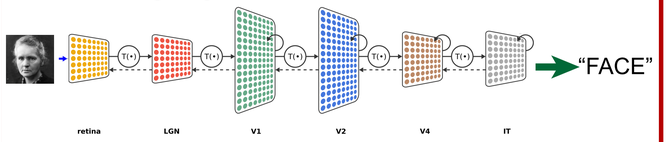
Task-Driven Convolutional Recurrent Models of the Visual System

Aran Nayebi[1]*, Daniel Bear[2]*, Jonas Kubilius[7,8]*, Kohitij Kar[5], Surya Ganguli[4,9], David Sussillo[9], James J. DiCarlo[6,7], Daniel L. K. Yamins[2,3,5]

Neurosciences PhD Program[1], Department of Psychology[2], Computer Science[3], Applied Physics[4], and Wu Tsai Neurosciences Institute[5], Stanford University, Stanford, CA 94305
Department of Brain and Cognitive Sciences[6] and McGovern Institute for Brain Research[7], MIT, Cambridge, MA 02139
Brain and Cognition[8], KU Leuven, Leuven, Belgium and Google Brain[9], Google, Inc., Mountain View, CA 94043

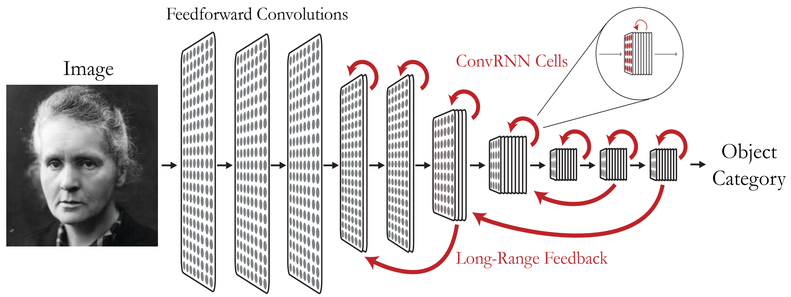
Motivation

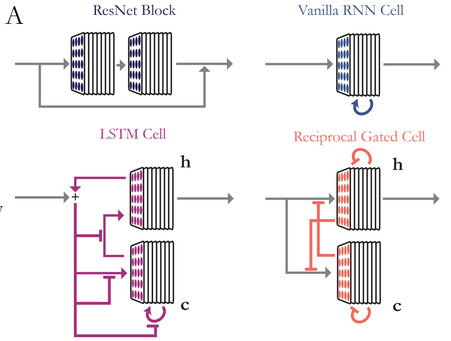
The primate ventral visual stream performs object recognition with a series of feedforward computations. However, it is also rich in local and long-range feedback connections between neurons.

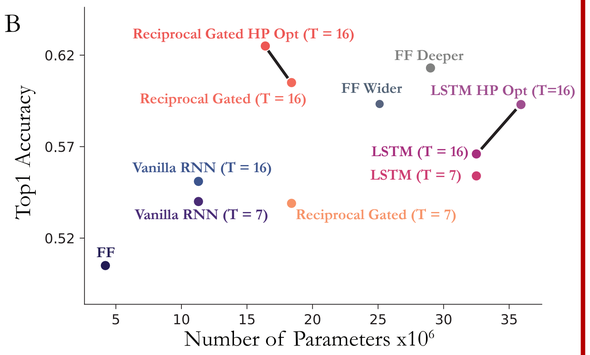


For many images, object category can be decoded from neural population feedforward responses (Majaj, Hong, et al. 2015). But for some images, decoding of object category is delayed (Kar et al., 2018), suggesting an unexplained role for feedback computations. These are not included in feedforward convolutional neural network (CNN) models of the ventral stream.. We therefore augmented CNNs with recurrence (ConvRNNs.)

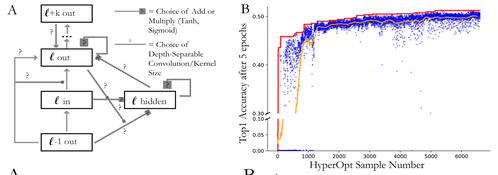
Designing New Convolutional RNNs to Improve Object Recognition

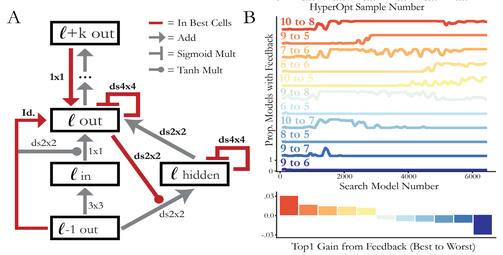




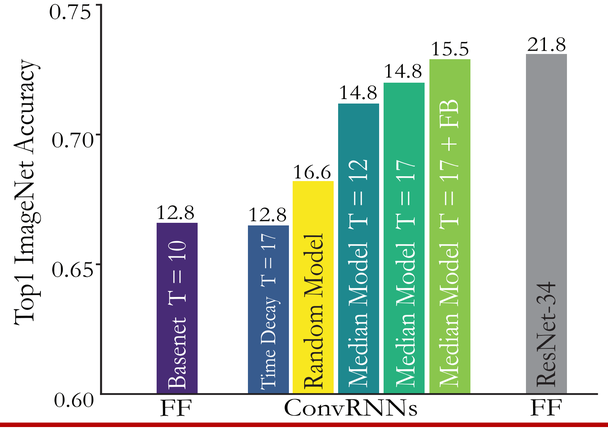


A Large-Scale Search Identified Better Recurrent Architectures

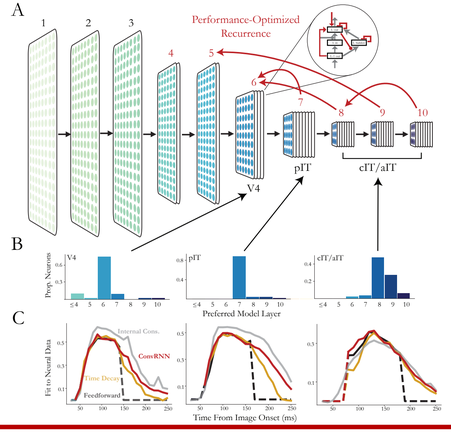




ConvRNNs Match Deeper CNNs on ImageNet Performance



Task-optimized ConvRNNs Predict Neural Dynamics in V4 and IT



Conclusions

- Standard RNN cells (e..g ConvLSTMs) do not improve ImageNet performance when installed in a feedforward model
- The Reciprocal Gated Cell, which includes both gating and bypassing, allows ConvRNNs to match performance of deeper models and larger CNNs
- ConvRNNs explain temporal dynamics in visual cortex better than simpler recurrent models

Next steps include:

- optimize ConvRNNs for other tasks and robustness
- explore more sophisticated decoding from dynamics
- continue search for better recurrent motifs to augment stronger CNN backbones
- compare ConvRNN outputs to primate behavior

This project has received funding from the James S. McDonnell foundation (grant #220020469, D.L.K.Y.), the Simons Foundation (SCGB grants 543061 (D.L.K.Y) and 325500/542965 (J.J.D)), the US National Science Foundation (grant iis-ri1703161, D.L.K.Y.), the European Union's Horizon 2020 research and innovation programme (agreement #705498, J.K.), the US National Eye Institute (grant R01-EY014970, J.J.D.), and the US Office of Naval Research (MURI-114407, J.J.D.). We thank Google (TPUv2 team) and the NVIDIA corporation for generous donation of hardware resources.