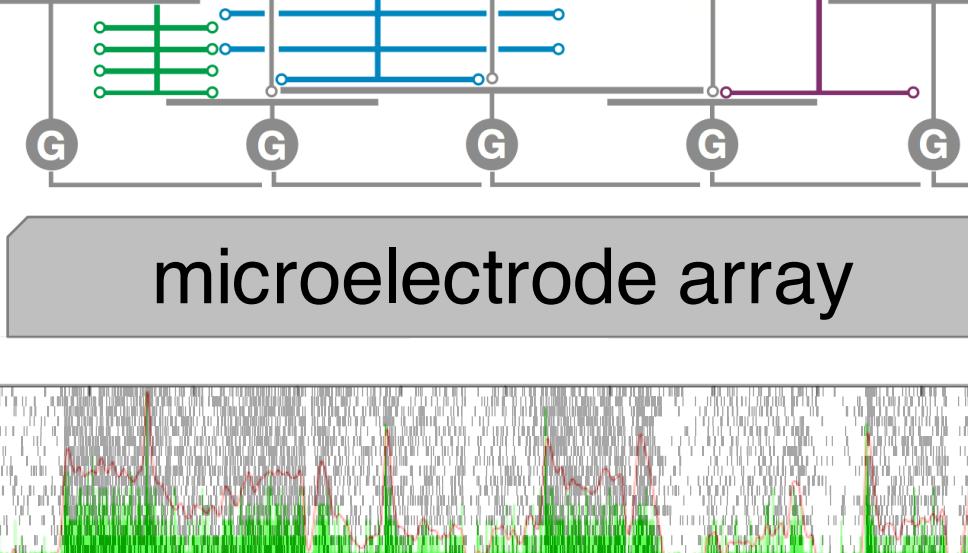
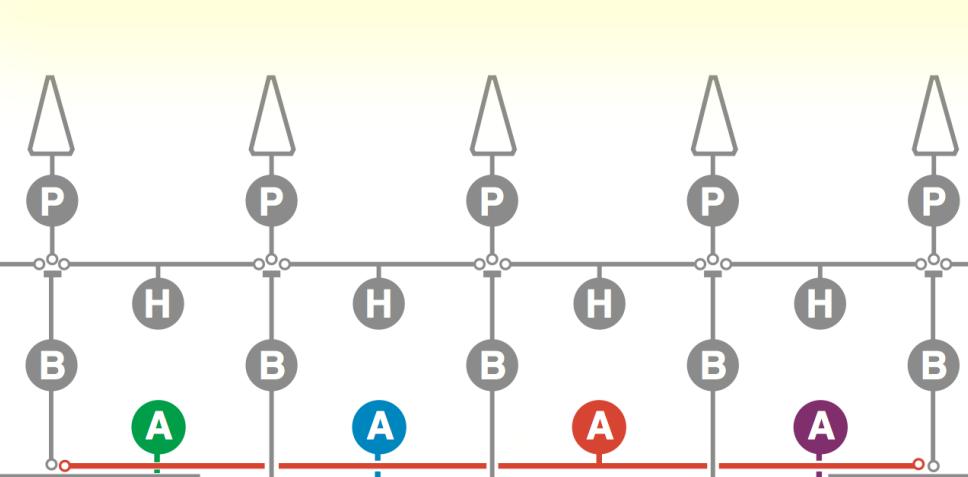
We demonstrate that deep neural network models are considerably more accurate than pre-existing models at modeling retinal responses to artificial and natural stimuli and generalize significantly better across stimulus types.

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

Methods



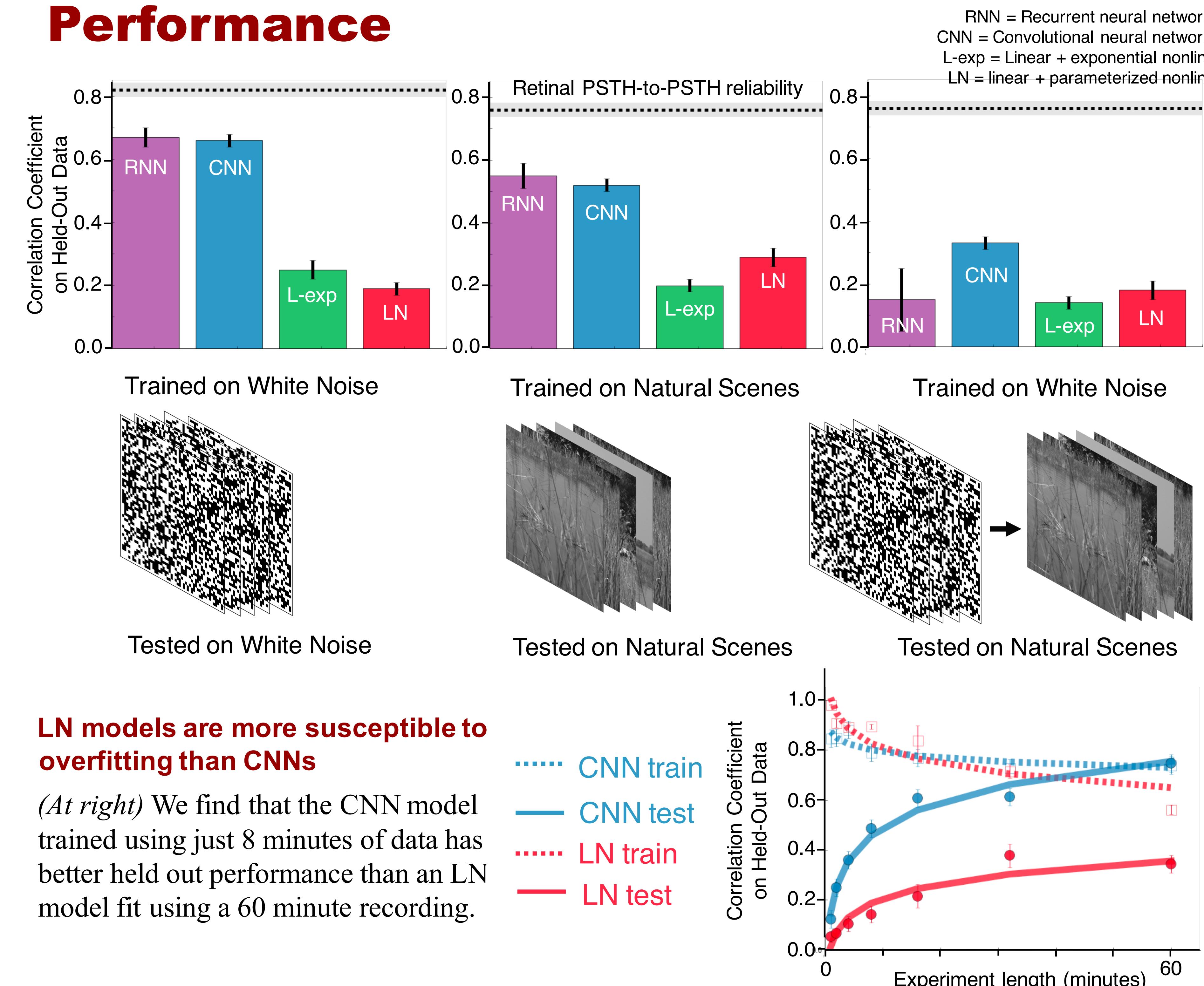
video monitor



under Poisson spike generation, 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.

Acknowledgements

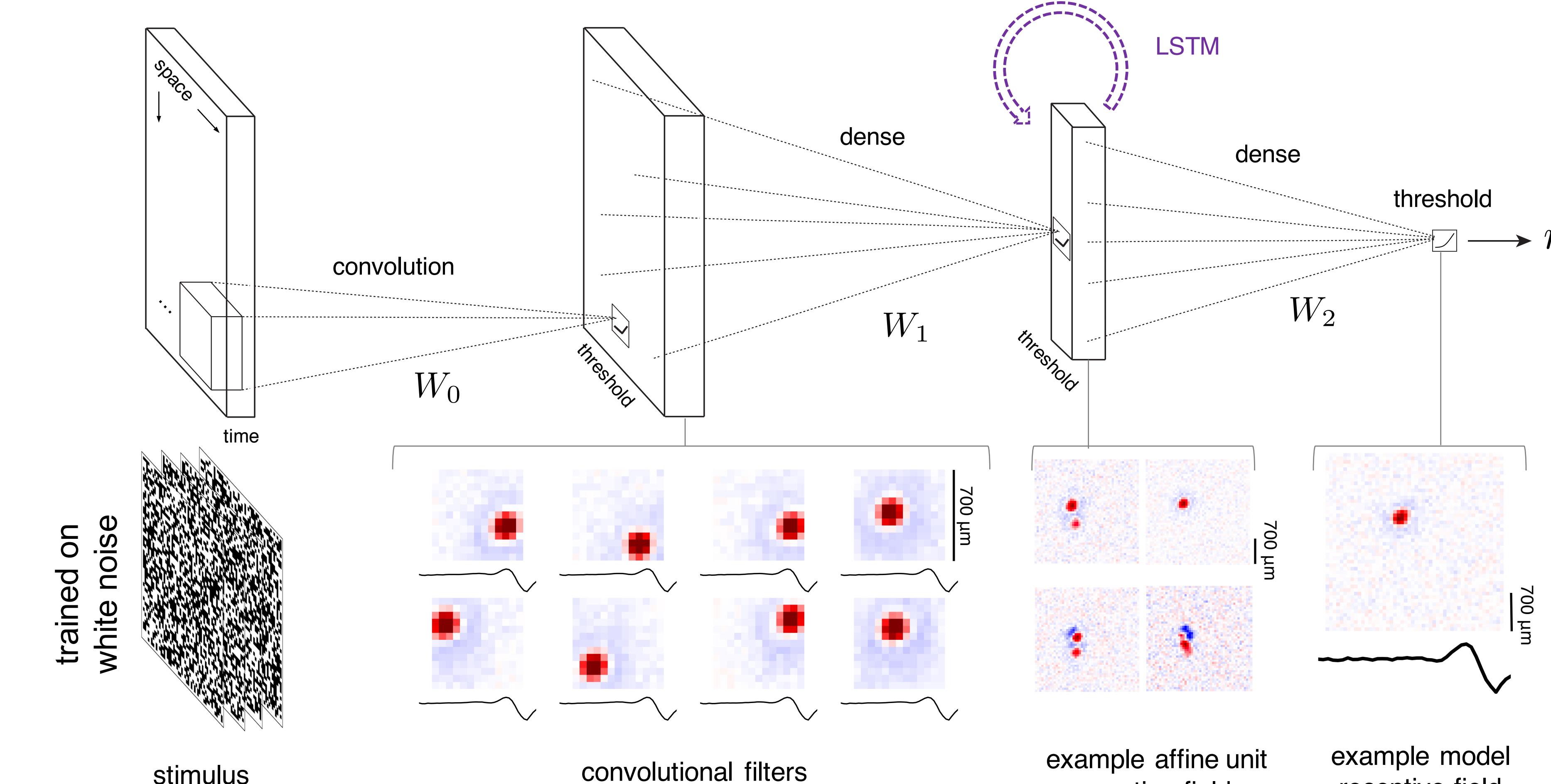
Performance



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

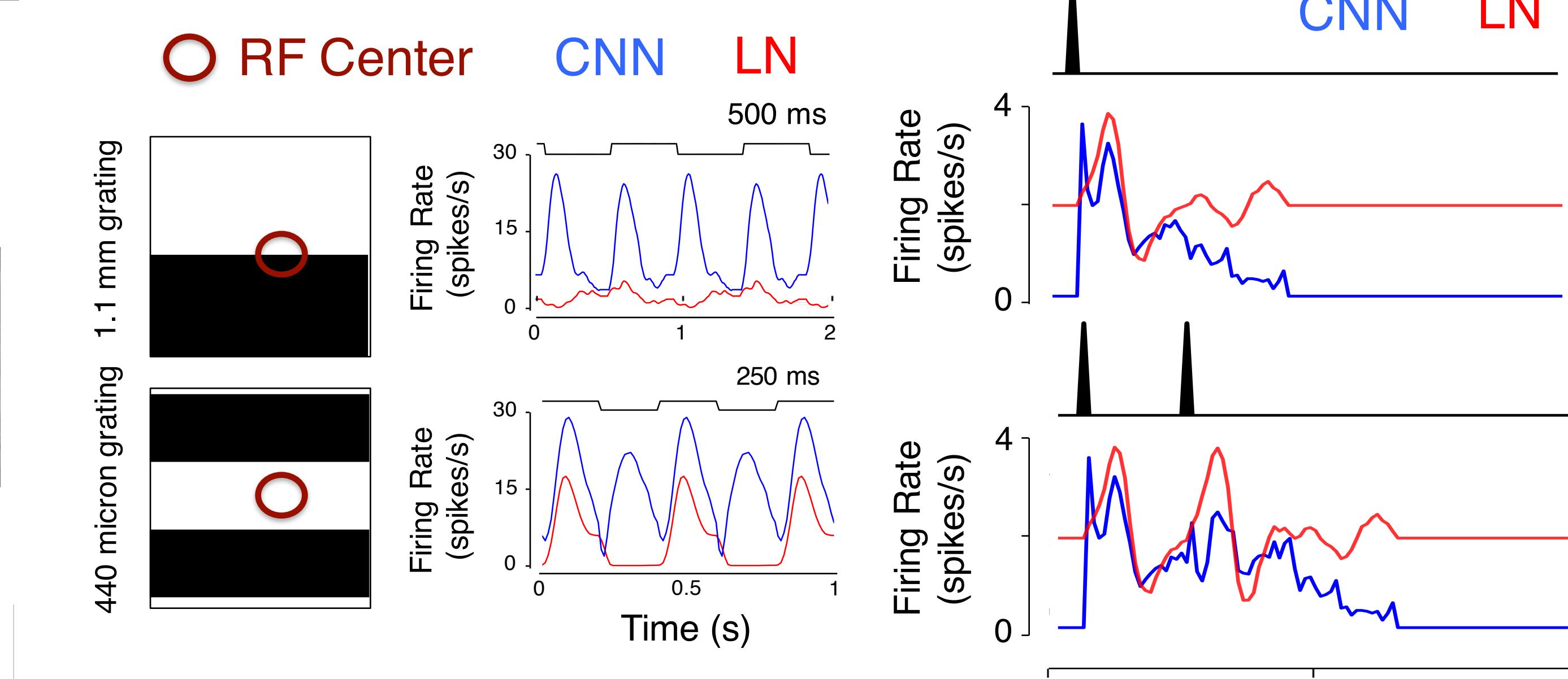


Convolutional filters learn features that have similar spatial bandwidth as bipolar cells; these are pooled to yield inter-units with spatially localized receptive fields. The overall receptive field of the model reproduces the spike-triggered average of the ganglion cell.

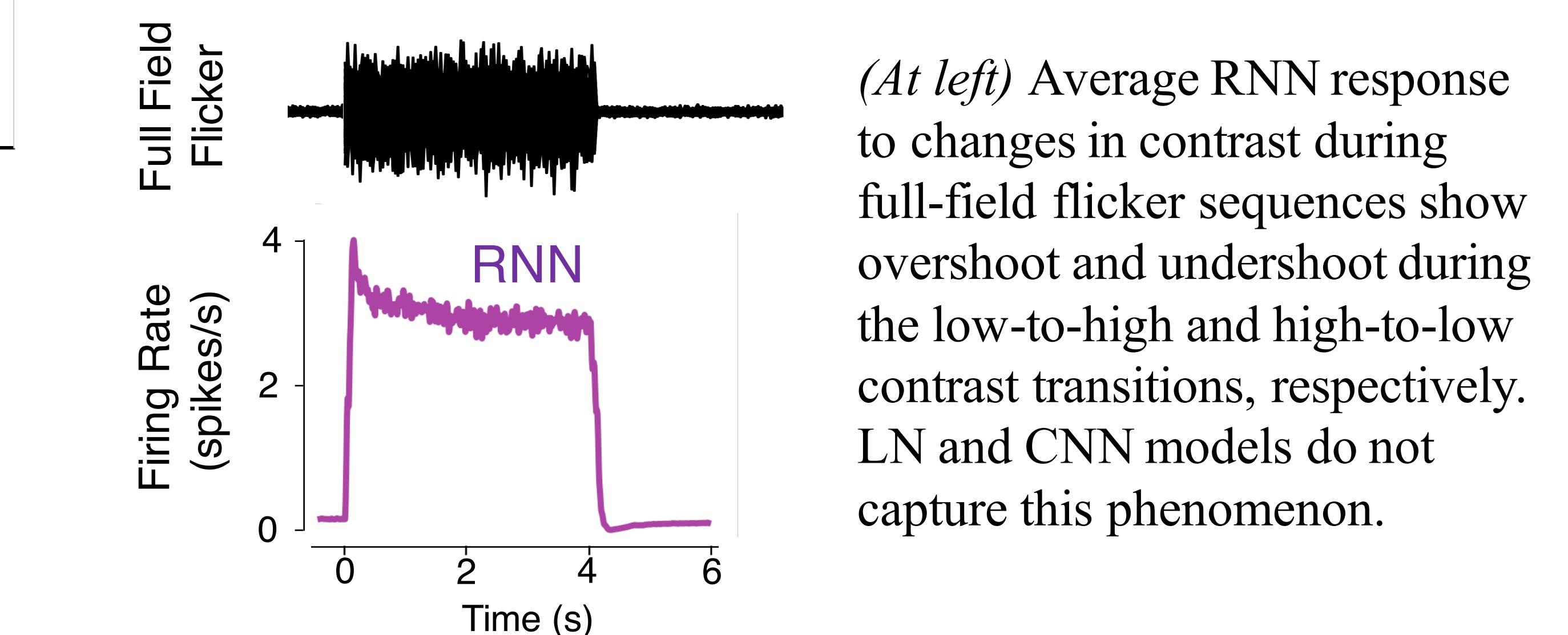
Structured stimuli responses

Frequency doubling with reversing gratings

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



RNNs trained on natural scene stimuli exhibit contrast adaptation



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, lm McIntosh@stanford.edu

