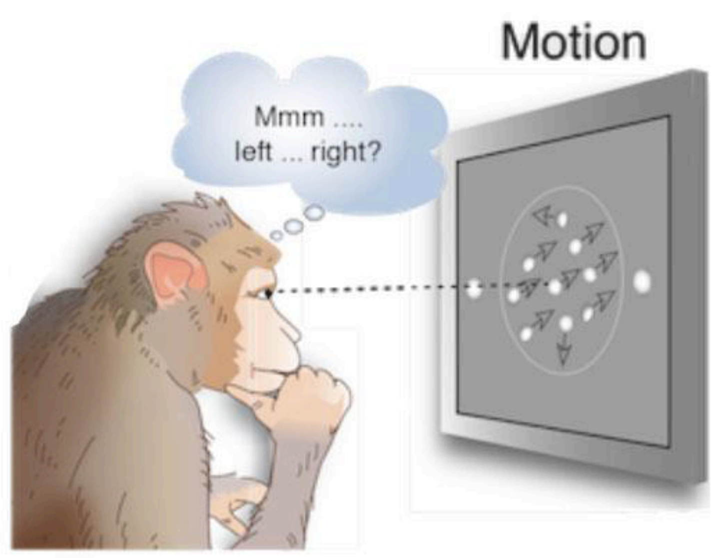
Applying 3D Deep Neural Networks to Human Psychophysics

Daniel Birman, Dylan Cable, Steeve Laquitaine

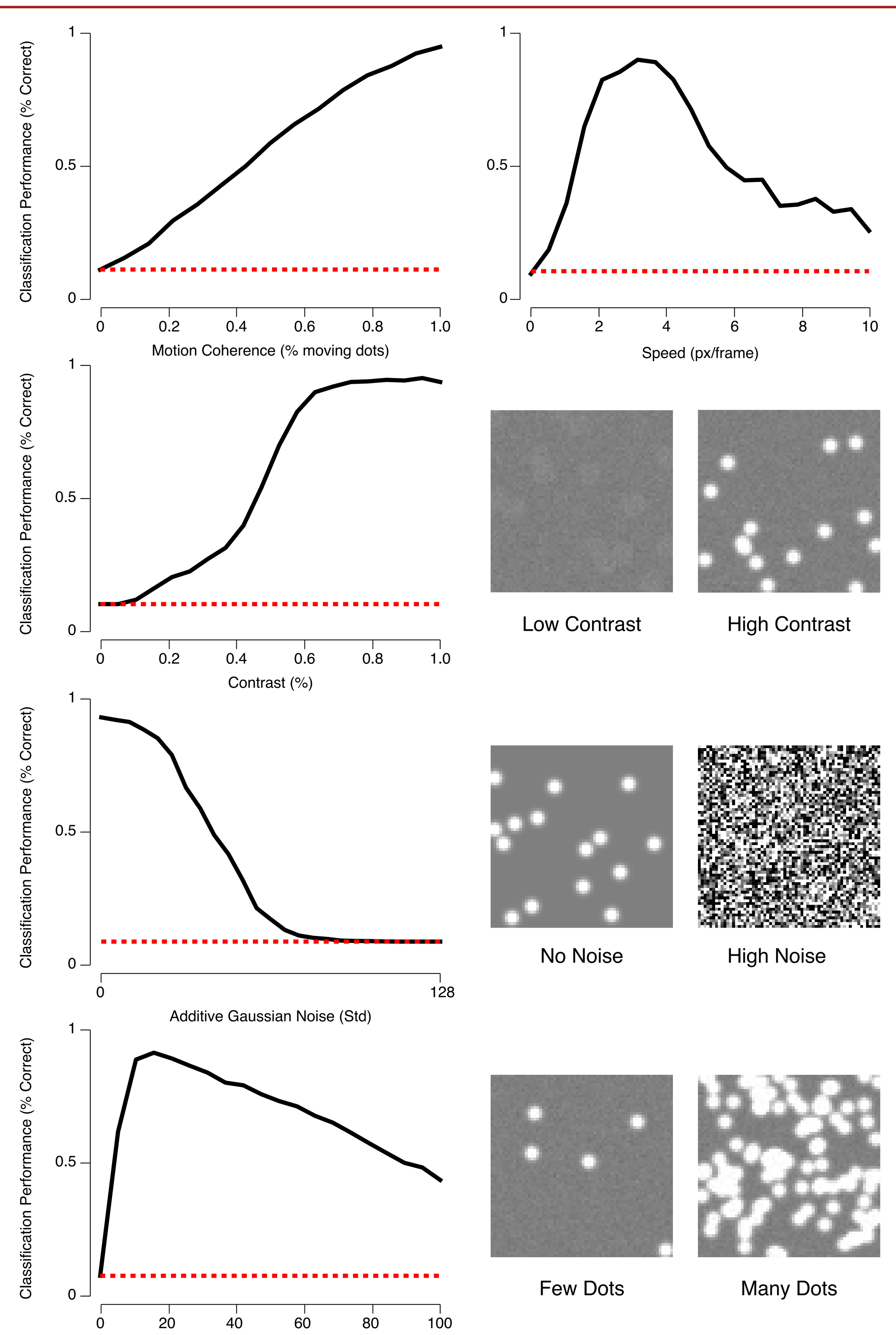
Stanford
Department of Psychology

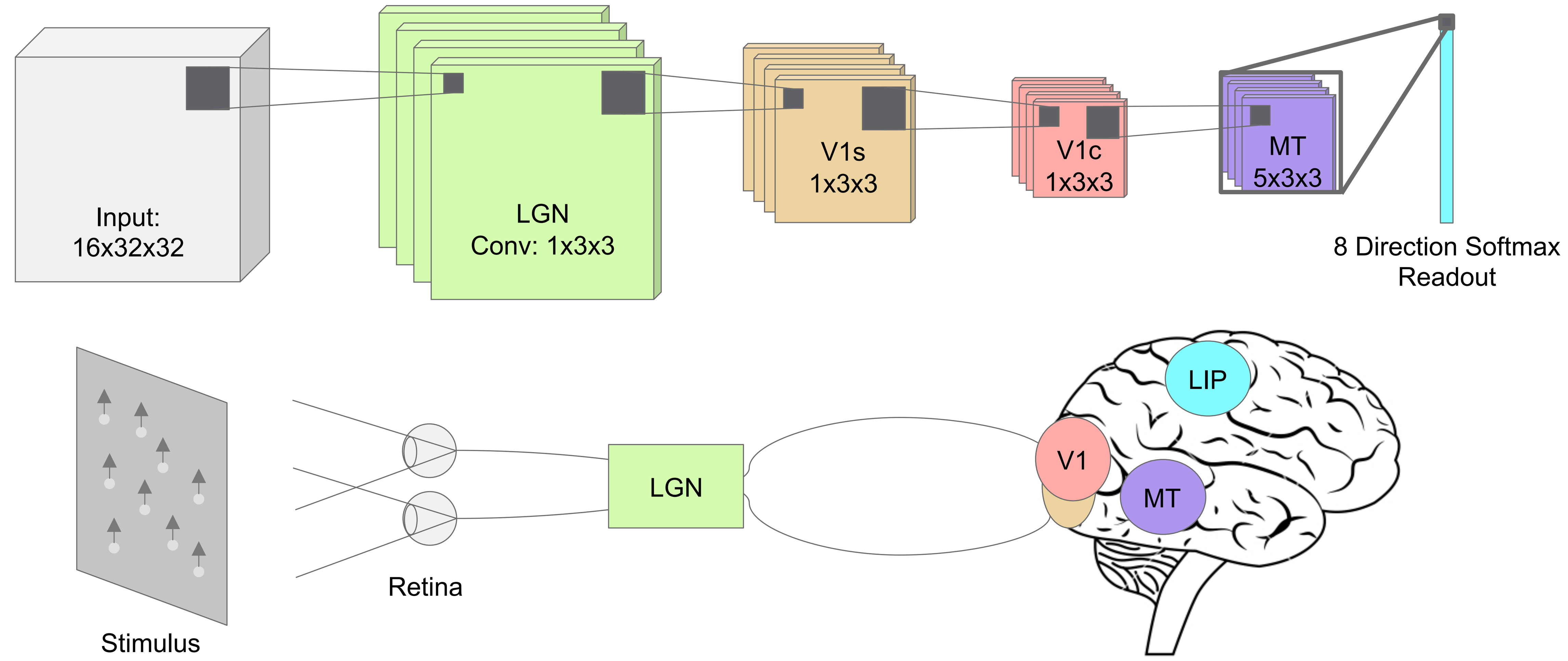
Motivation: Why model human psychophysics?

Training models of the human nervous system **to perform the same task** as humans is a useful tool to help us understand human behavior (Yamins et al. *PNAS* 2014). We built a neural network based on the human visual stream up to cortical area MT. We trained our model to complete a simple behavioral task: discriminate direction. Our model is a modernization of an older model without constraints on the characteristics of internal feature representations (Simoncelli et al. *PNAS* 1996).

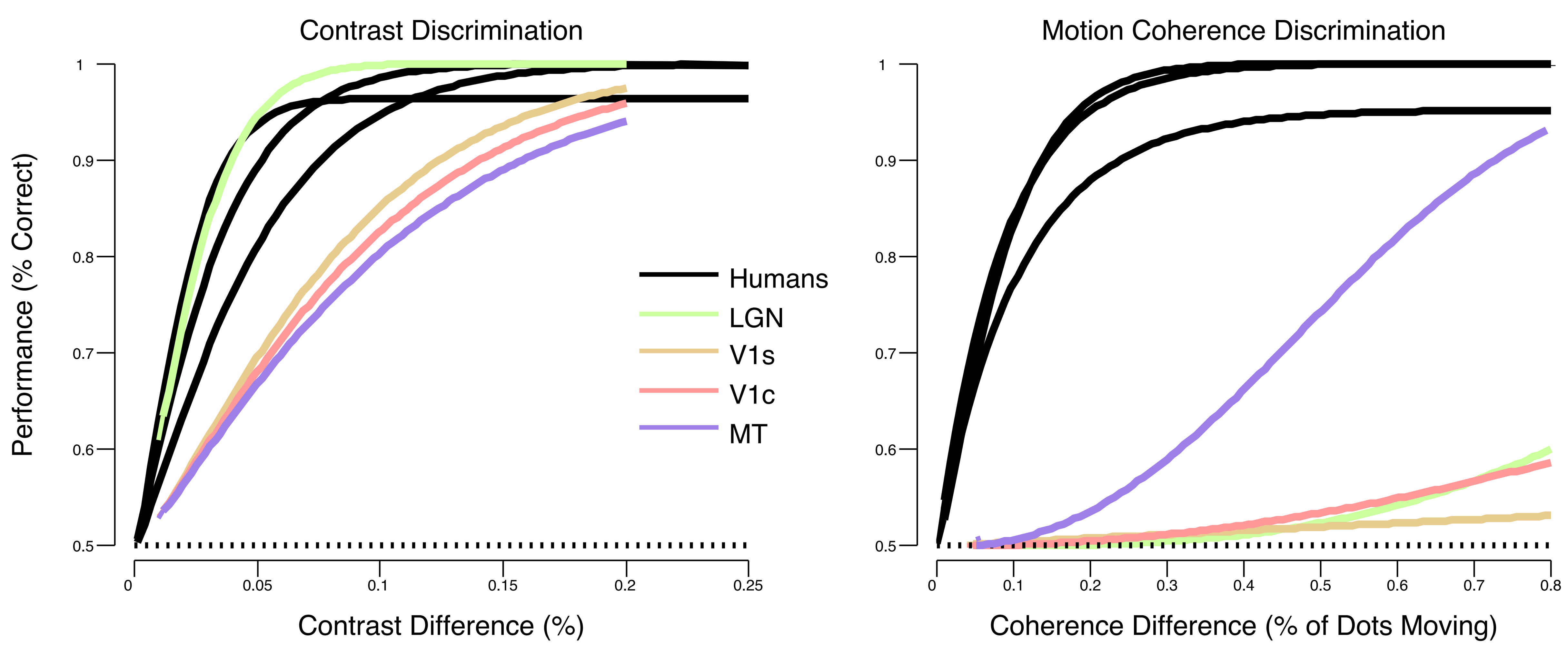


Validation: Classifying 8 direction motion



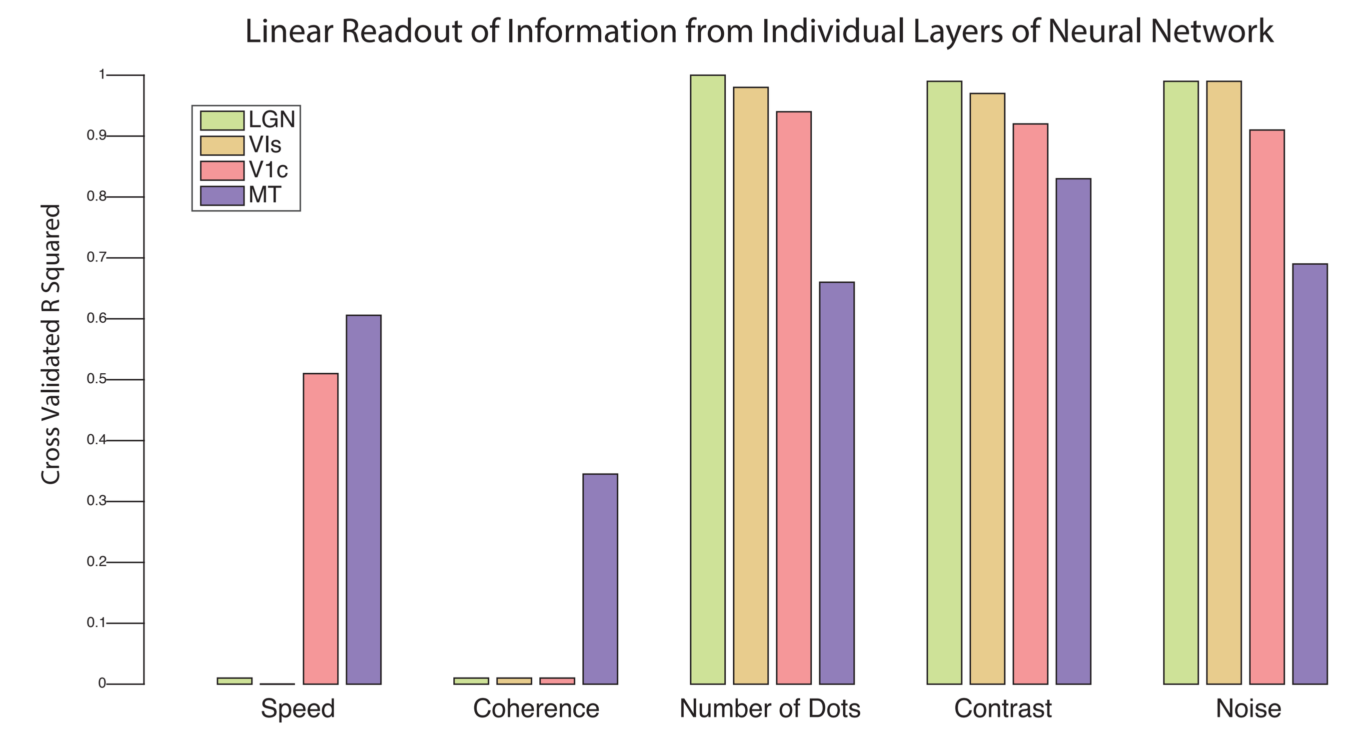


Humans vs. Machines: Performance on an untrained psychophysics task



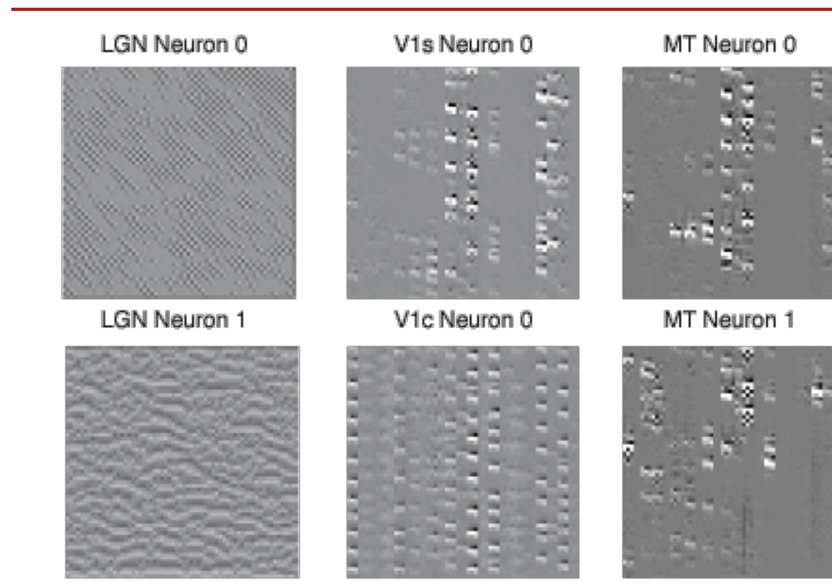
We used our trained model to perform two basic psychophysical tasks to compare its performance against human behavior. We looked at **contrast discrimination** and **motion coherence discrimination**, two tasks that can be done on the same stimulus that we used to train our model (see our website for stimulus examples). Above we plot the amount of stimulus strength difference needed to discriminate which of two patches of dots had higher contrast or greater motion coherence. We compare these results against a model-based linear readout from specific layers. We found that for contrast both human and model performance are qualitatively similar, while for motion coherence our model does not appear to capture the necessary features to decode motion strength in a linear manner--we expect this is due to the nonlinear representation of motion coherence in the MT layer of our model.

Readout: Does our model represent important motion features?



To understand the internal features represented by our model we computed regressions predicting feature strength from layer outputs. To reduce dimensionality we averaged outputs across both space and time. We found that, although not explicitly trained for them, our model represented aspects of our motion stimulus in specific layers.

Feature Inversion



GIF videos of our motion stimuli and feature inversions are available on our website.

Gardner Lab @Stanford University

Acknowledgments
Thank you to the CS231n staff and our TA Lane McIntosh for helpful comments on earlier iterations of our project!

Contact: dbirman@stanford.edu, dcable@stanford.edu, steeve@stanford.edu

http://gru.stanford.edu/doku.php/deepmotion