Task-Driven Convolutional Recurrent Neural Network Models of Dynamics in Higher Visual Cortex

SFN 2019

2019.10.21

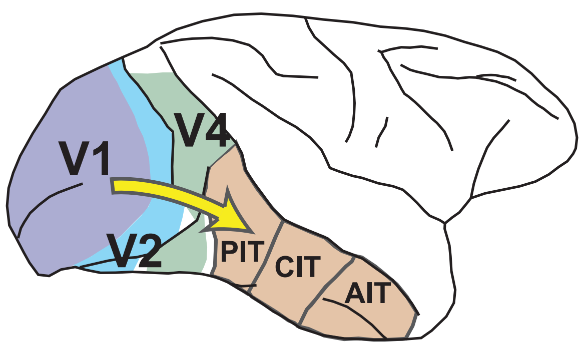
Aran Nayebi
Neurosciences PhD Program
Stanford University

Stanford
Neuro AI Lab
Stanford Neuroscience and Artificial Intelligence Lab

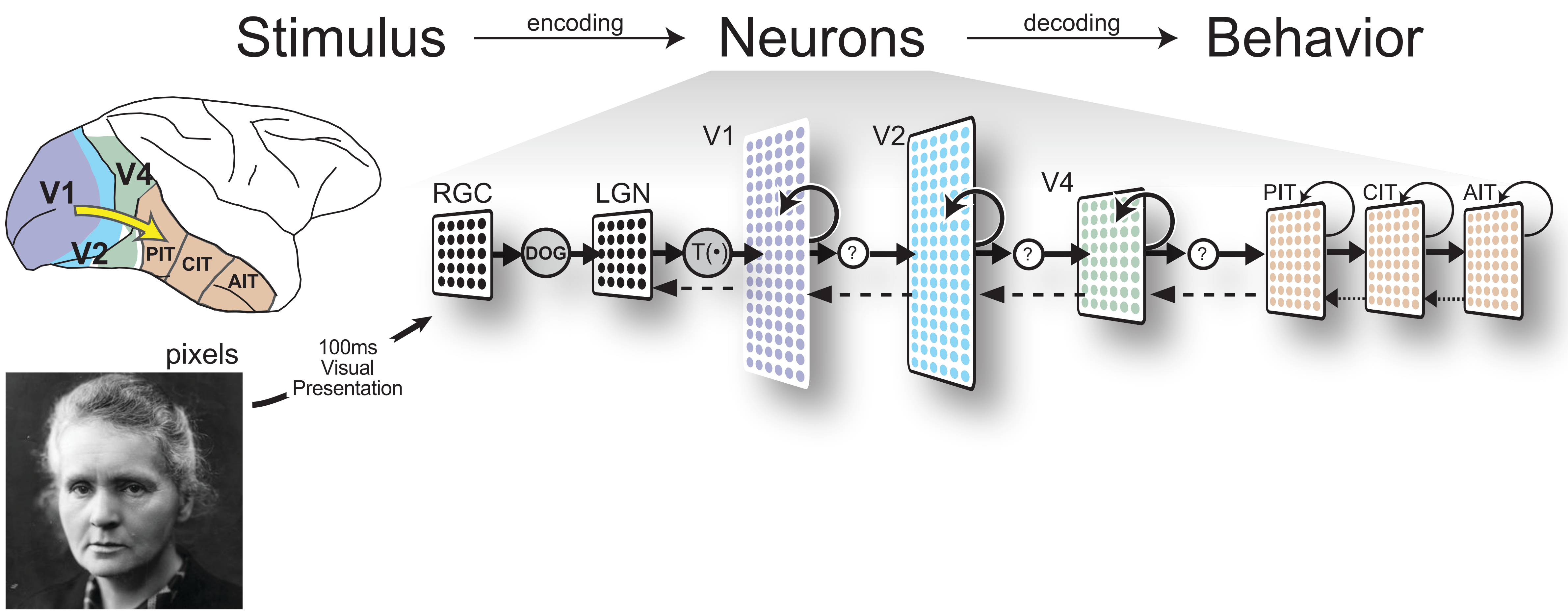
Object Recognition is Hard (But Easy for Us)

…thanks to the Ventral Stream

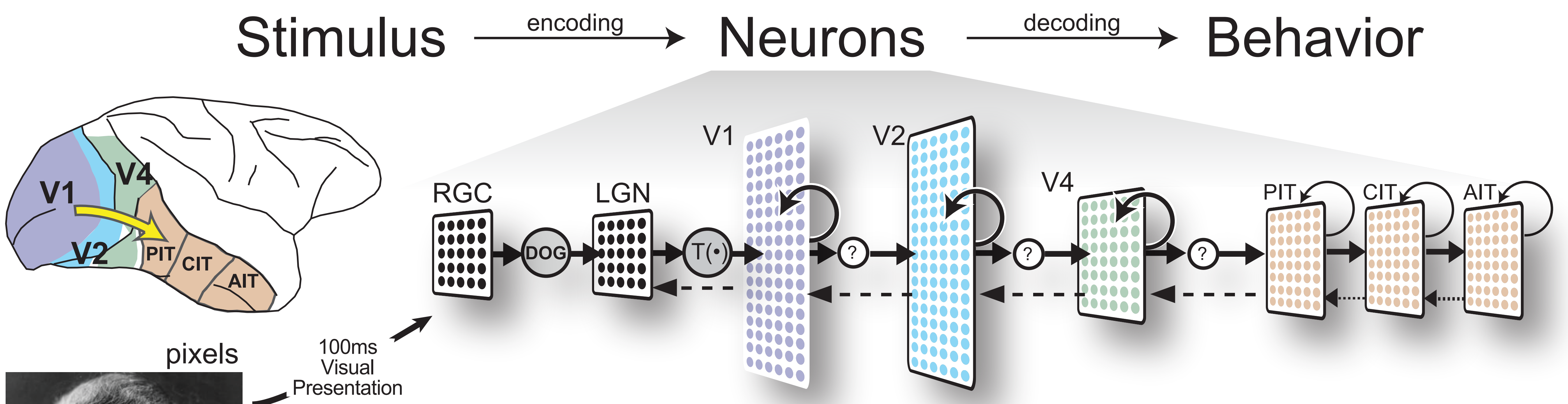
Stimulus $\xrightarrow{\text{encoding}}$ Neurons $\xrightarrow{\text{decoding}}$ Behavior

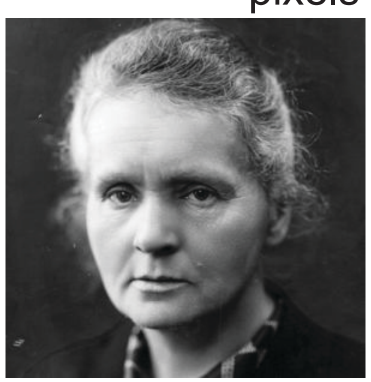


…thanks to the Ventral Stream



CNNs as Models of Object Recognition





Convolutional Neural Networks (CNNs)

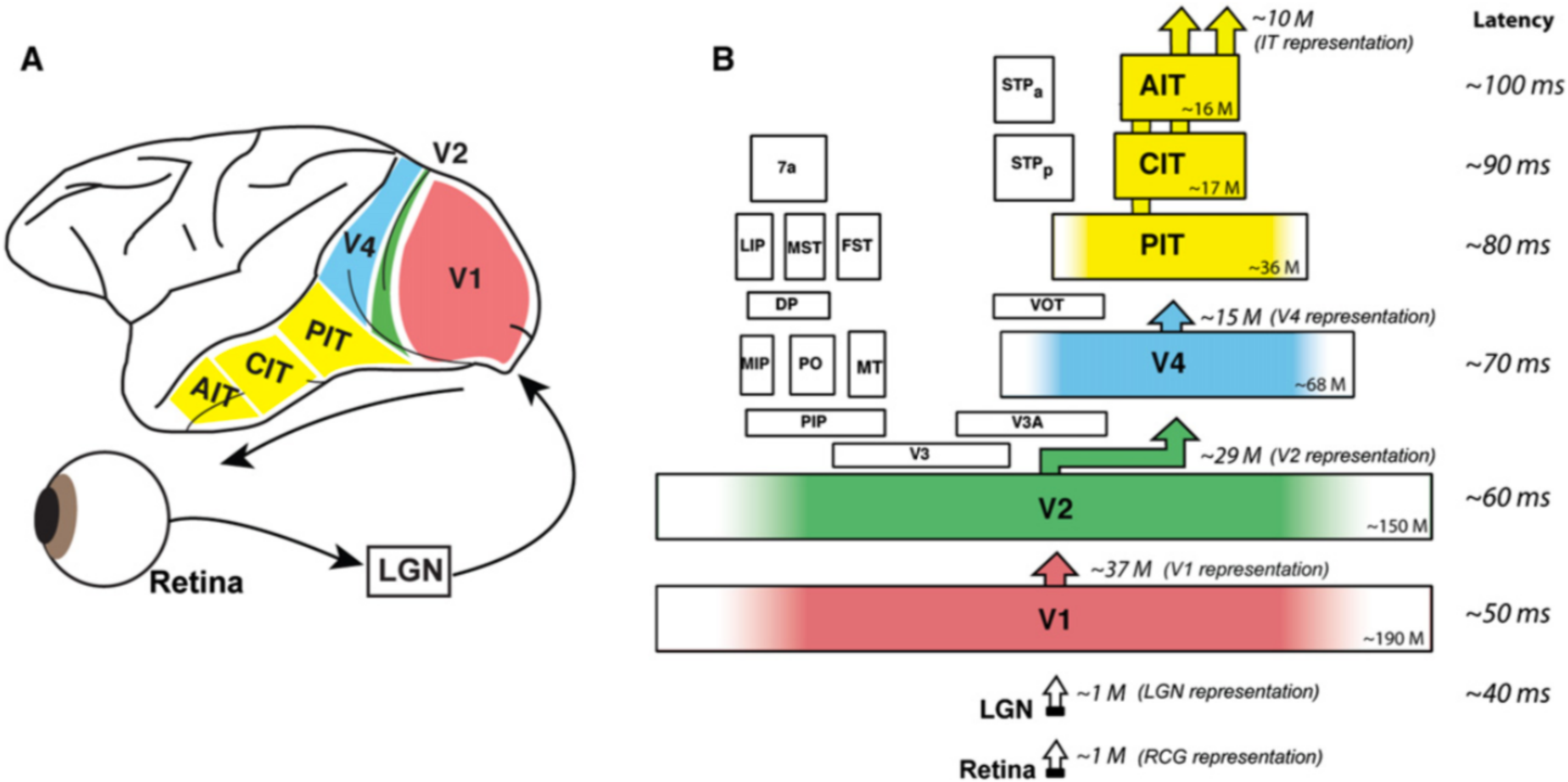
Fukushima, 1979; Lecun, 1995

CNNs are inspired by visual neuroscience:

1) **hierarchy**

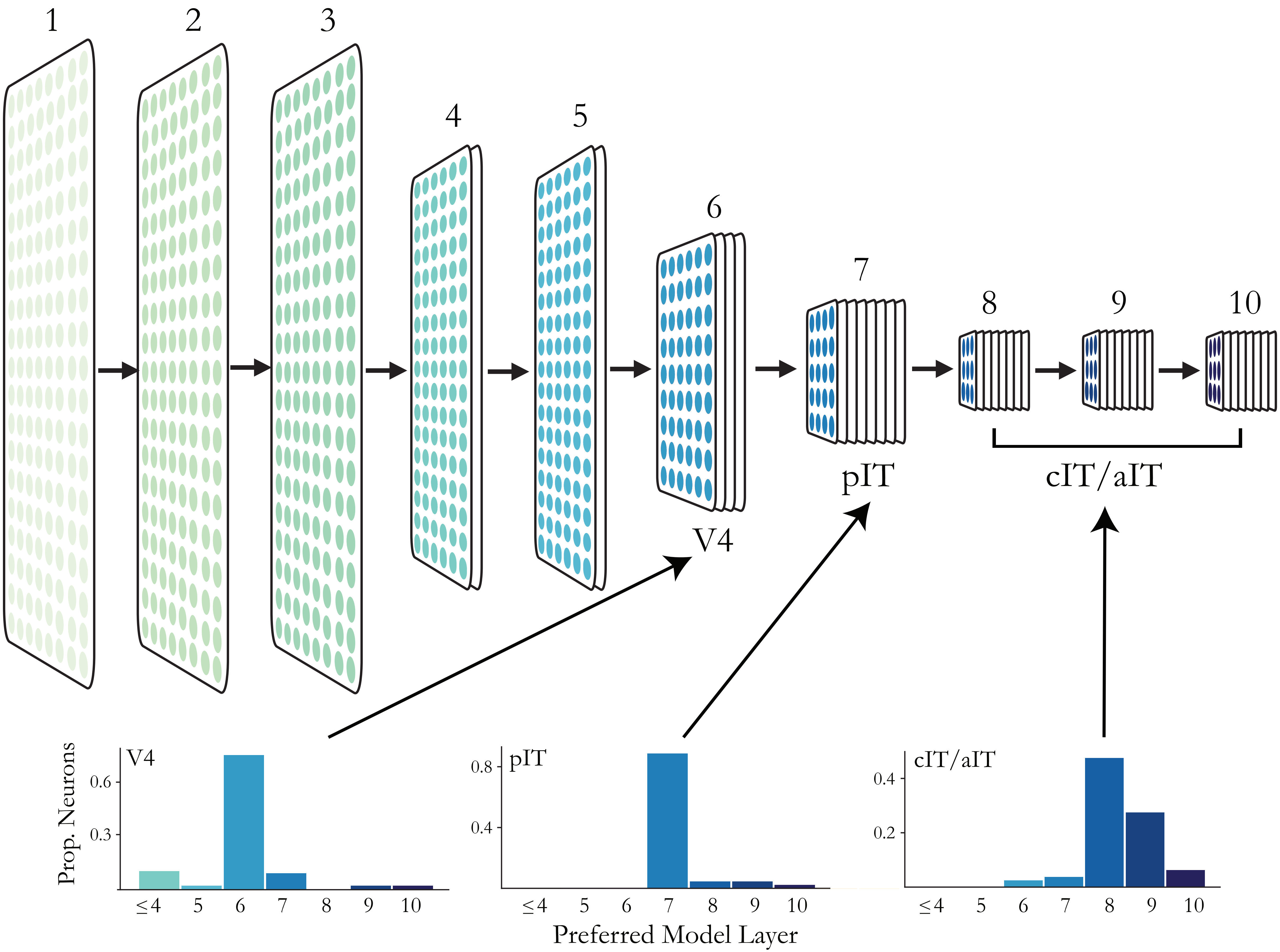
2) **retinotopy** (spatially tiled)

~10-12 "Layers" Plausible based on anatomy and timing

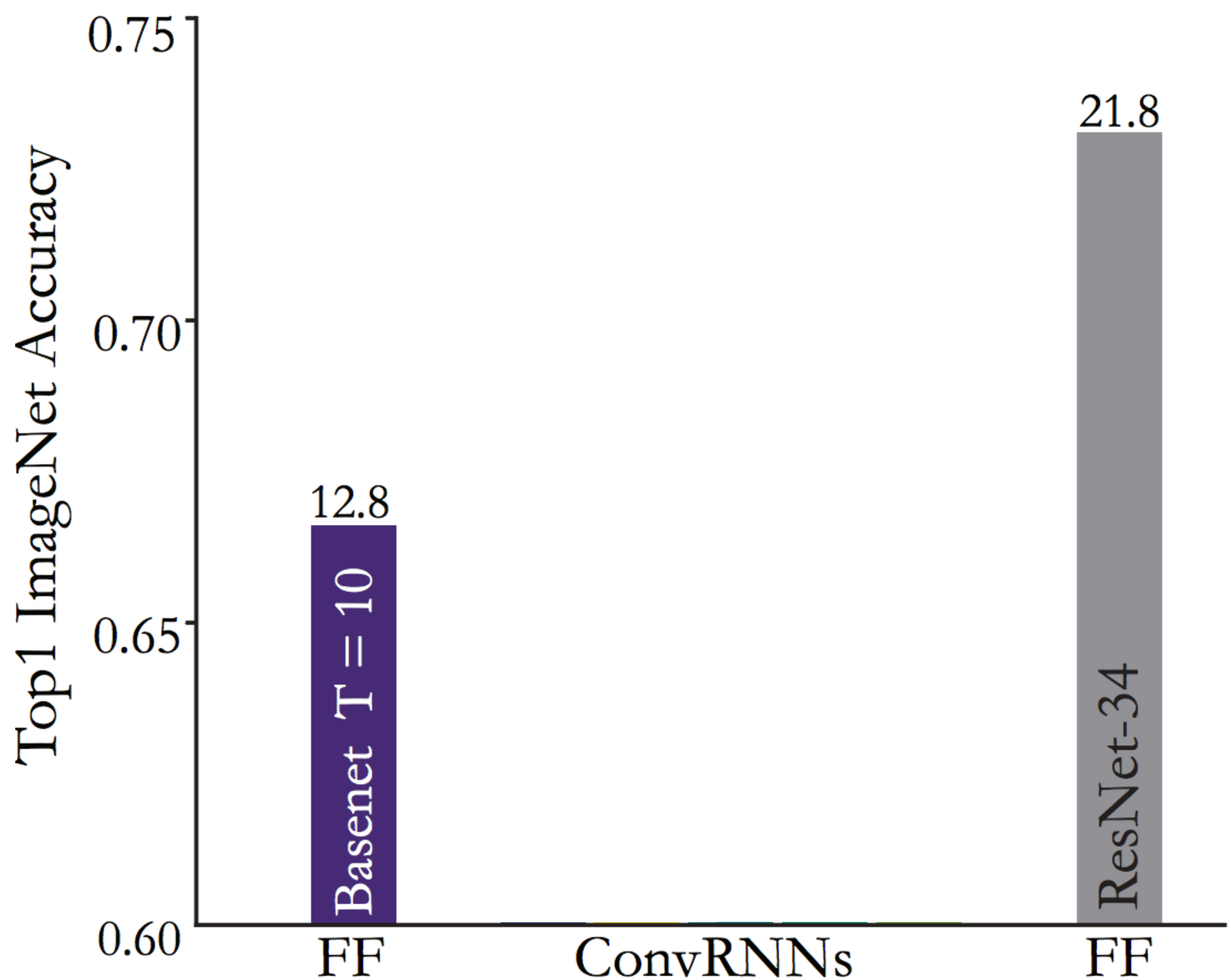


DiCarlo, Zoccolan, & Rust 2012

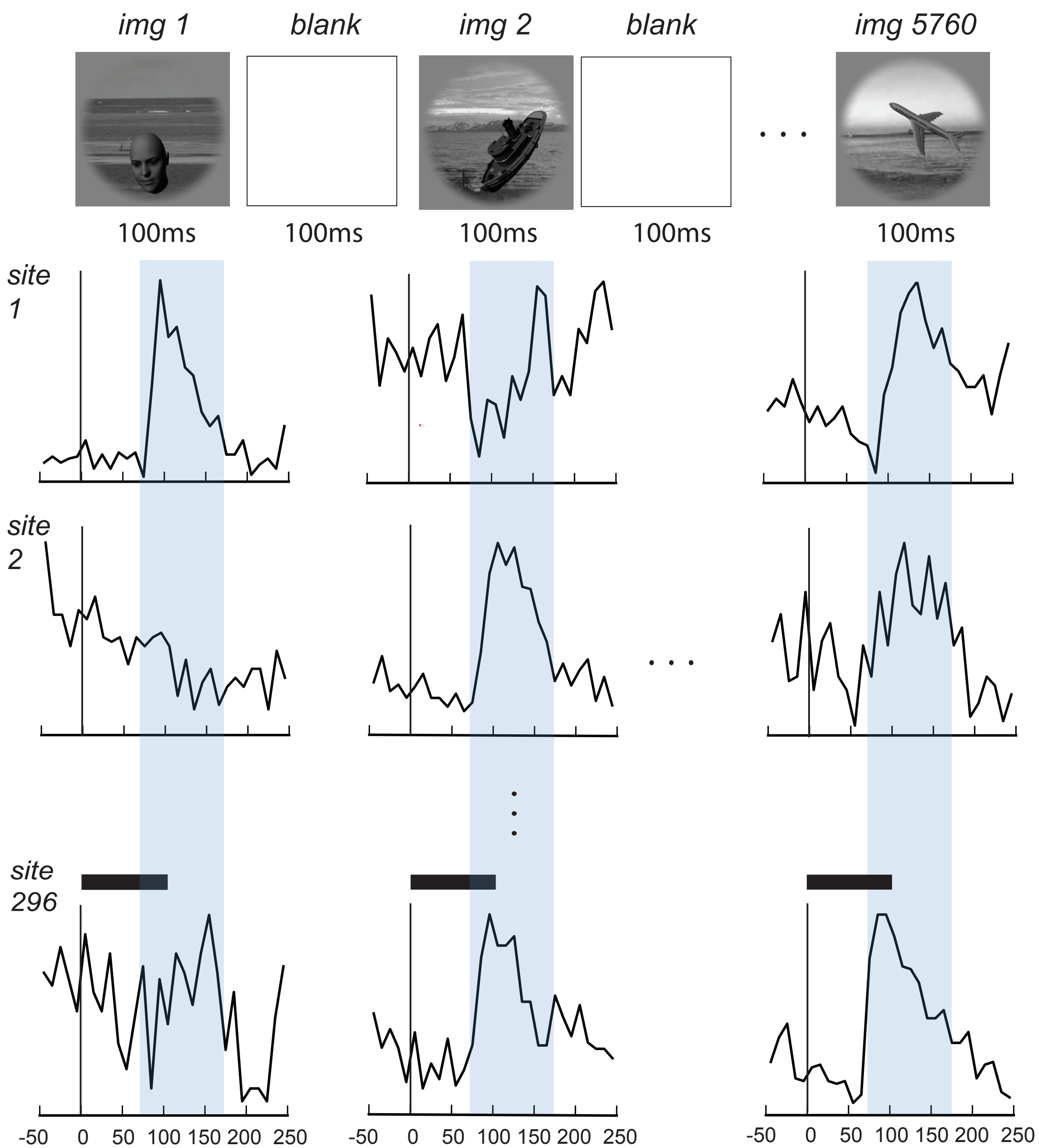
~10-12 "Layers" Plausible based on anatomy and timing



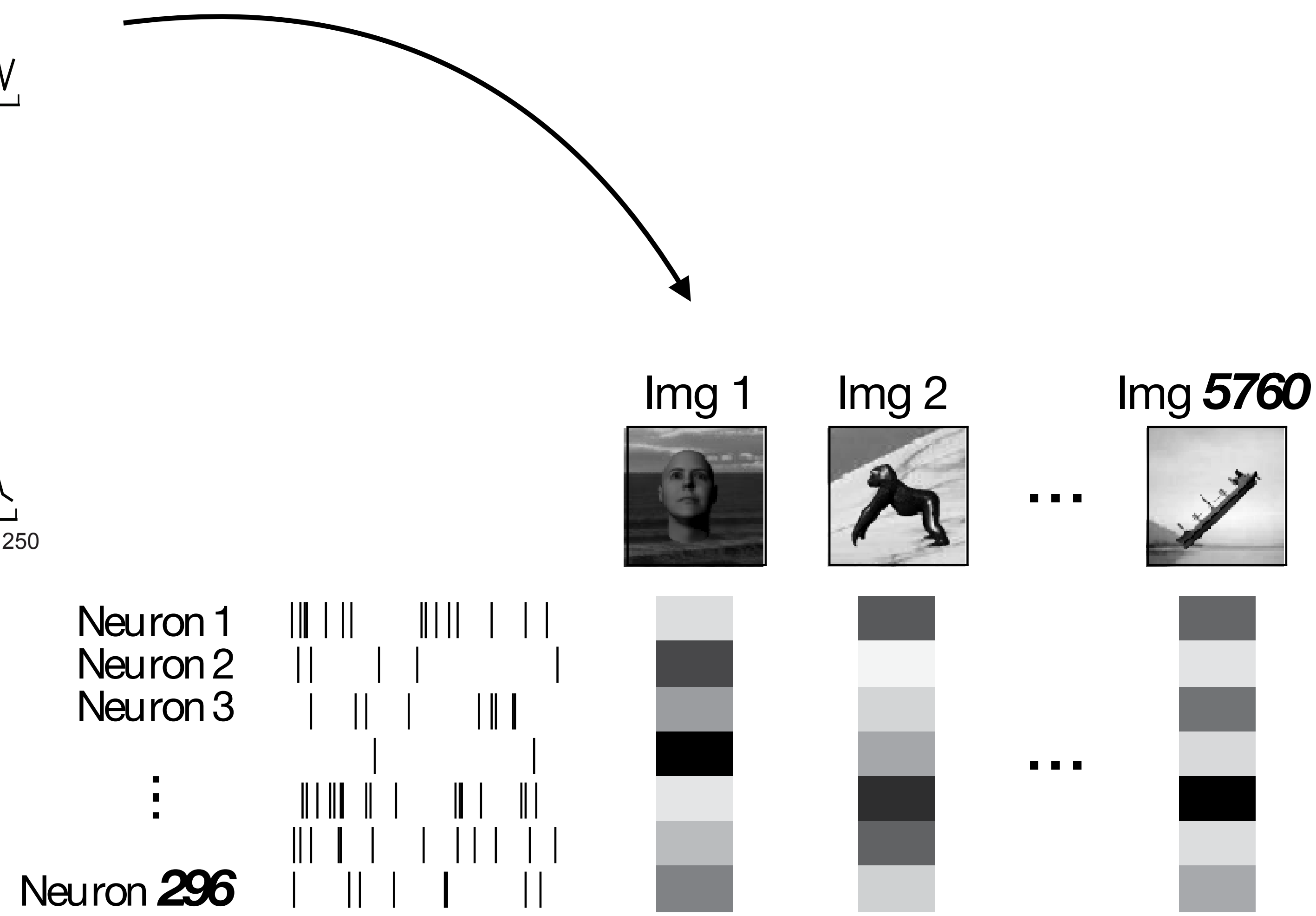
…But, such Networks Are Far From Human Performance



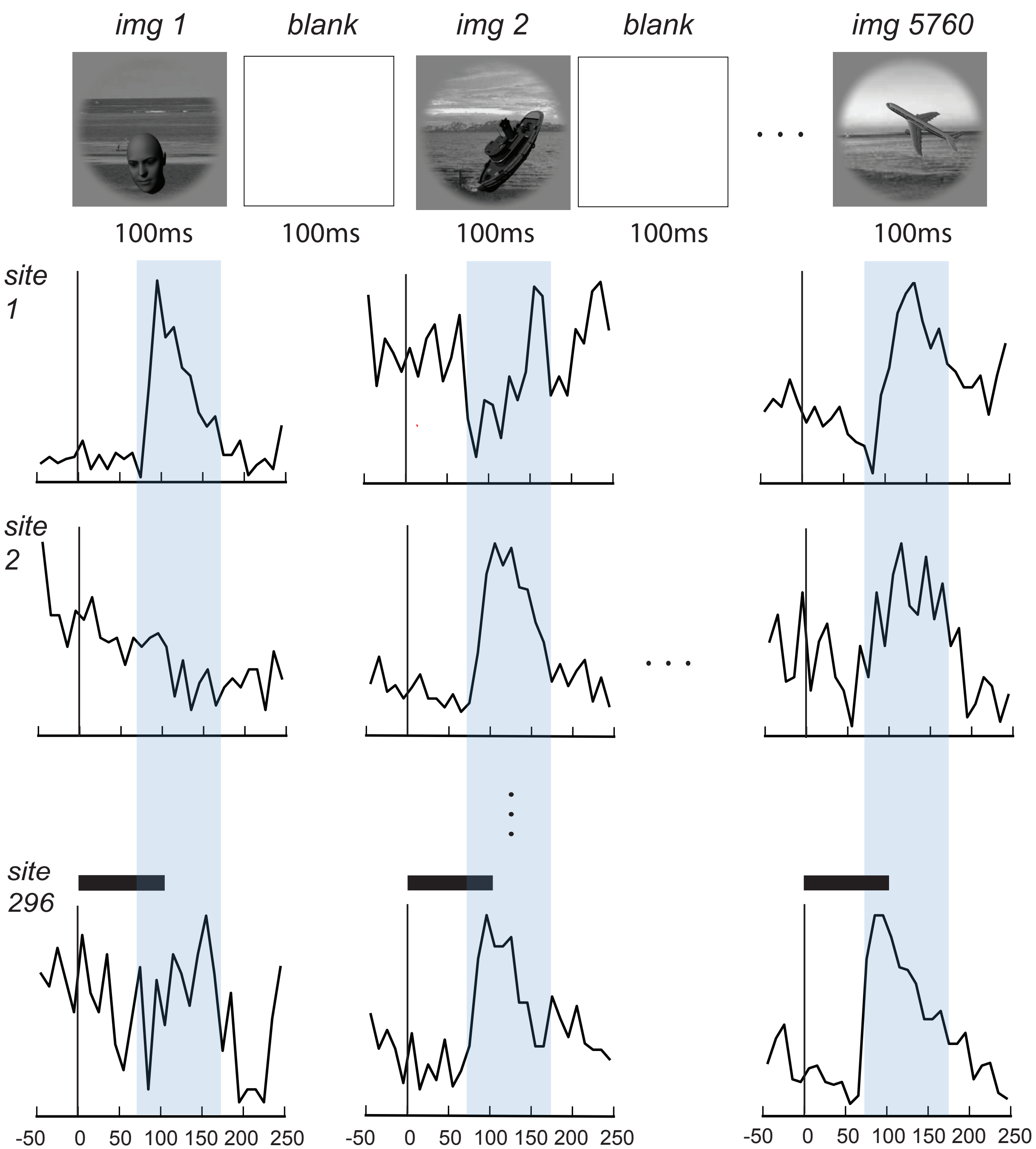
So far, only explaining temporal average of responses



e.g. Binned spike counts 70ms-170ms post stimulus presentation

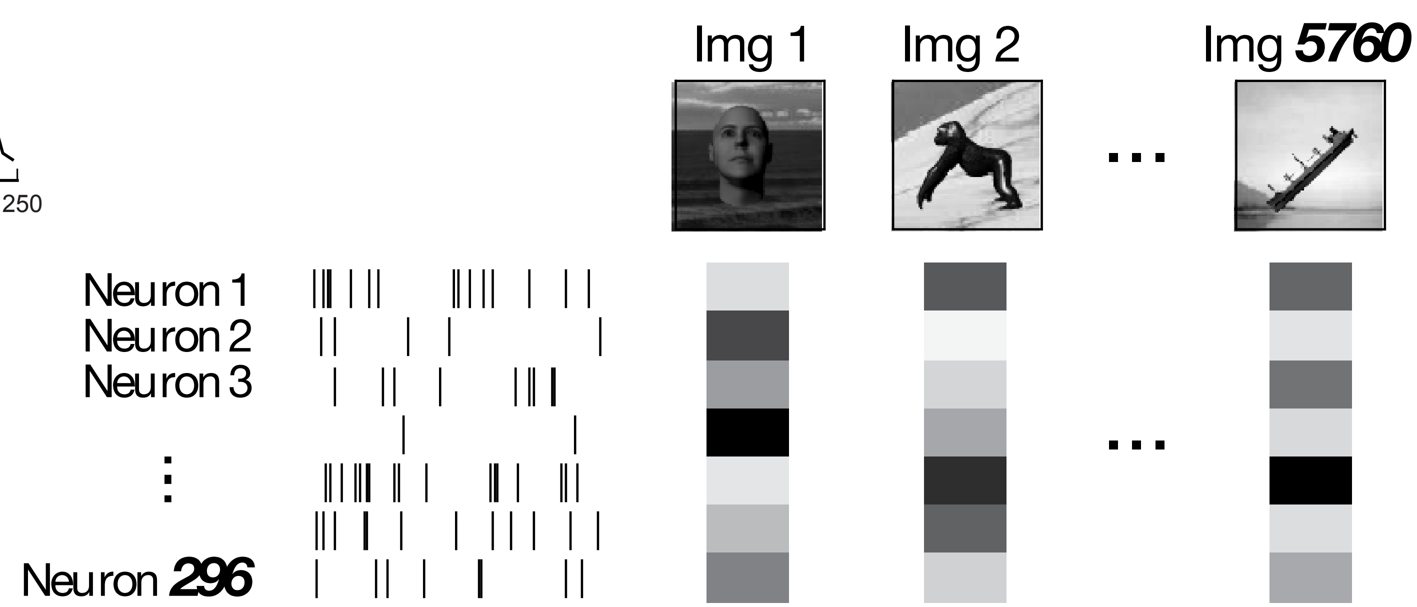


So far, only explaining temporal average of responses



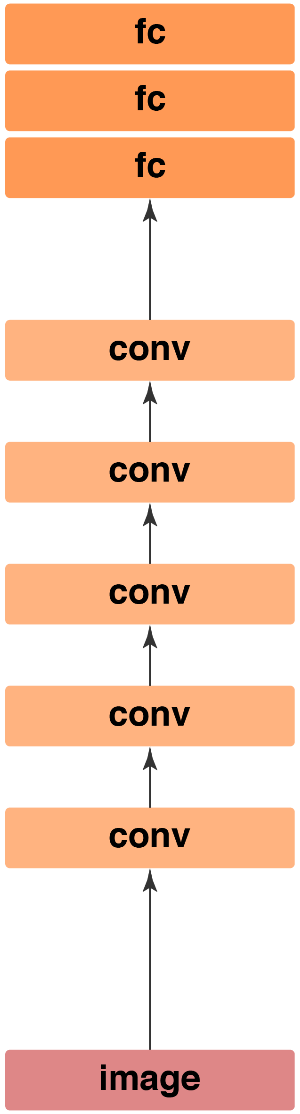
e.g. Binned spike counts 70ms-170ms post stimulus presentation

but actually the data is highly reliable at much finer grain — 10ms bins

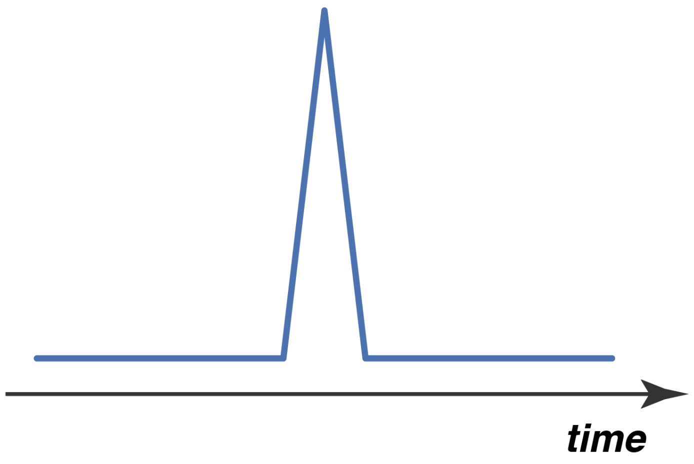


Trajectory Possibilities

Simple feedforward networks simple dynamics:

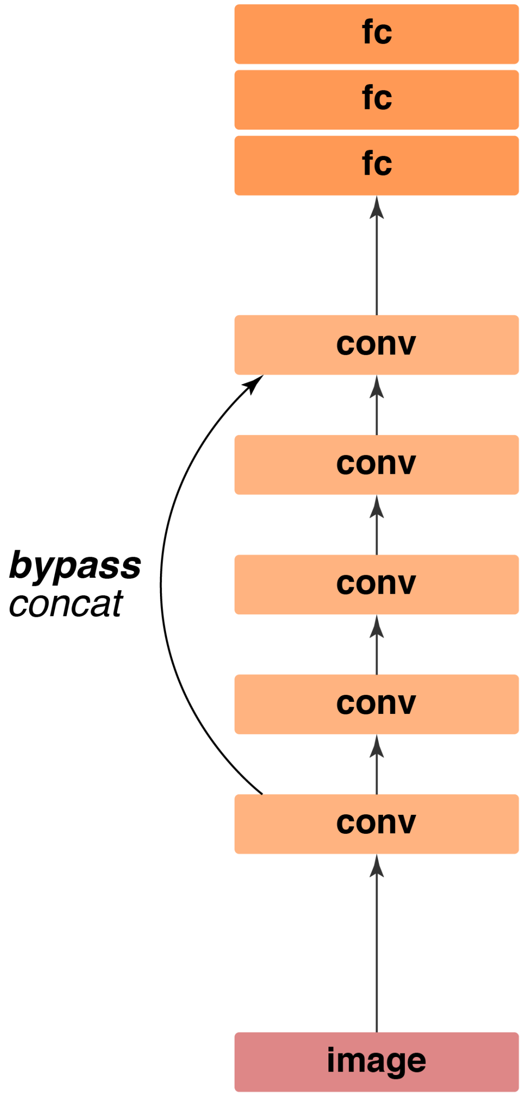


courtesy Jonas Kubilius

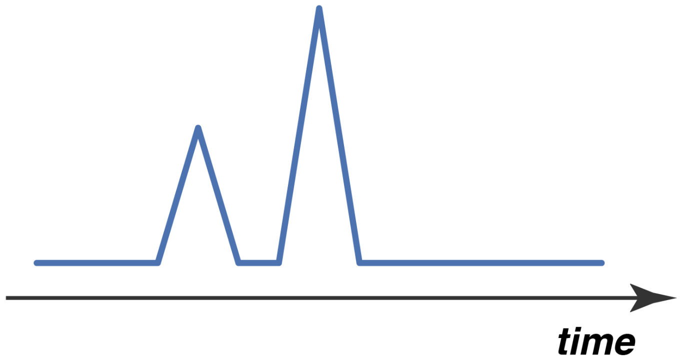


Trajectory Possibilities

Dynamics more interesting with bypasses:

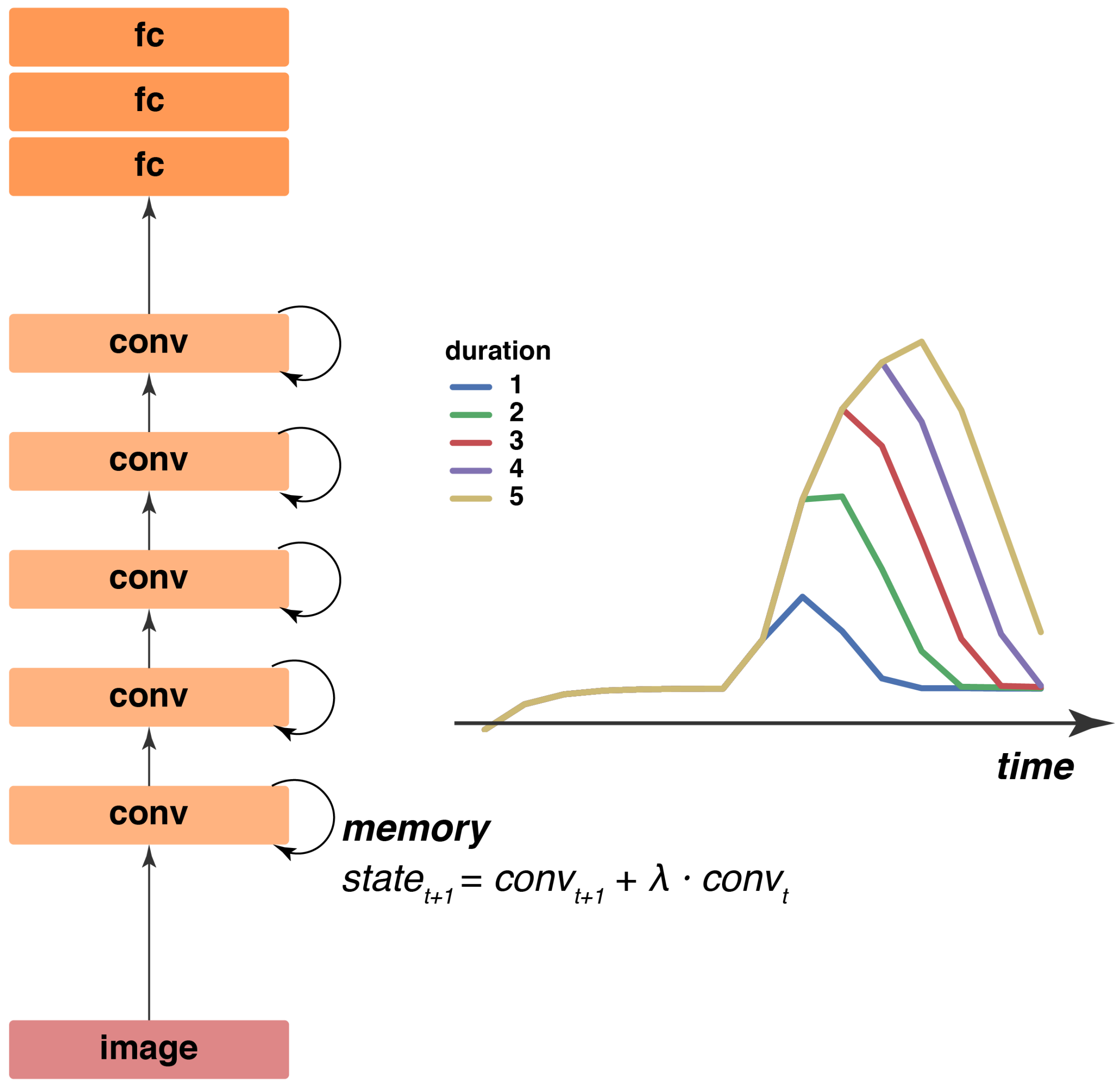


courtesy Jonas Kubilius



Trajectory Possibilities

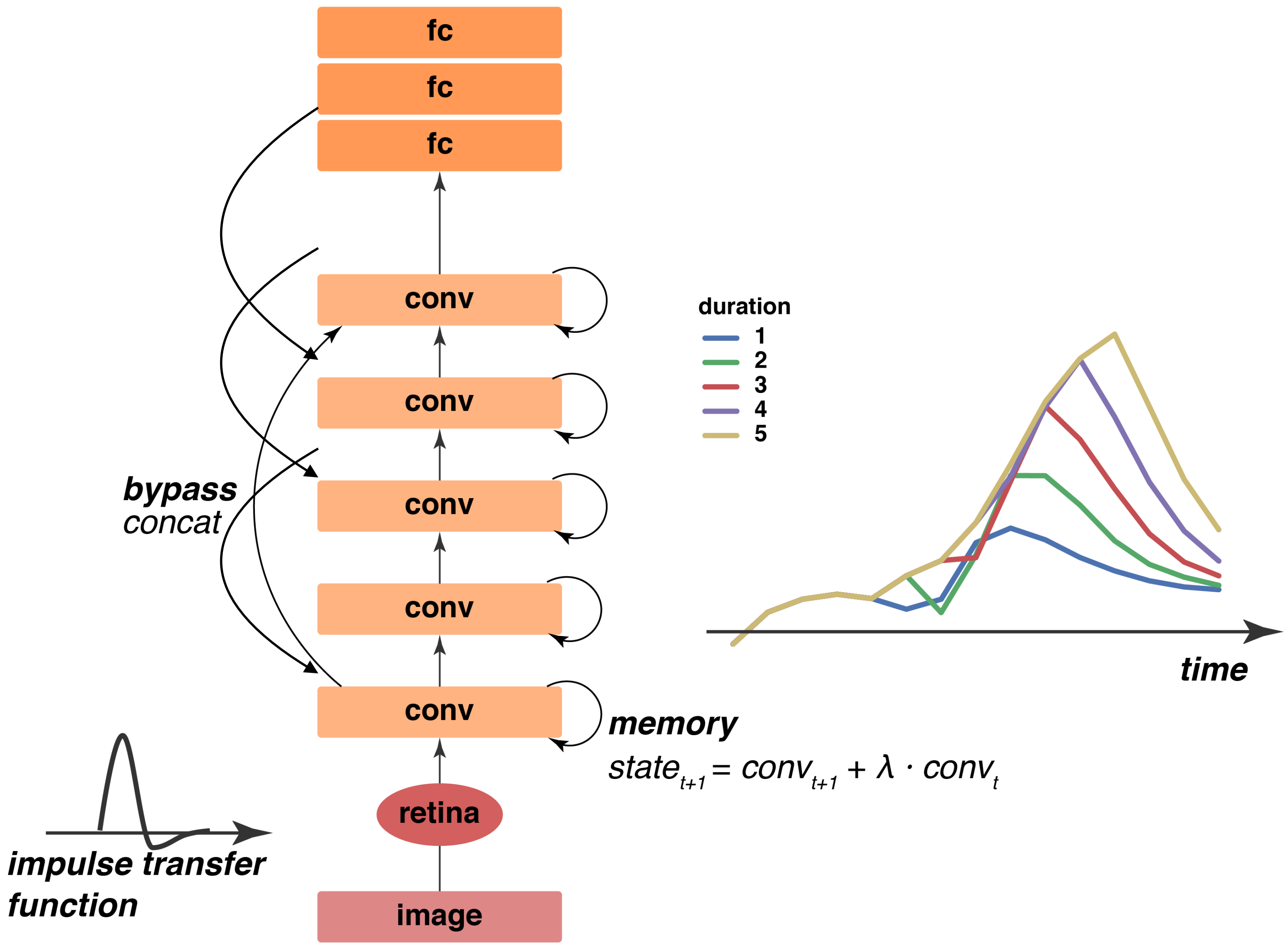
Dynamics more interesting with bypasses, local recurrence:



courtesy Jonas Kubilius

Trajectory Possibilities

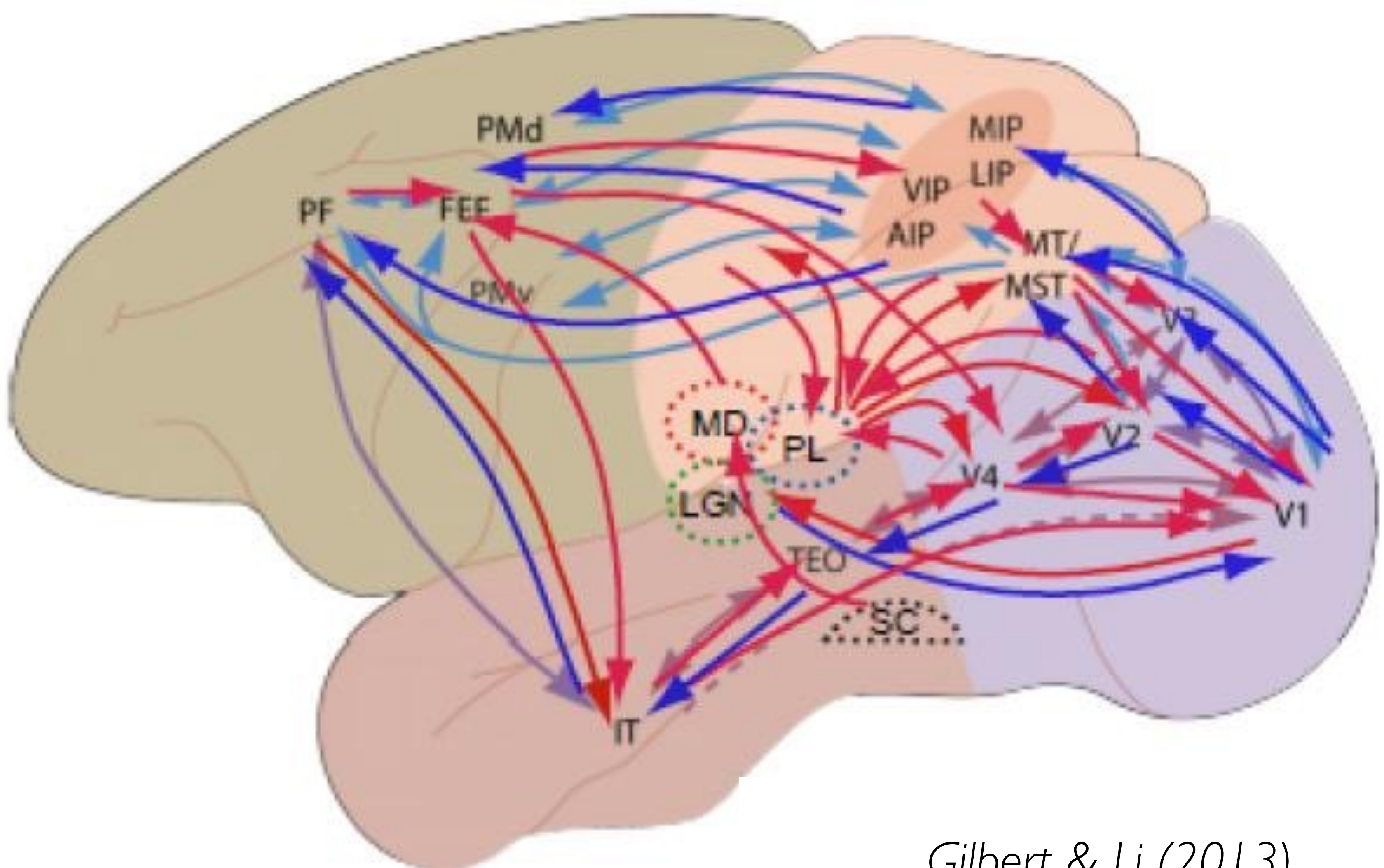
Dynamics more interesting with bypasses, local recurrence, long-range feedback:



courtesy Jonas Kubilius

Dynamics result from recurrence

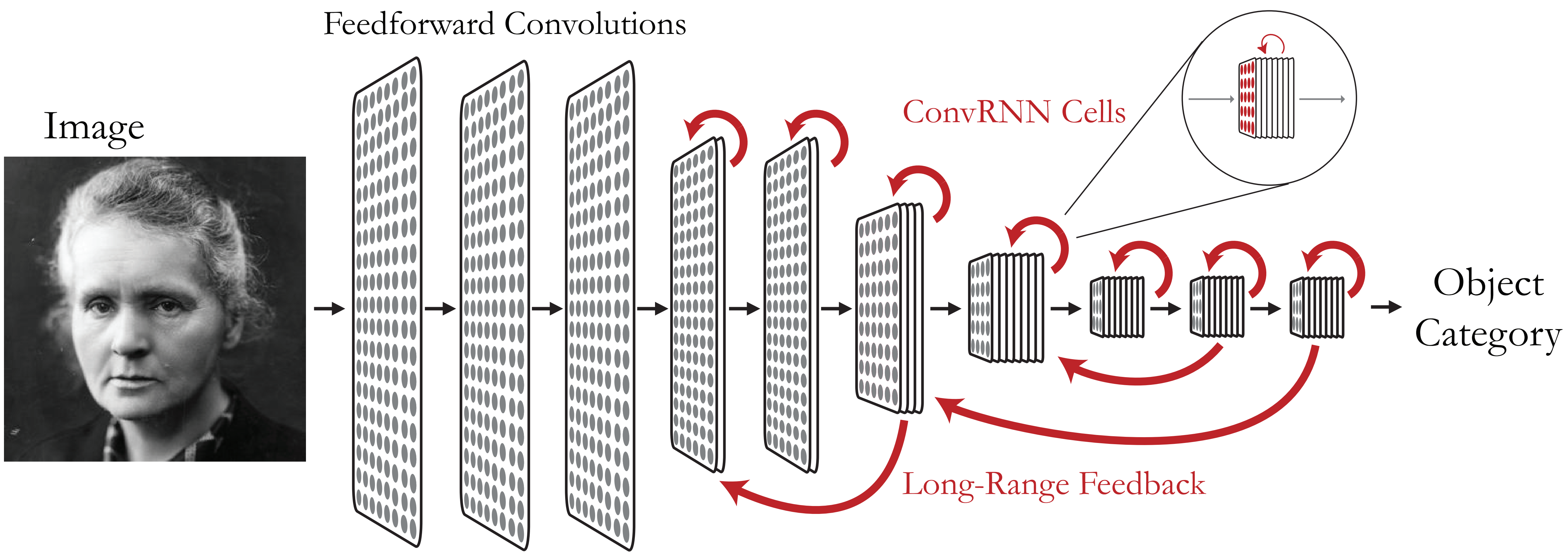
Feedbacks are everywhere anatomically:



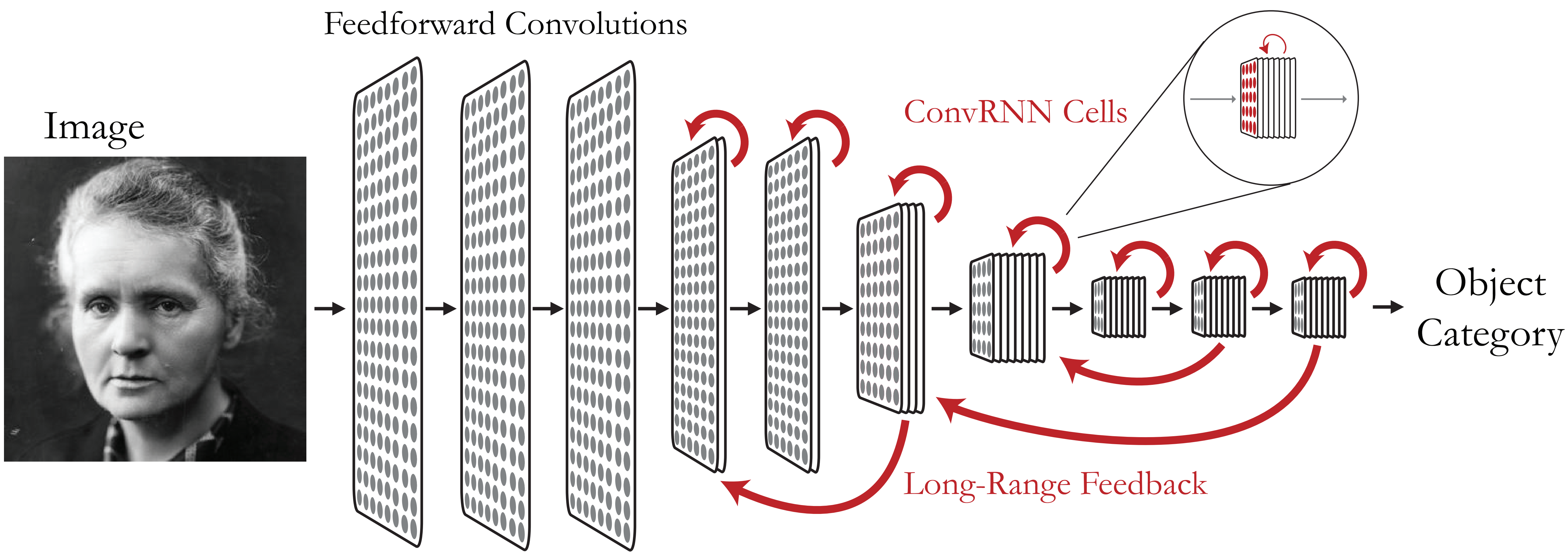
Gilbert & Li (2013)

. . . but what are they for?

Convolutional Recurrent Neural Networks (ConvRNNs)

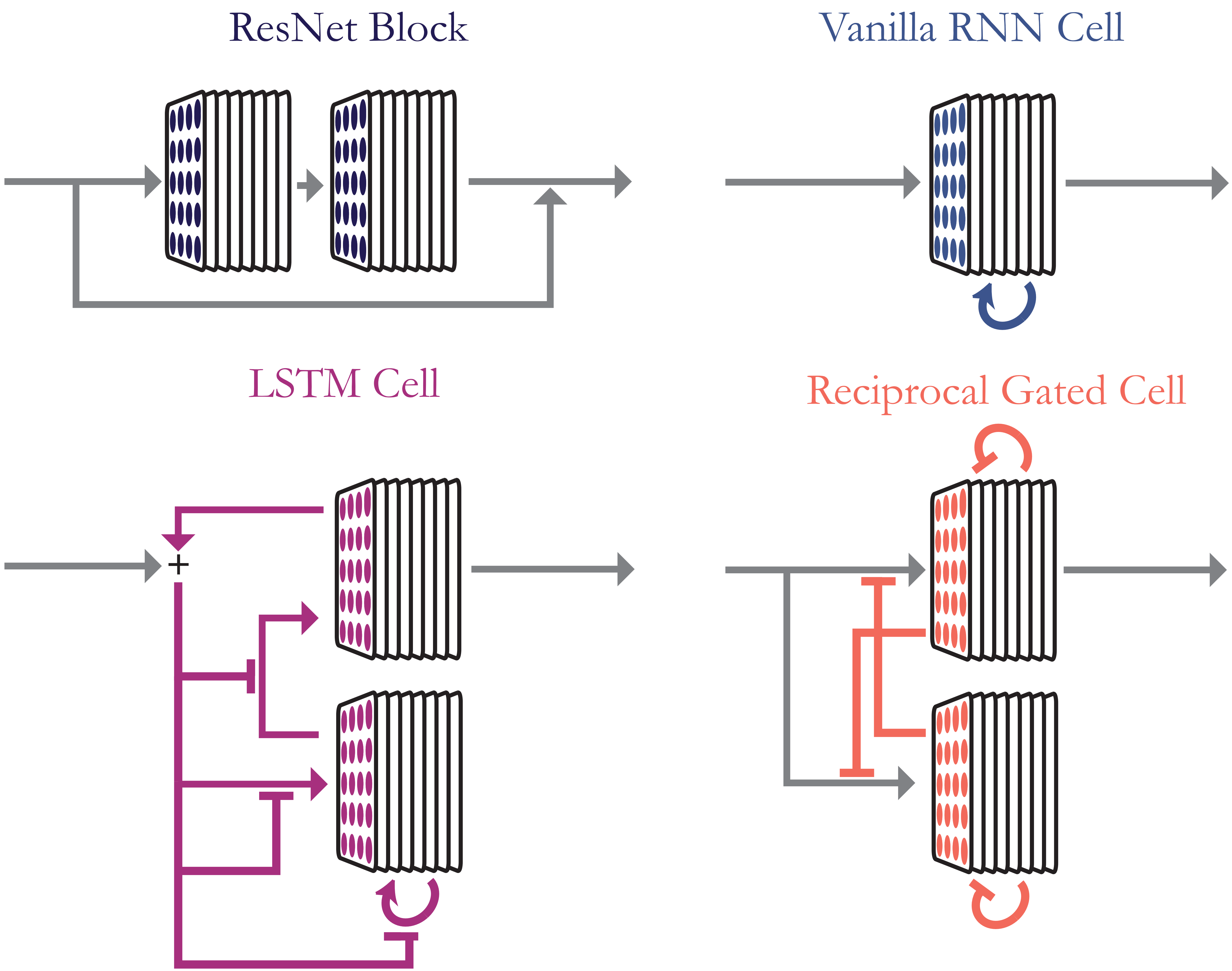


Improving ImageNet Performance with ConvRNNs



Each time-step (10 ms) is treated equally — including feedforward steps

Many Choices of Local Recurrence



Principles of Local Recurrence

Two complementary principles:

(1) gating = multiplication by input-dependent tensor w/ values in [0, 1]

(2) bypassing = when recurrent cell is in 0 state, input is unchanged ("performance preserving")

Principles of Local Recurrence

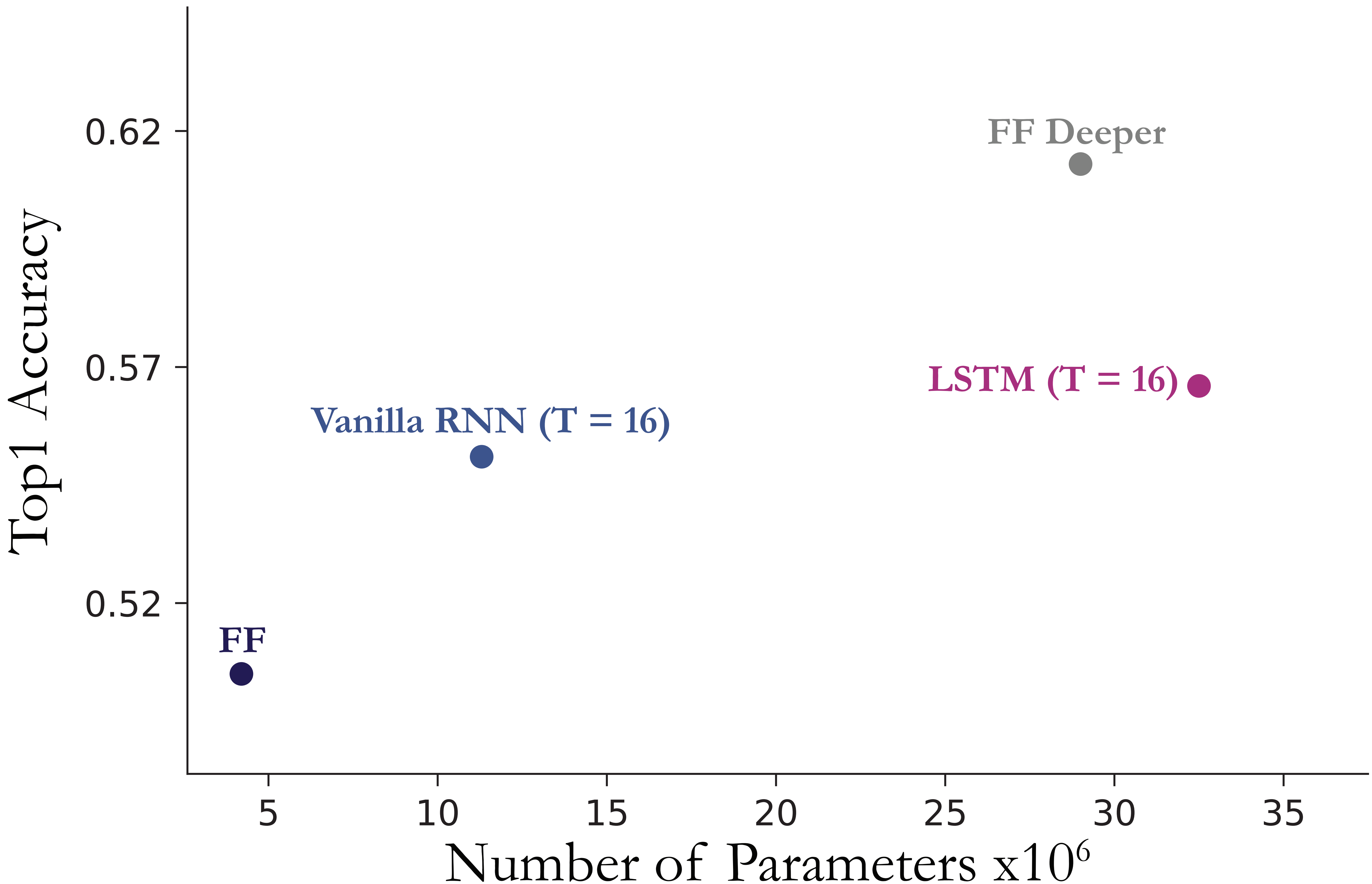
Two complementary principles:

(1) gating = multiplication by input-dependent tensor w/ values in [0, 1]

(2) bypassing = when recurrent cell is in 0 state, input is unchanged ("ResNet-like")

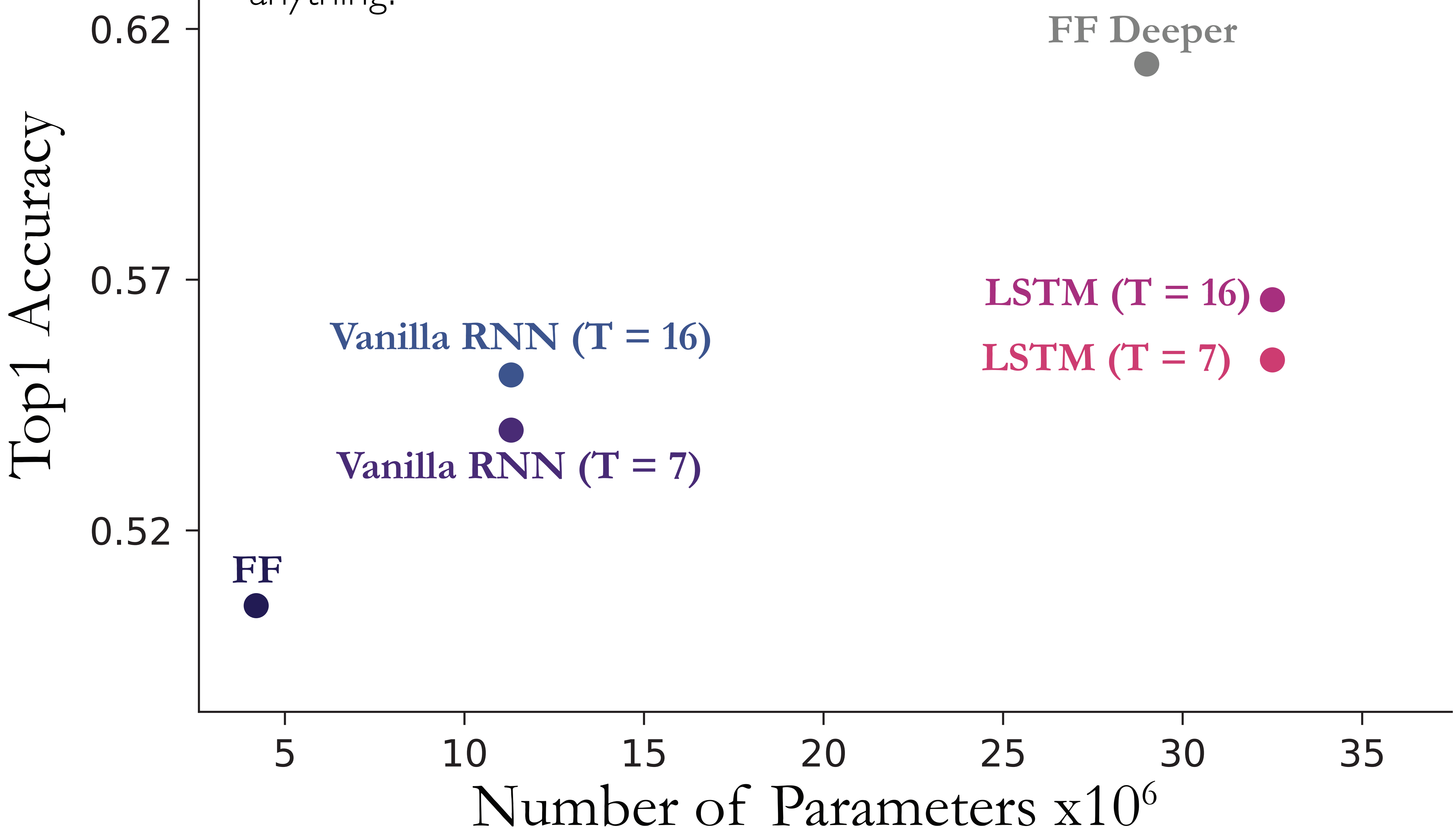
LSTM has **(1)** but not **(2)**; VanillaRNN has **(2)** but not **(1)**

Not All Local Recurrence is Equal

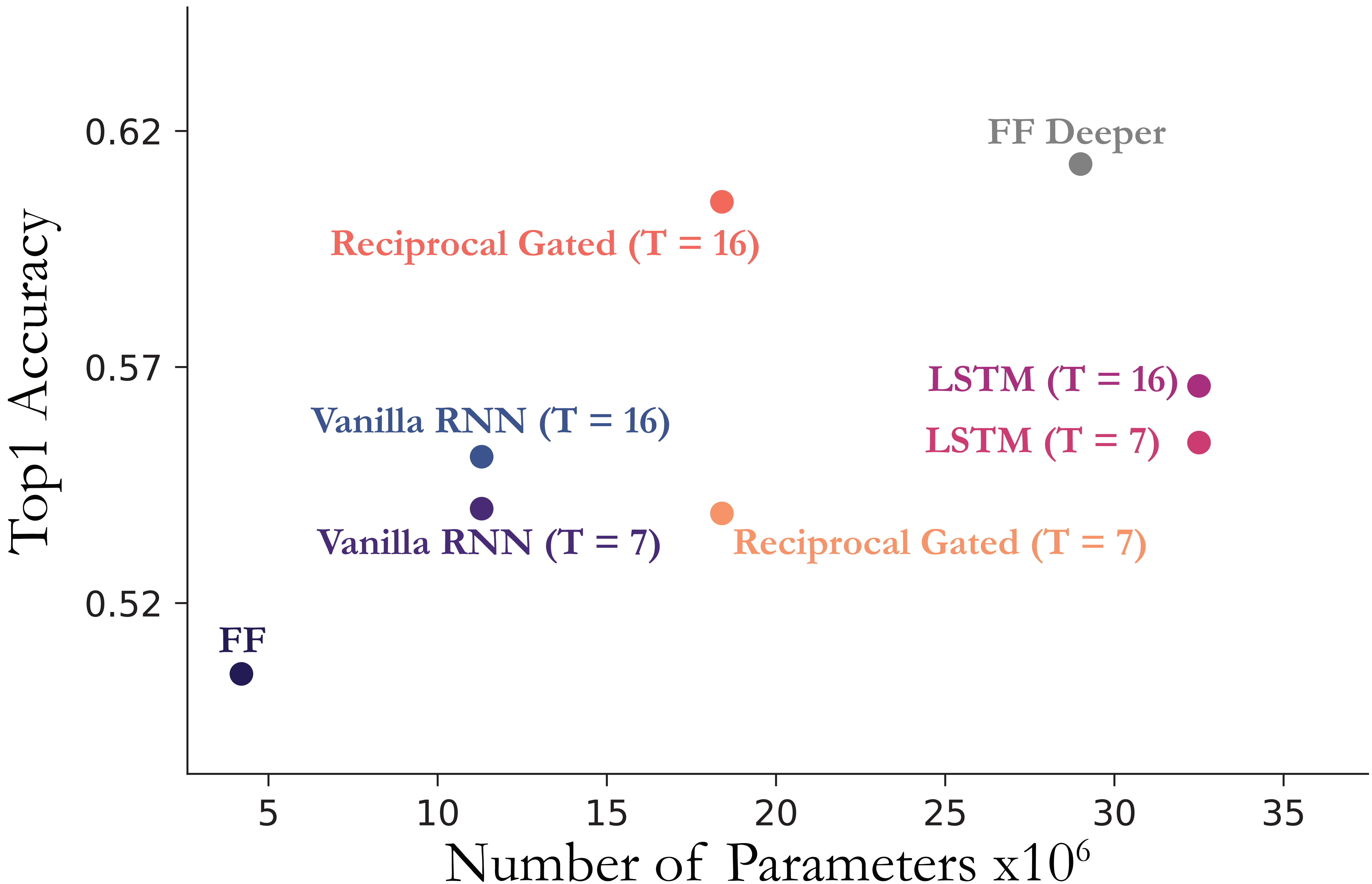


Not All Local Recurrence is Equal

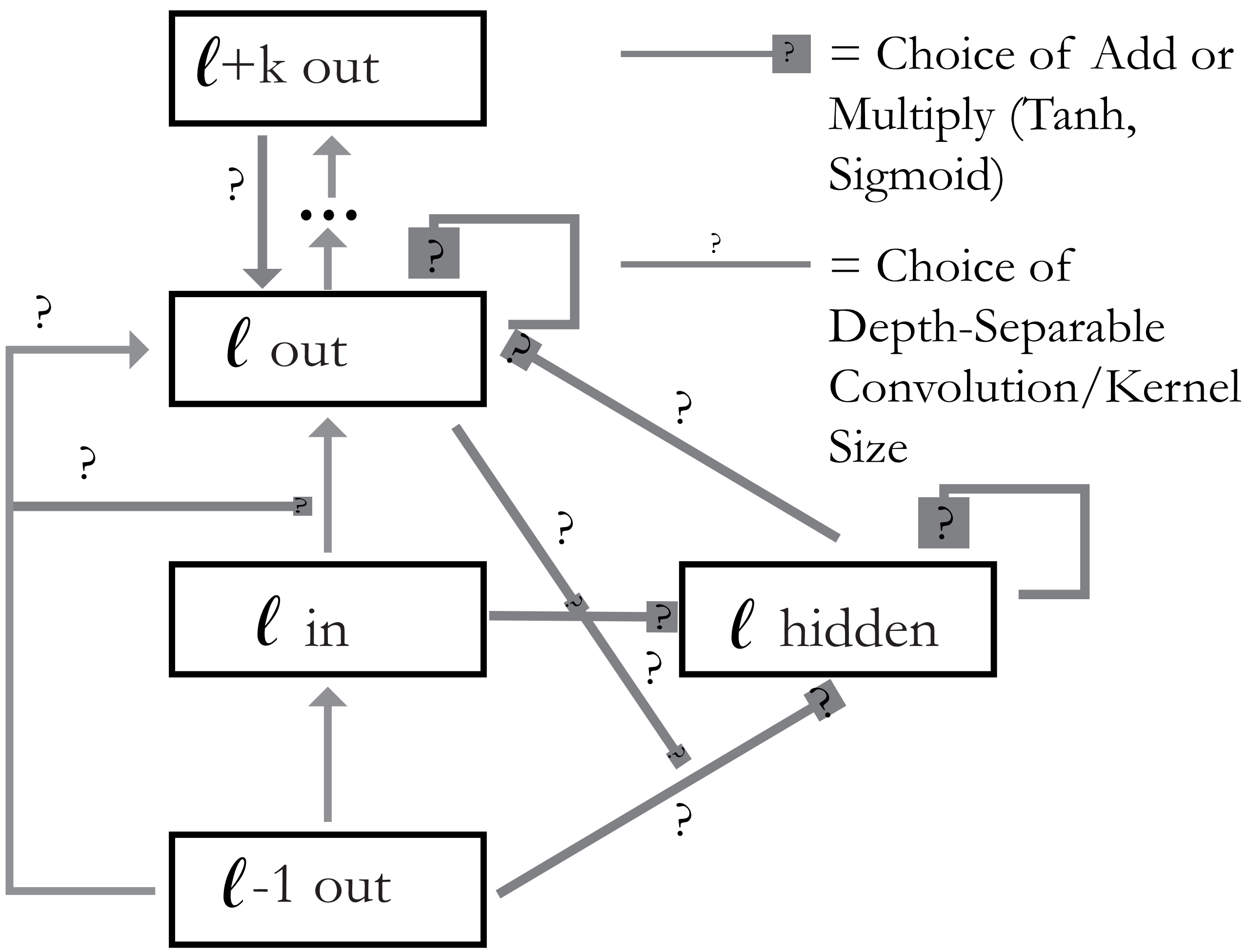
Control = shortest path through network using all units once gets at "is multi-interaction recurrence really adding anything?"



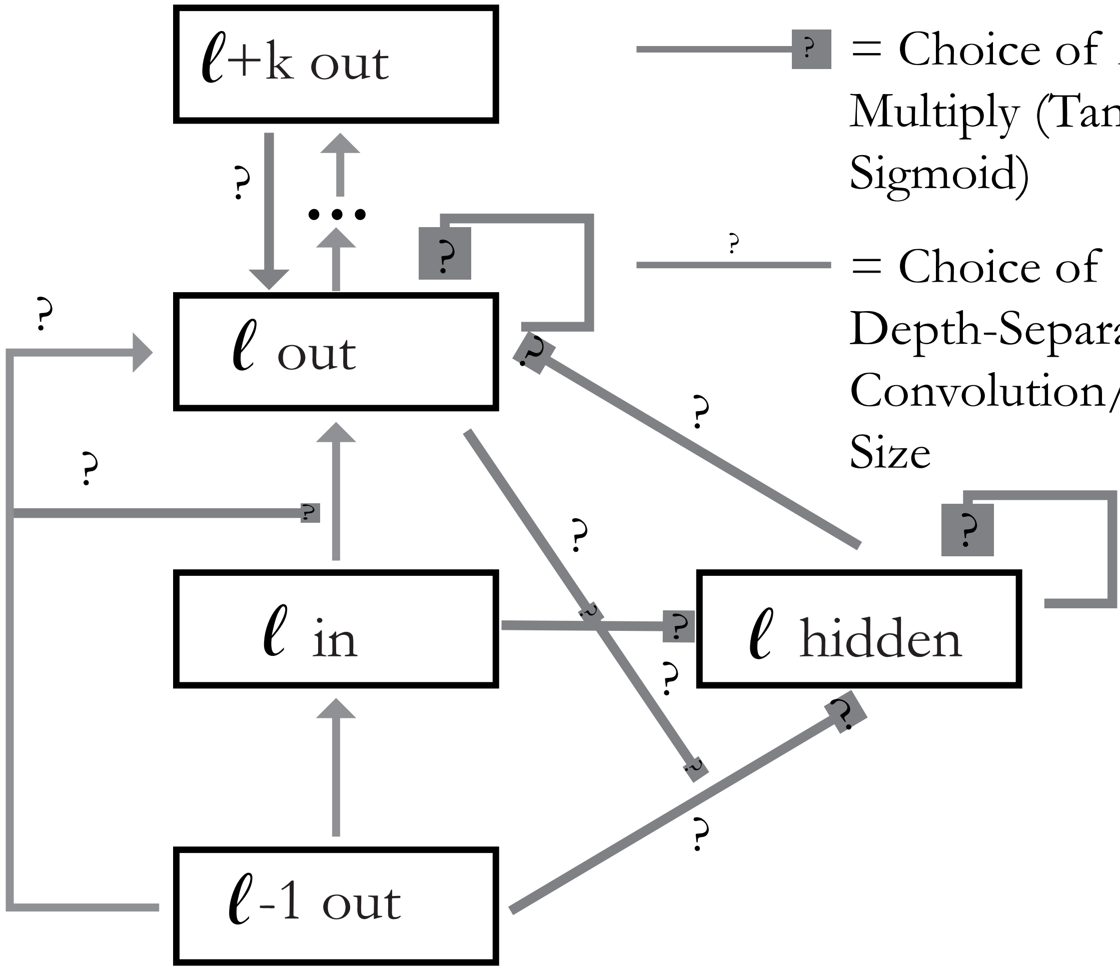
Not All Local Recurrence is Equal



Search Over Local and Global Recurrence

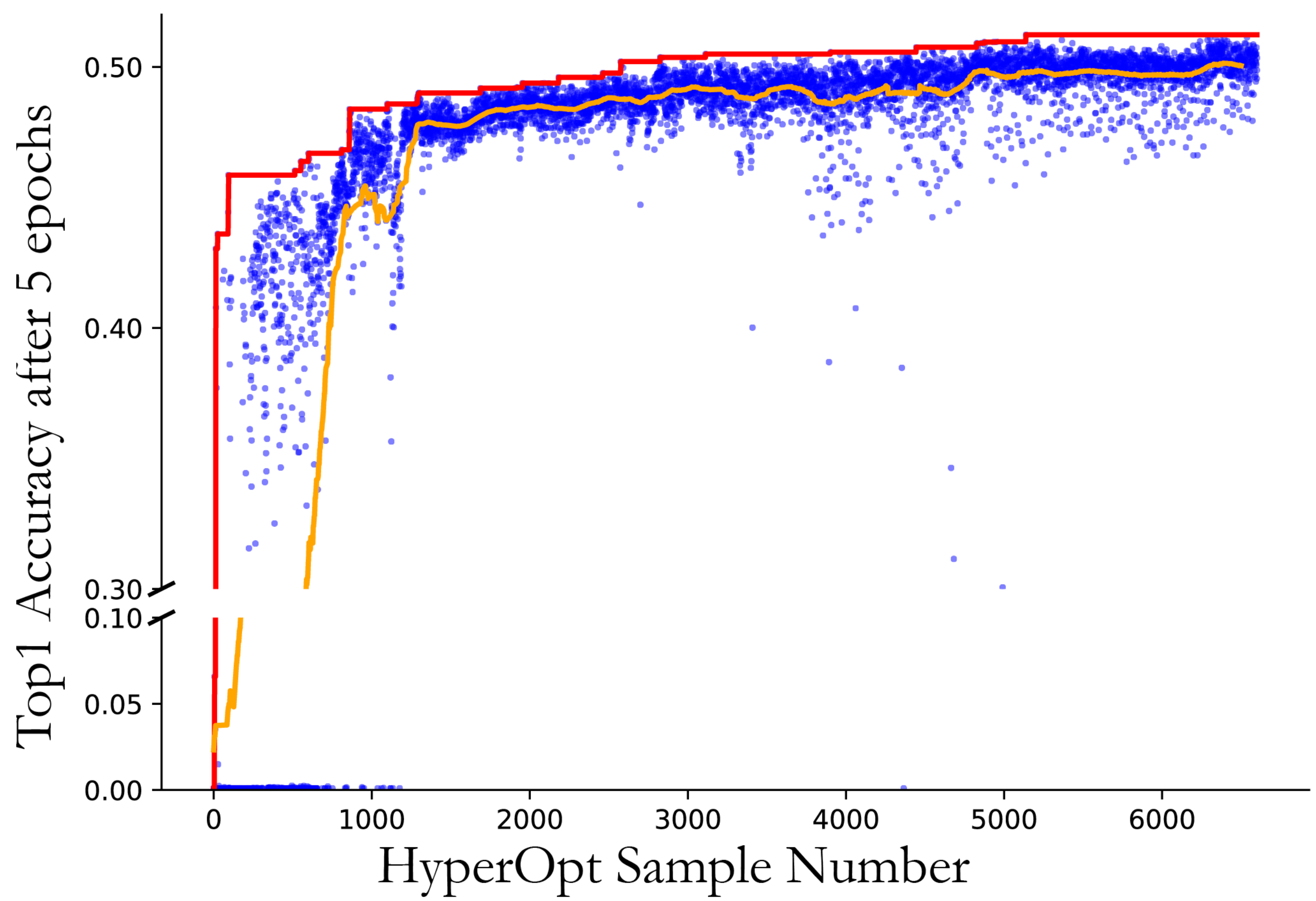


Search Over Local and Global Recurrence

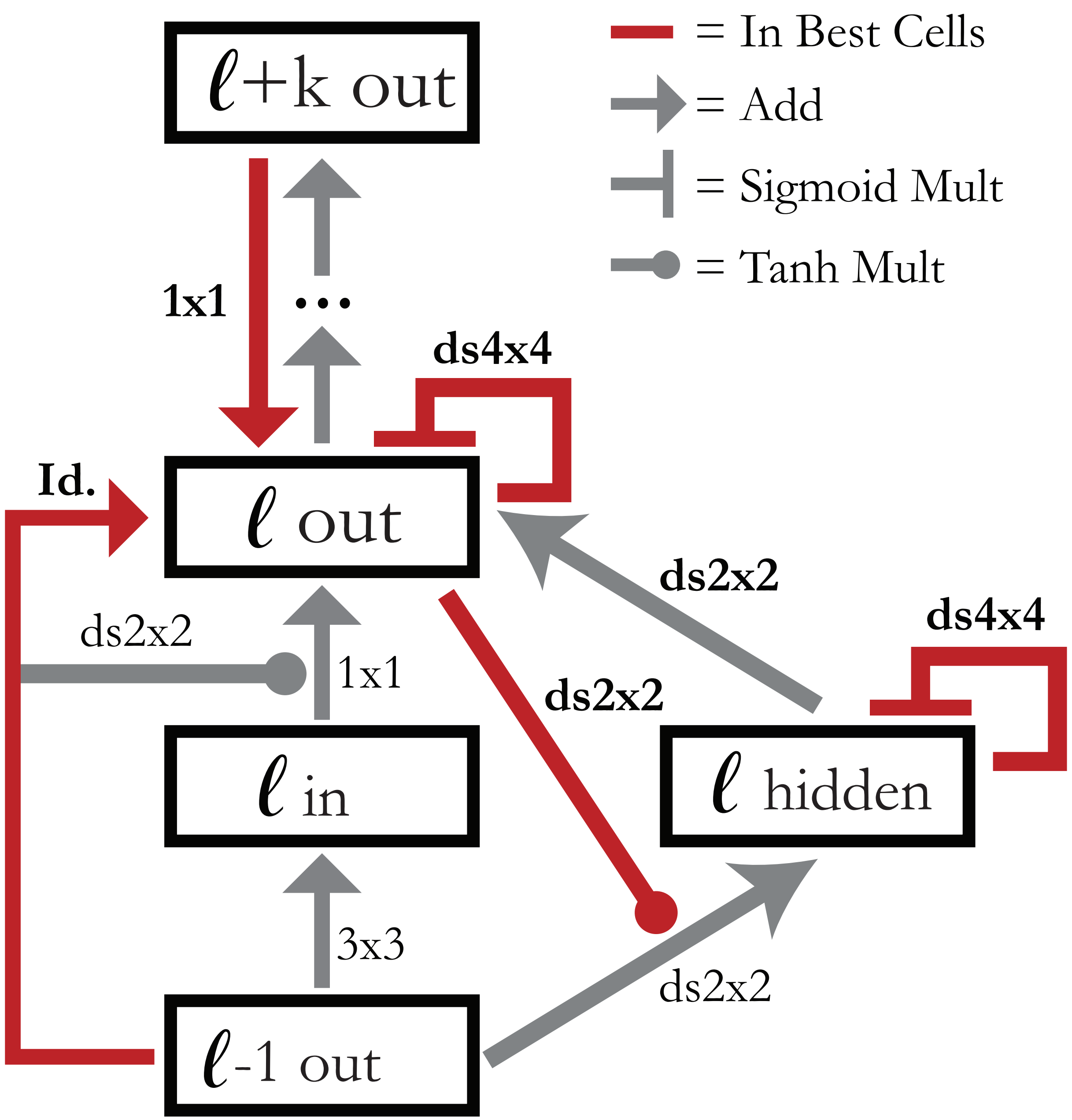


= Choice of Add or Multiply (Tanh, Sigmoid)

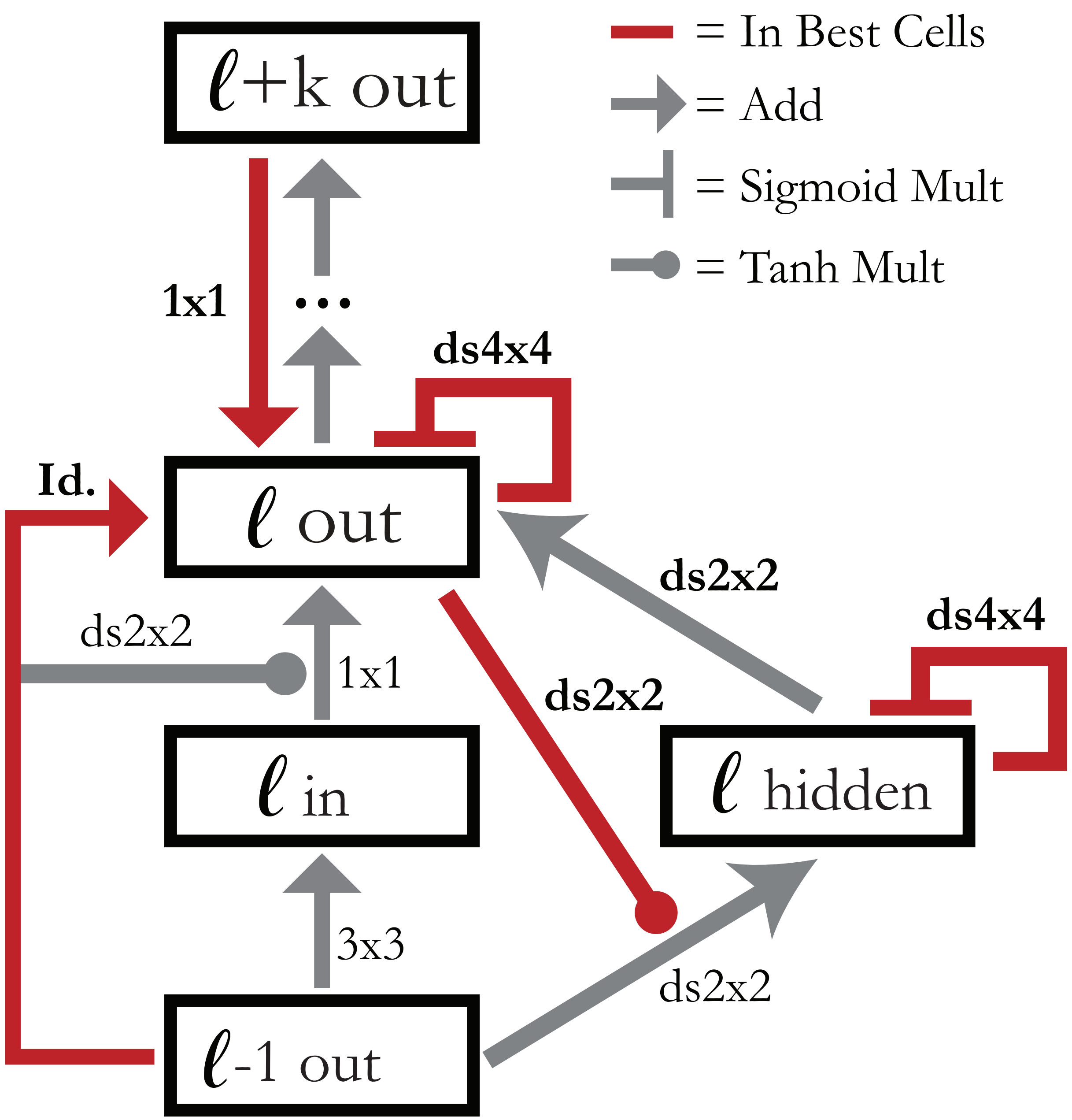
= Choice of Depth-Separable Convolution/Kernel Size



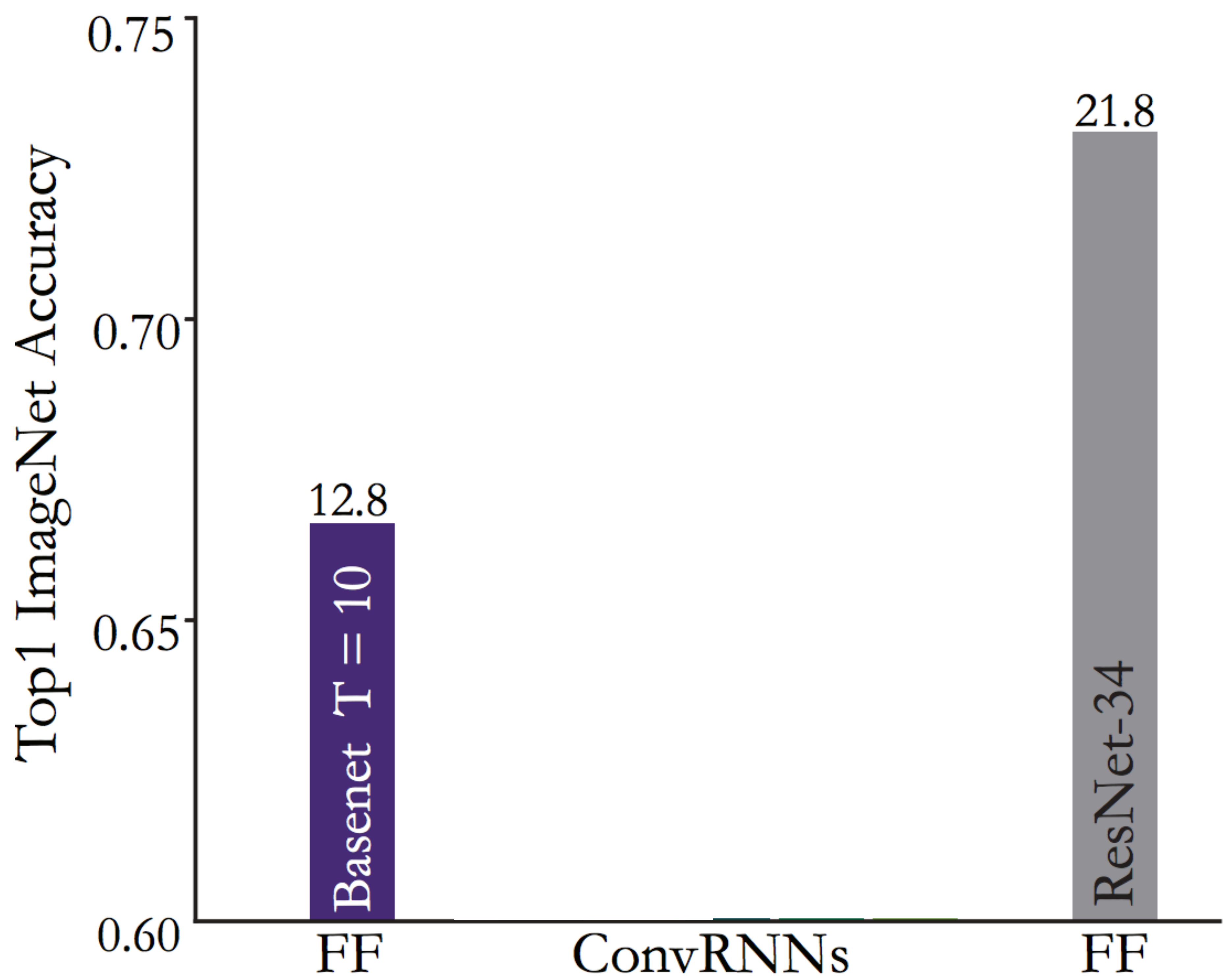
Emergent Local and Global Connectivity Patterns



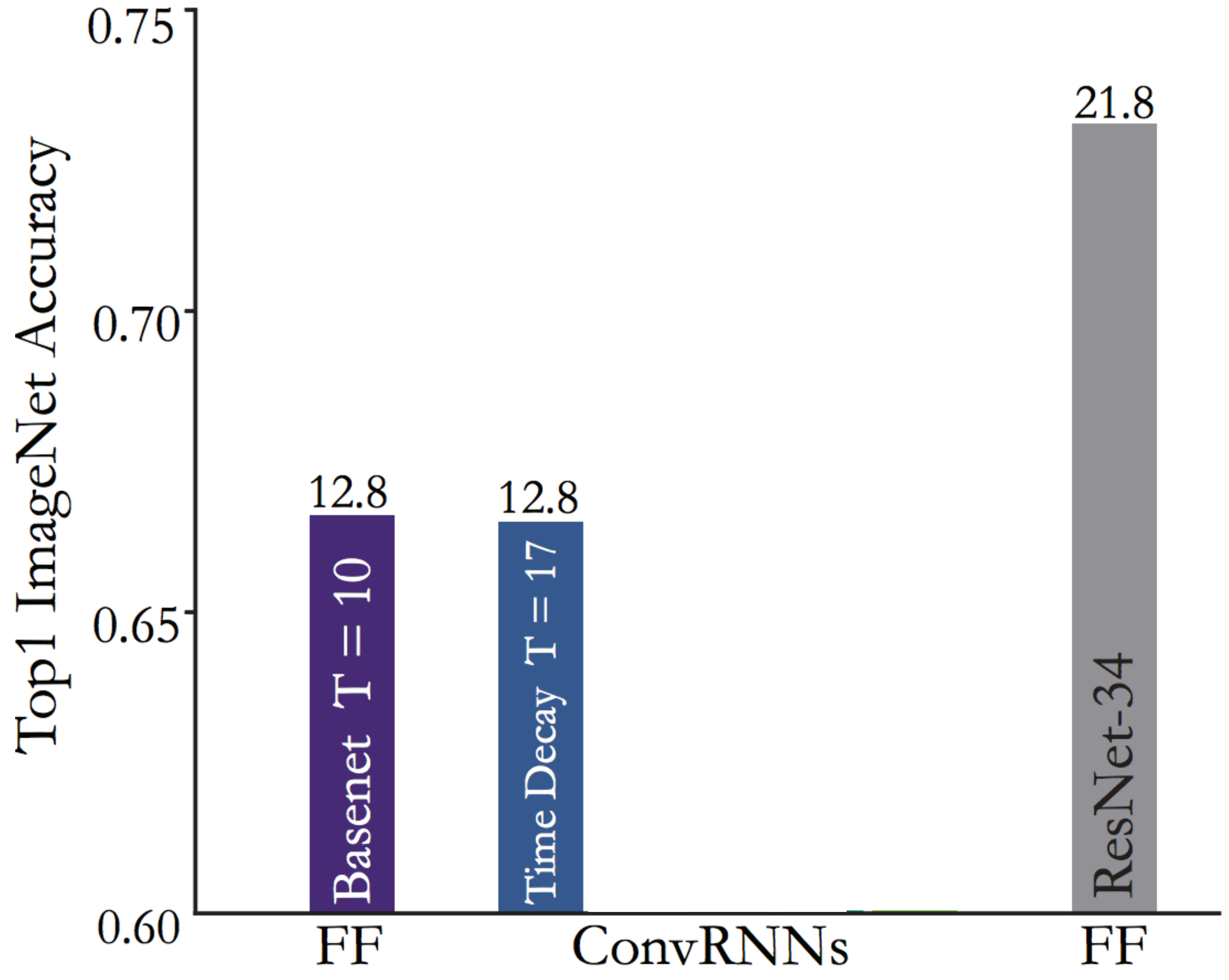
Emergent Local and Global Connectivity Patterns



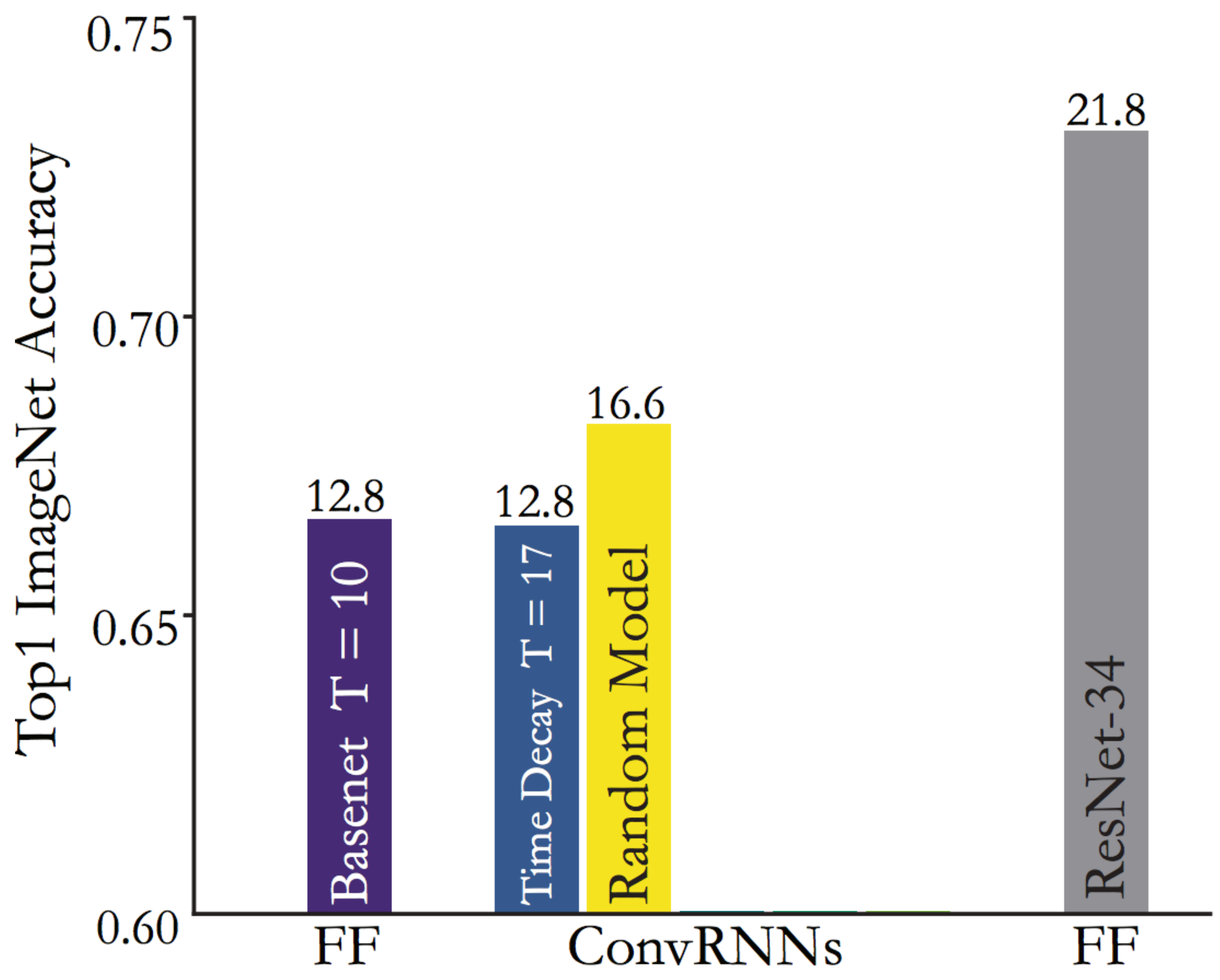
Improving ImageNet Performance with ConvRNNs



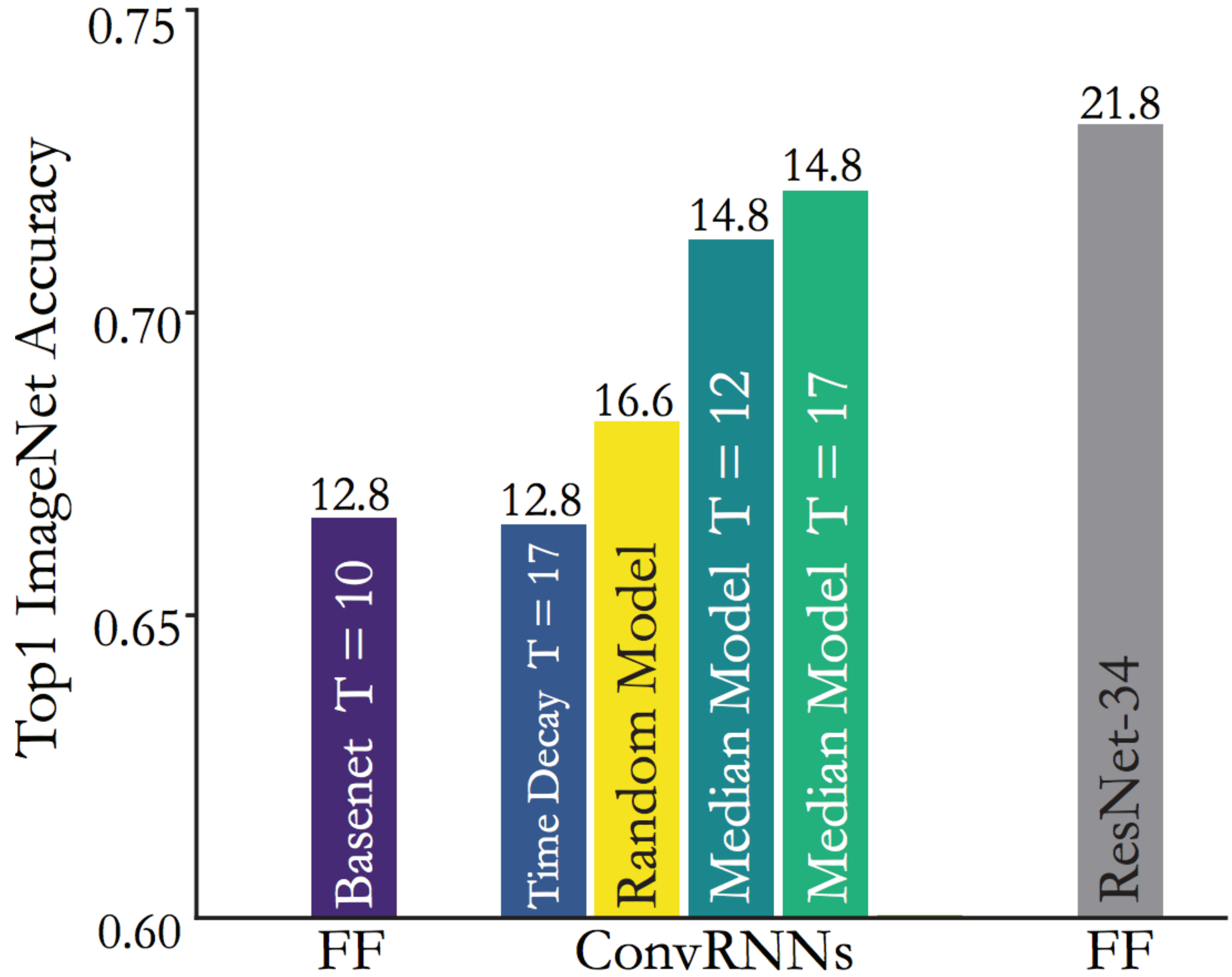
Improving ImageNet Performance with ConvRNNs



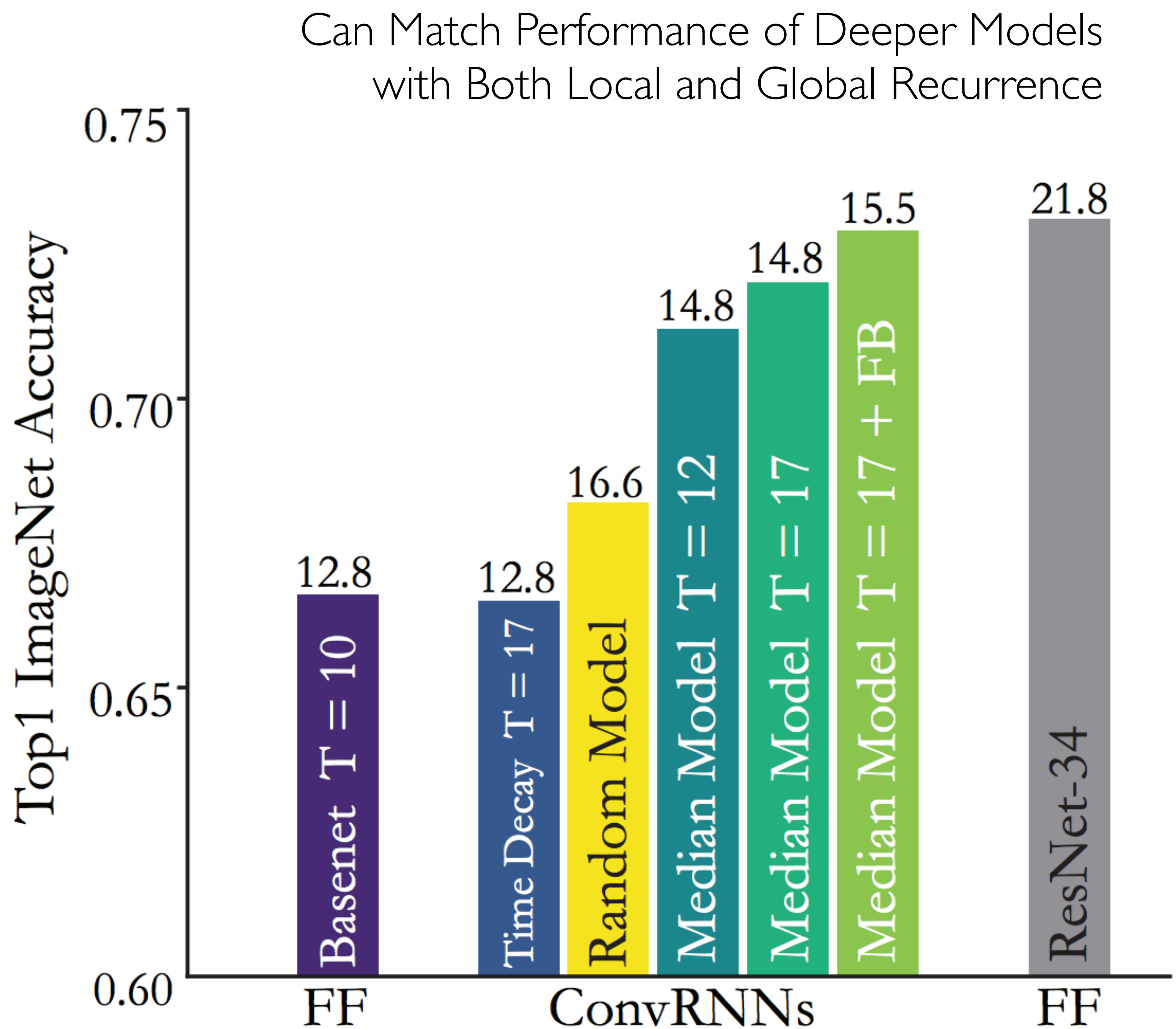
Improving ImageNet Performance with ConvRNNs



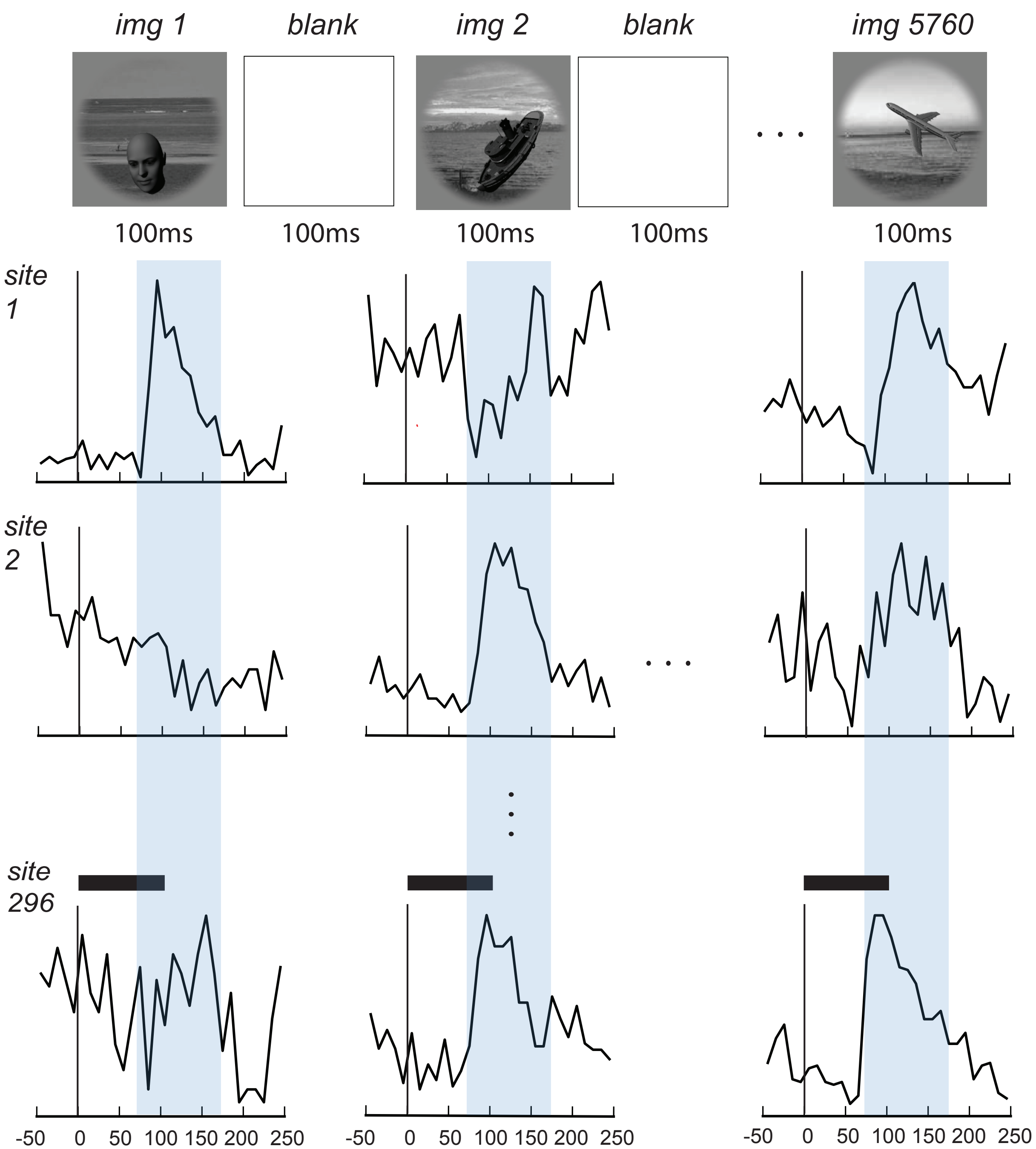
Improving ImageNet Performance with ConvRNNs



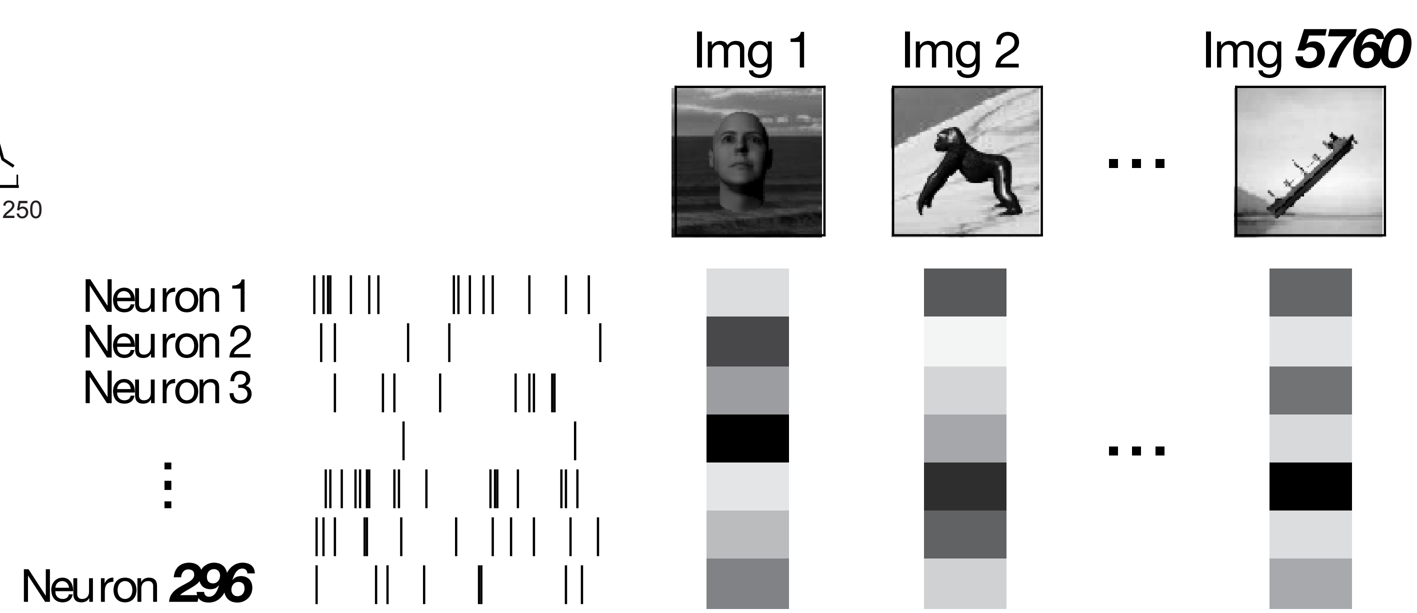
Improving ImageNet Performance with ConvRNNs



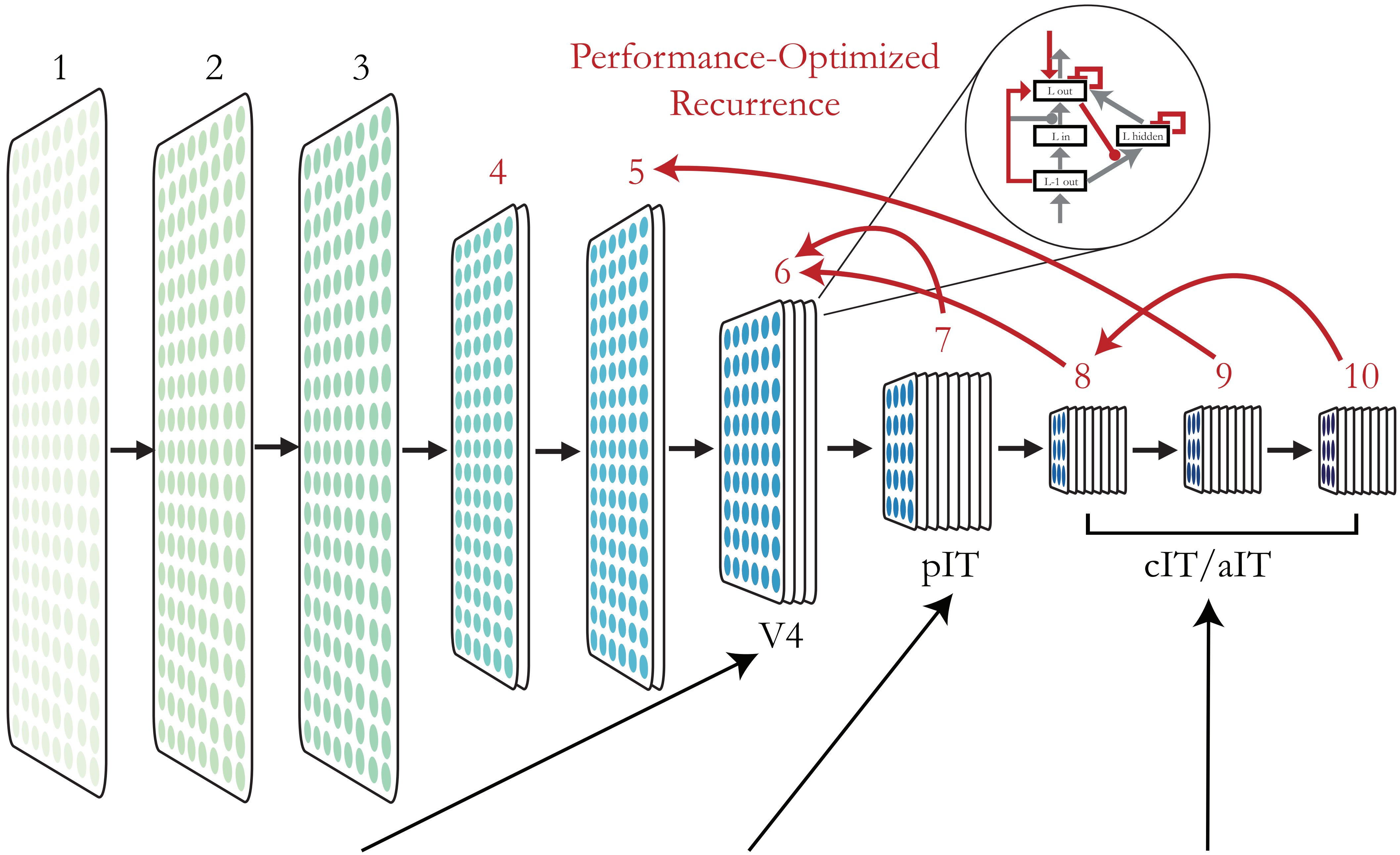
Neural Predictivity with ConvRNNs



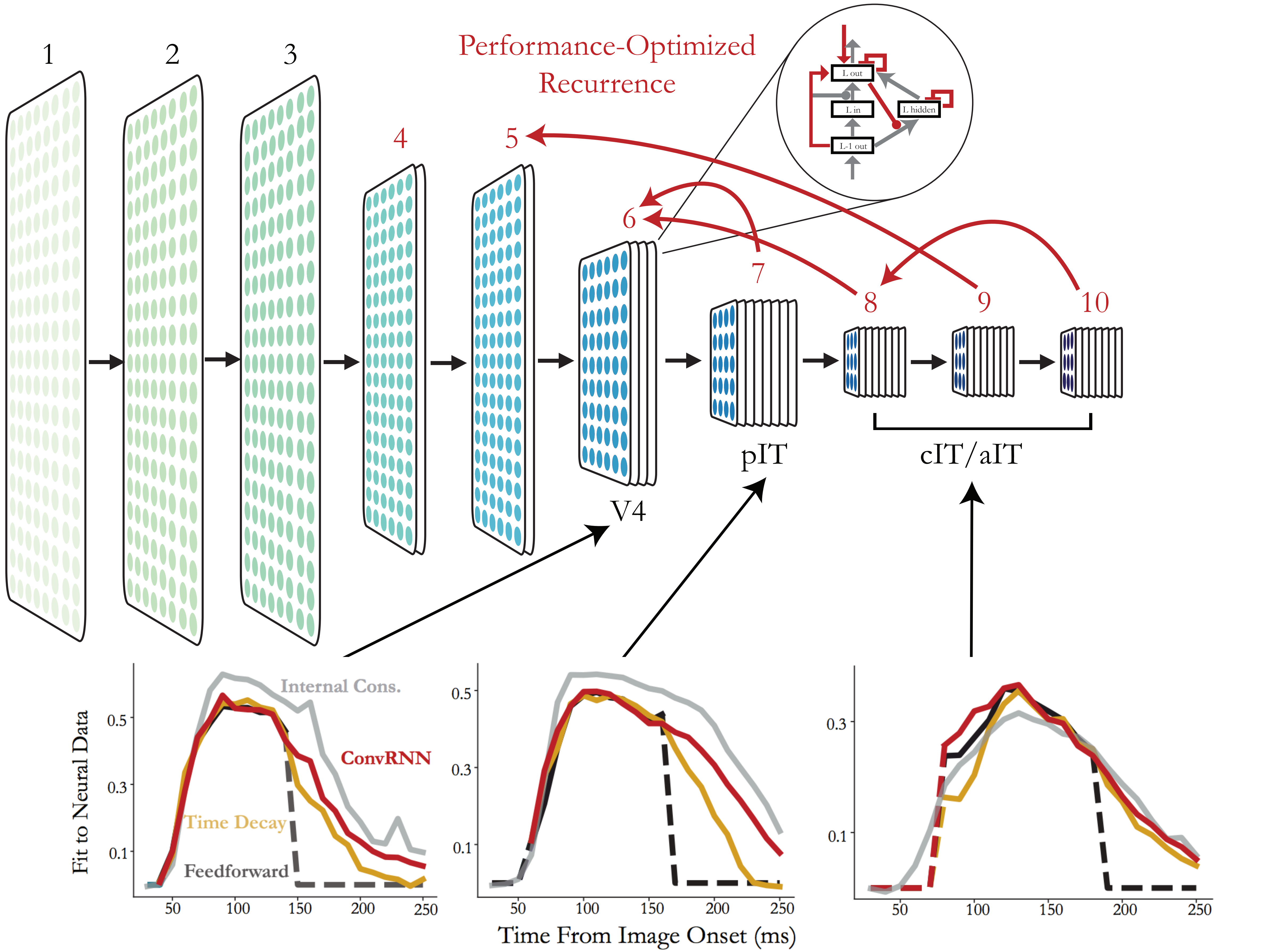
Binned spike counts 70ms-170ms post stimulus presentation



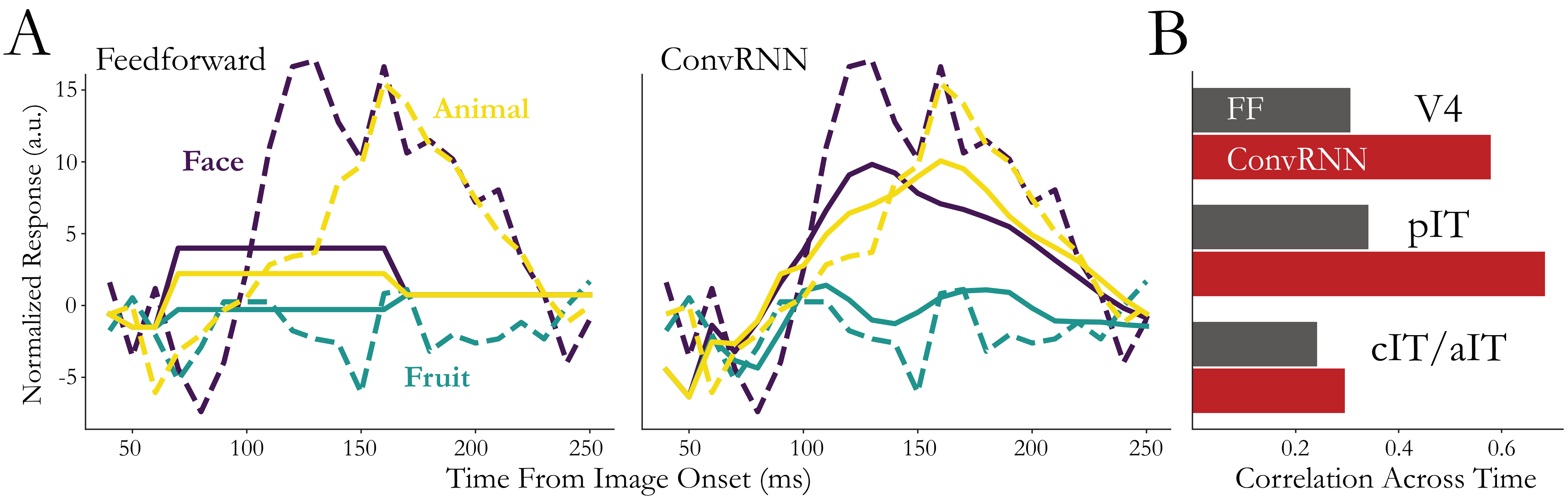
Improved Neural Fit with ConvRNNs



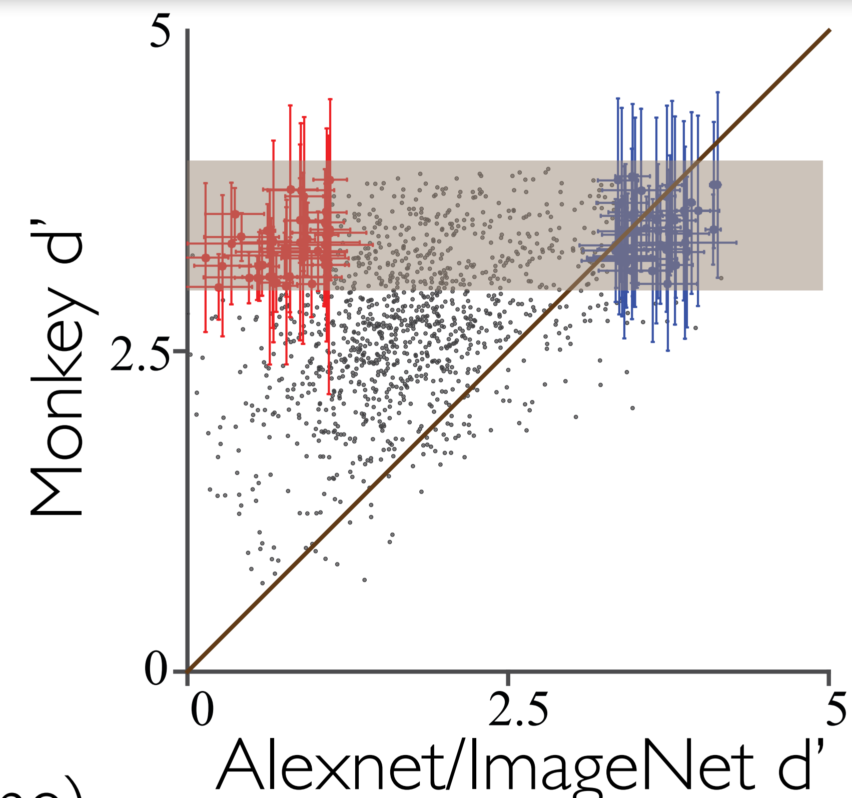
Improved Neural Fit with ConvRNNs



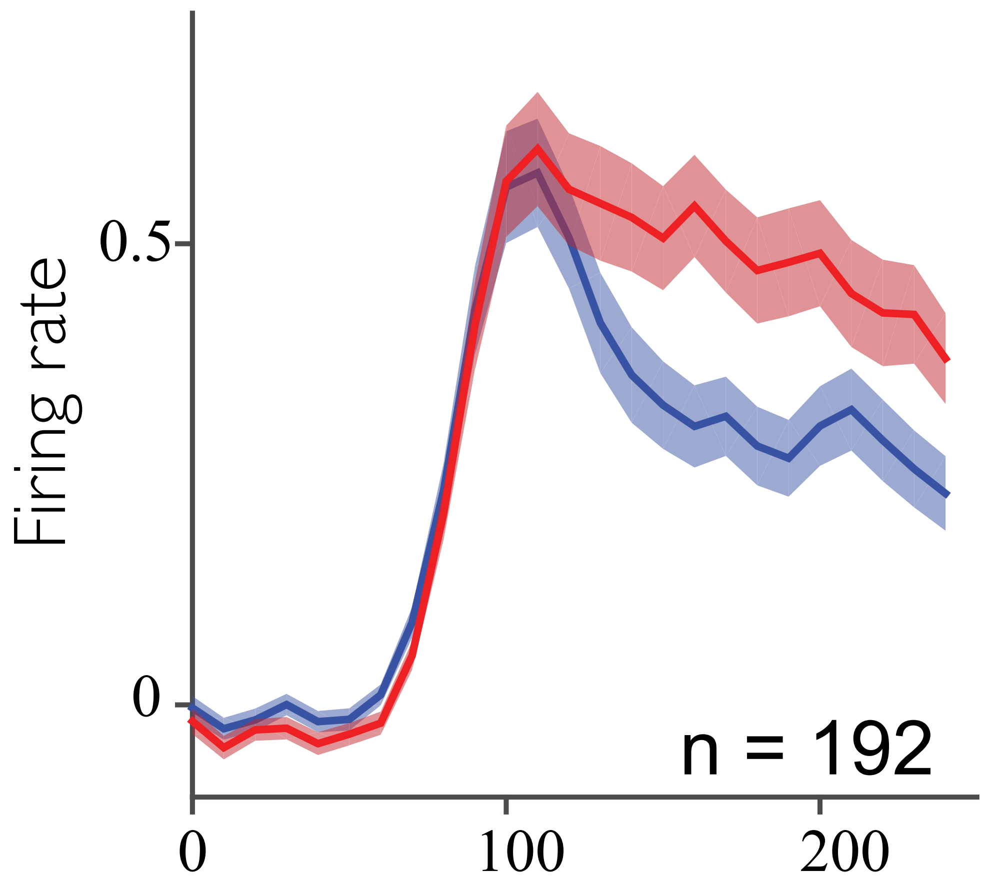
Improved Neural Fit with ConvRNNs

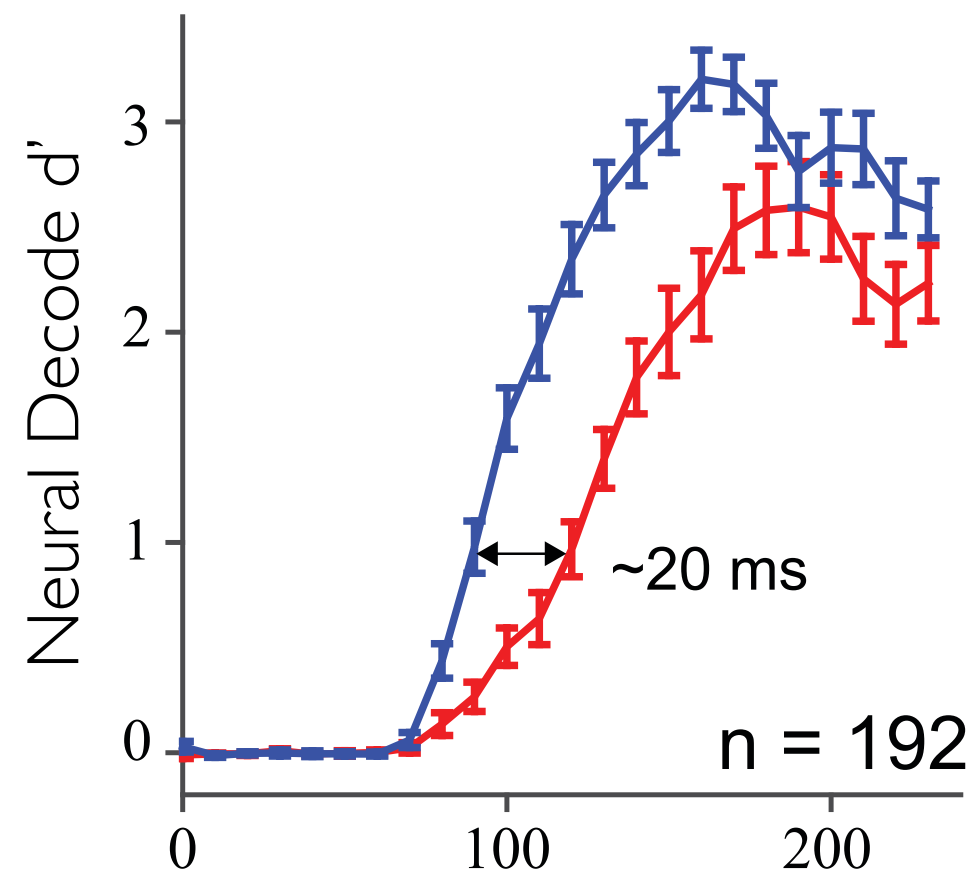


Behavioral Comparison



… even in range before 250ms (rough saccade time)

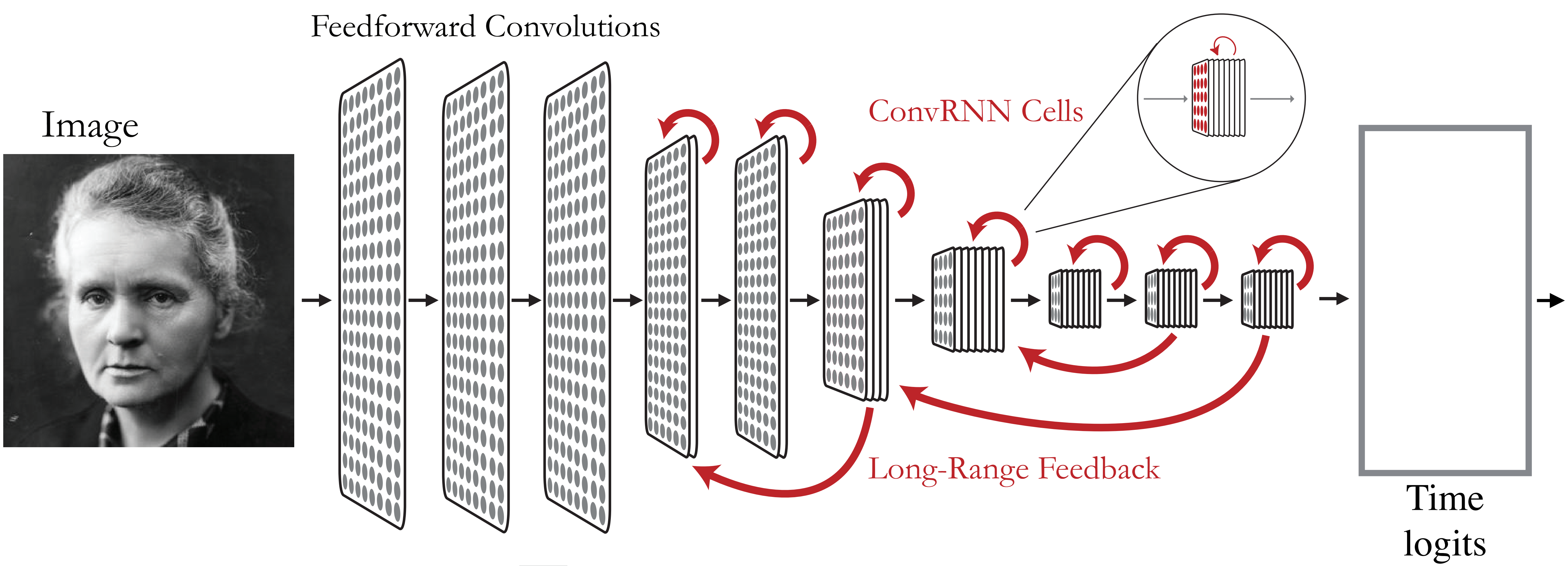


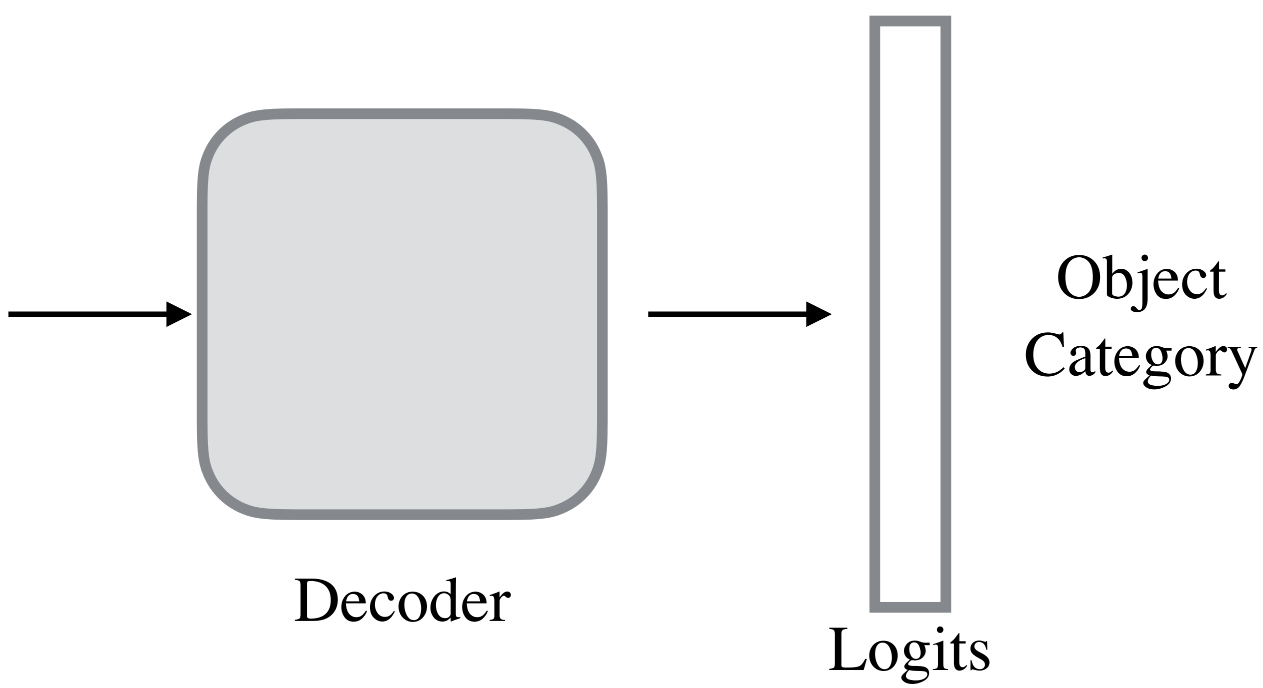


Time post-image onset (ms)

Kar et. al. (2018)

How do we model behavioral decoding?

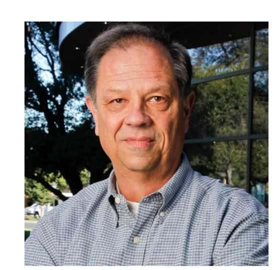




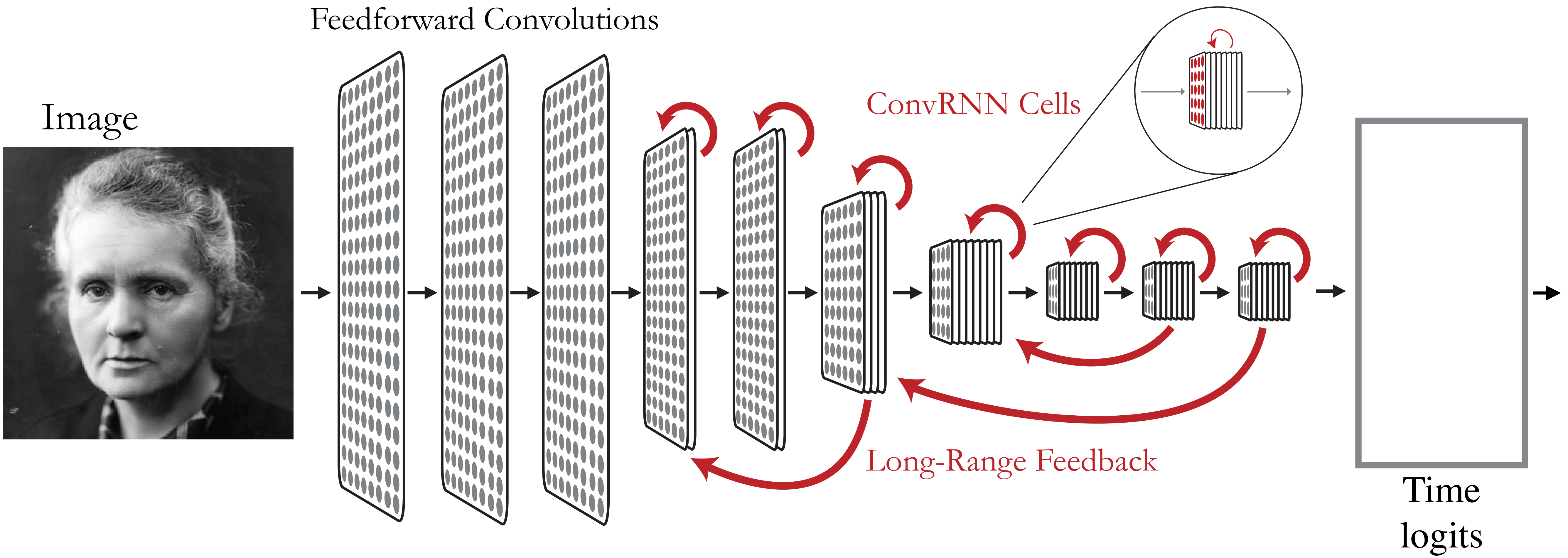
J Neurophysiol. 2001 Oct;86(4):1916-36.

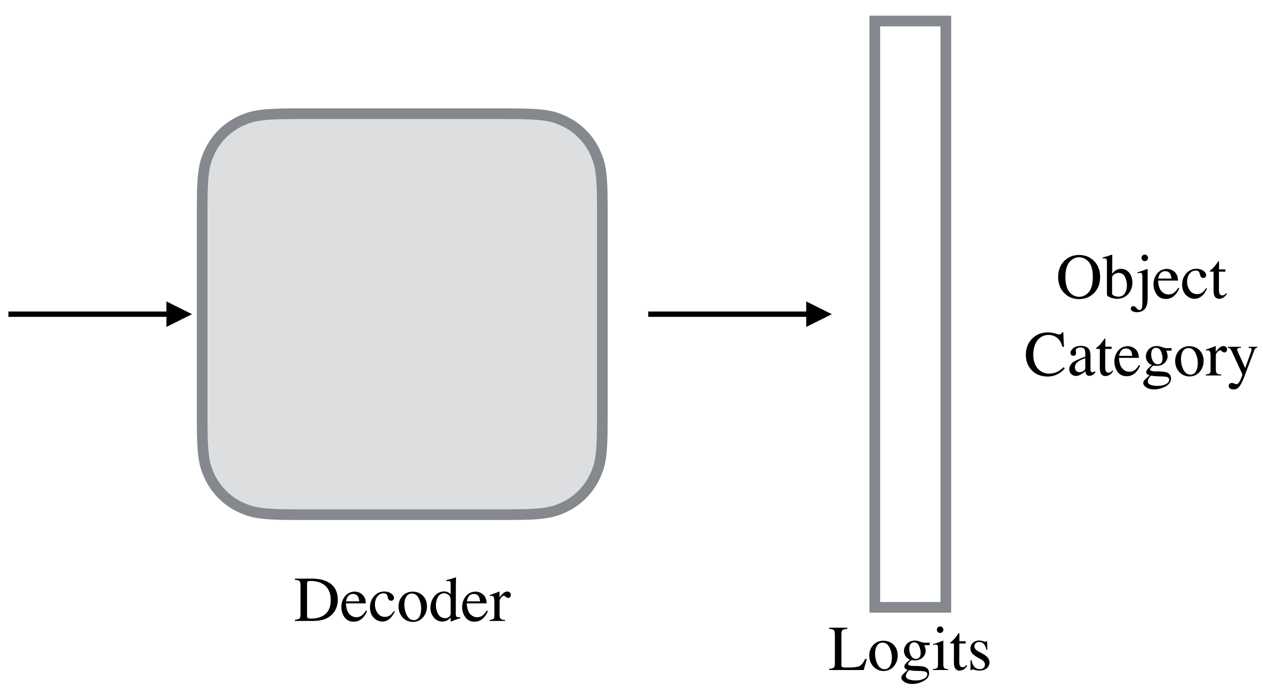
Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey.

Shadlen MN[1], Newsome WT.



How do we model behavioral decoding?





So how should we decode?

Simplest ideas for decoding

- **Final timestep:** simply use the logits vector from the last timestep.

- **Sum:** sum the logits across the time dimension.

- **Time average:** average the logits across the time dimension.

More ideas for decoding

- **Confidence weighting:** take a weighted average of the logits. Weights are proportional to the maximum confidence at each timepoint.

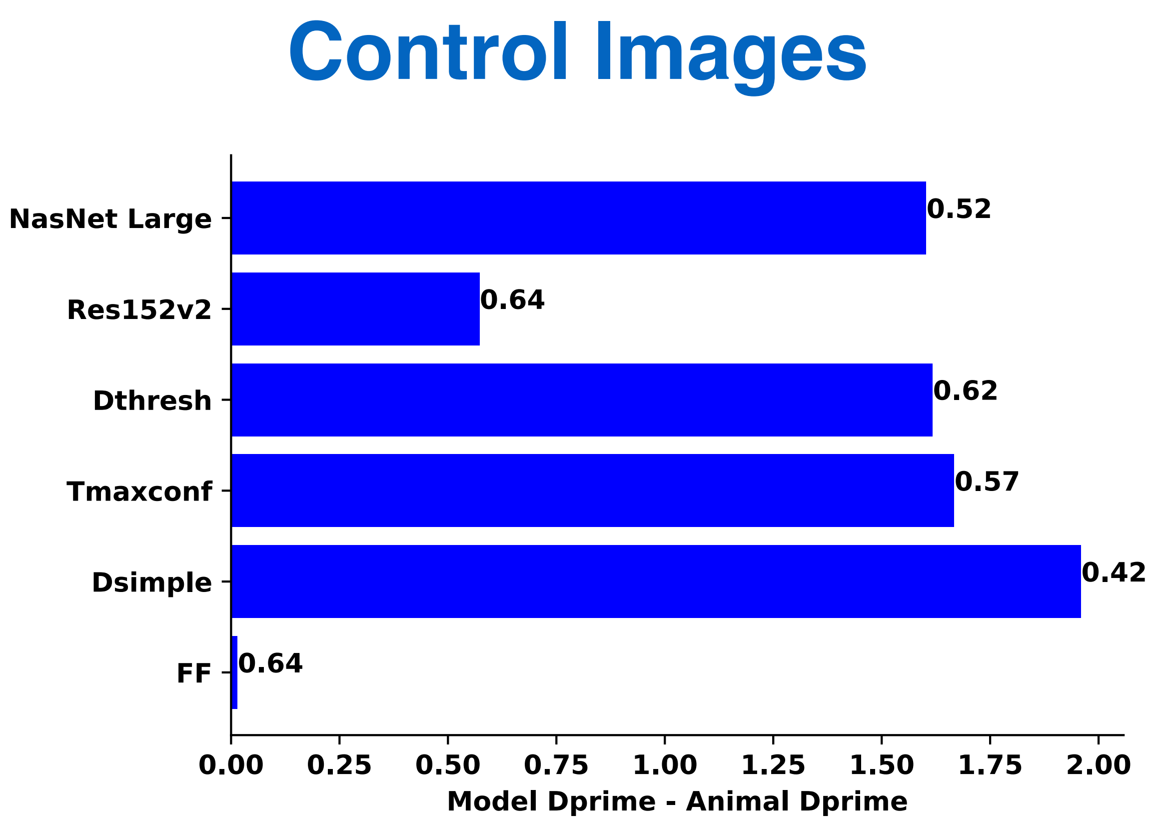
- **Weighted average variations:** allow for the weights to to take a more general form and make them trainable. Learn them end-to-end.

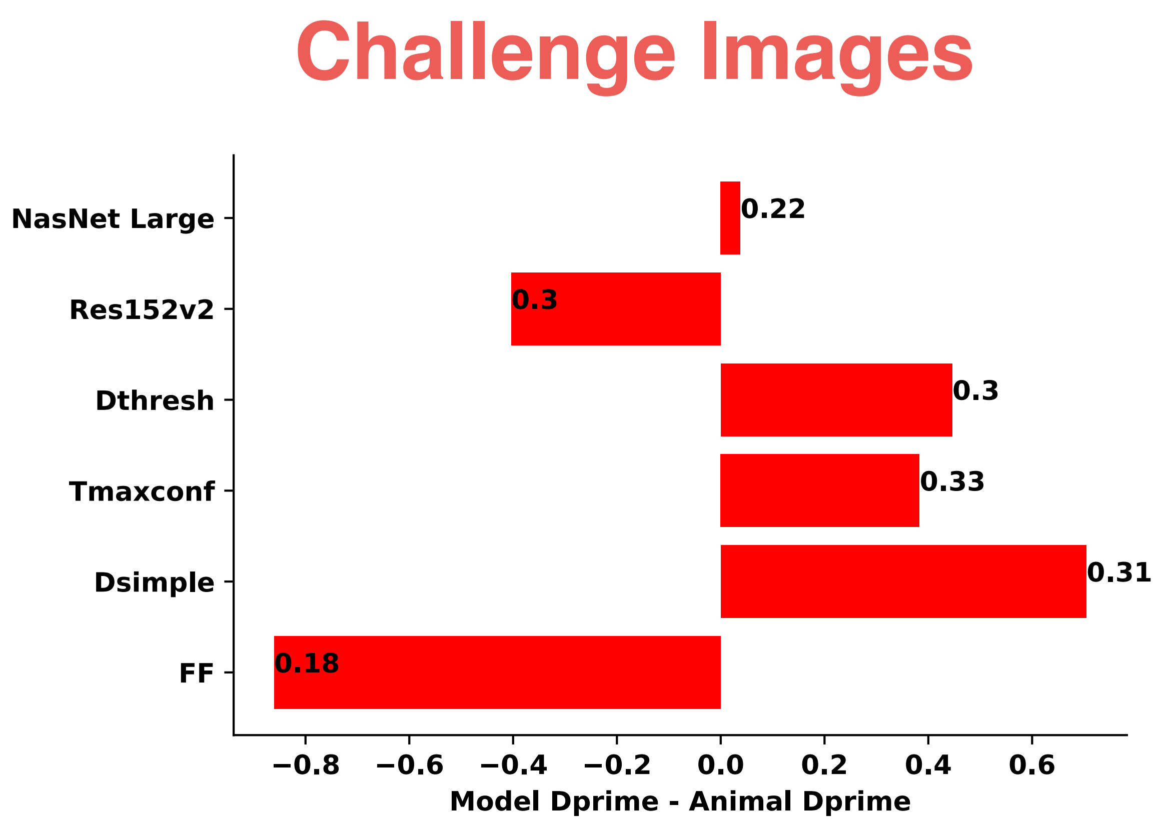
- **Maximum confidence:** select the logits from the timepoint which contains the most highest confidence value.

- **Confidence threshold:** select the logits from the first timepoint at which the maximum confidence crosses a given threshold. We can learn the threshold.

Javier Sagastuy

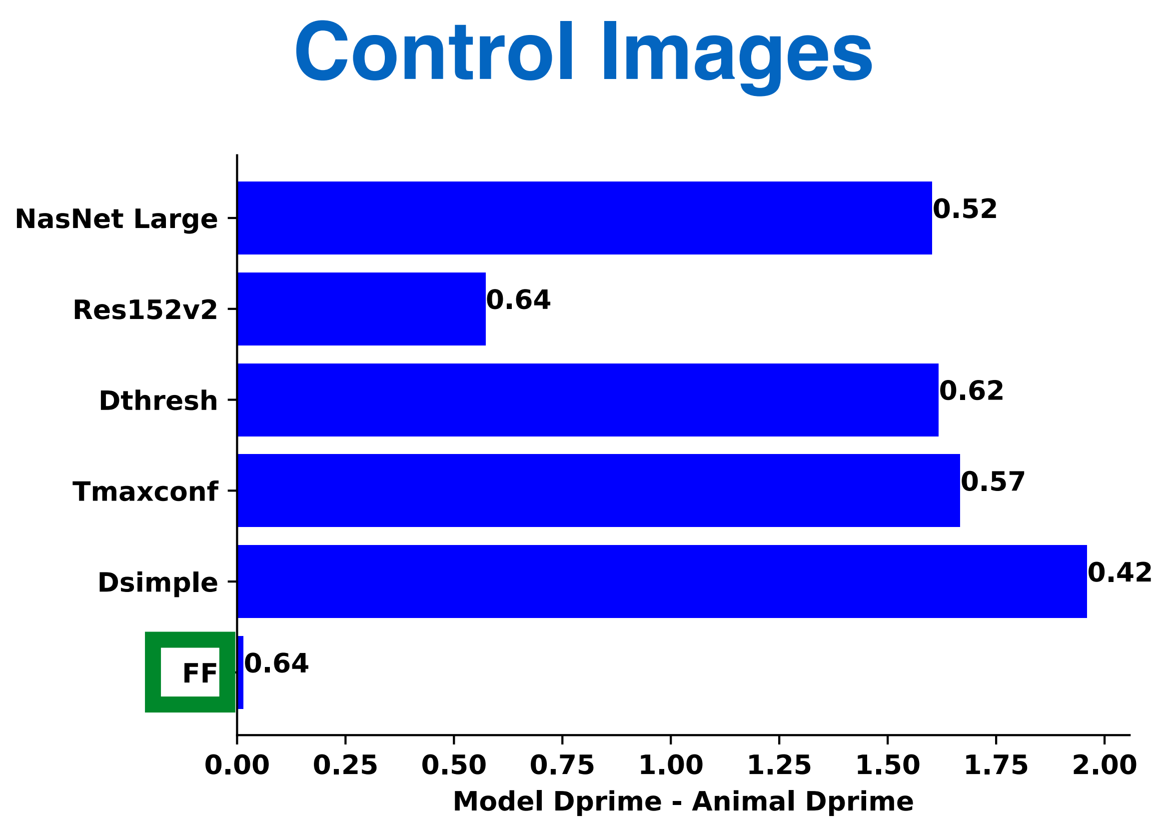
Recurrent models correlate better with animal "challenge" dprimes

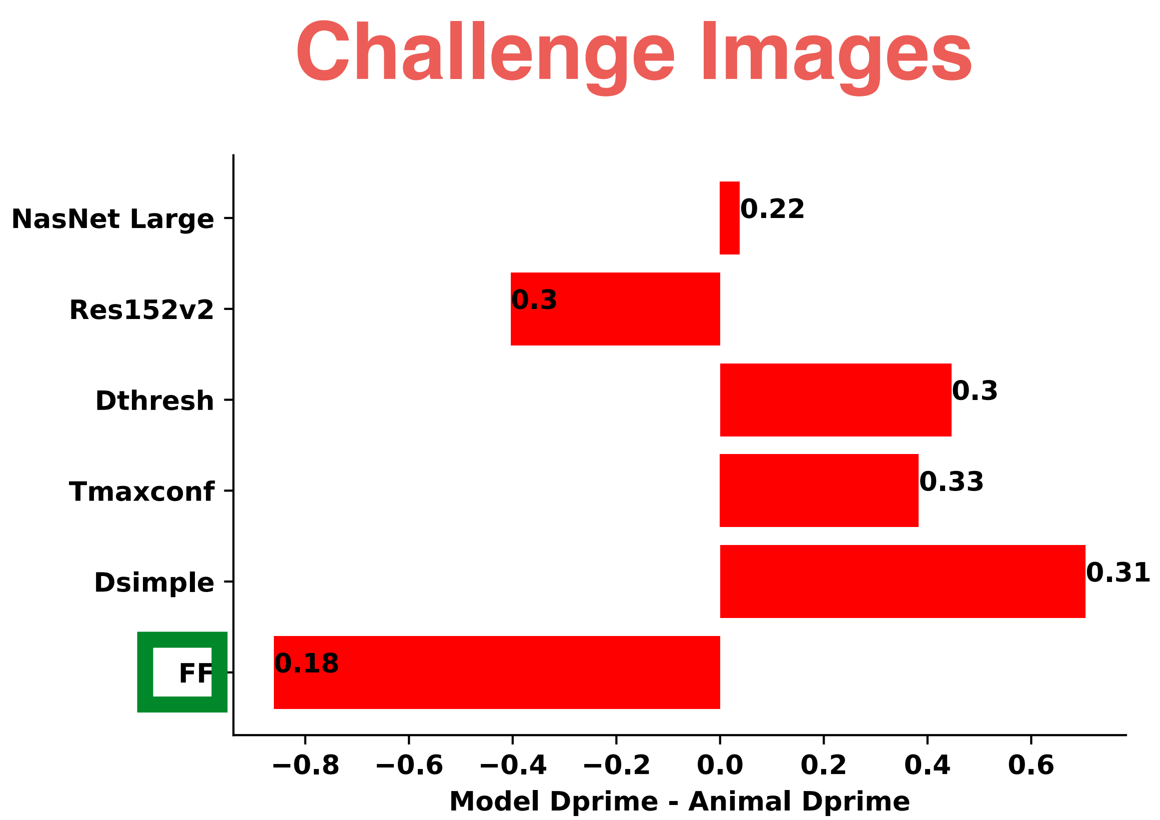




Recurrent models correlate better with animal "challenge" dprimes

Feedforward Base Model

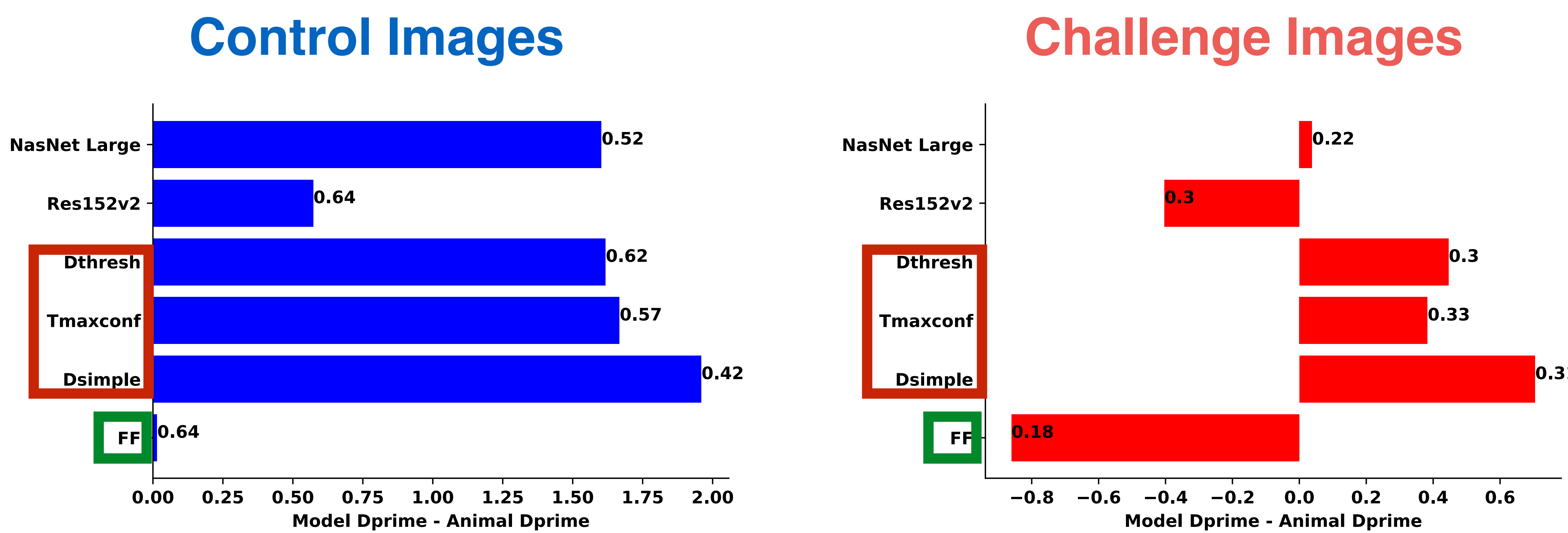




Recurrent models correlate better with animal "challenge" dprimes

Feedforward Base Model

ConvRNN + Decoder

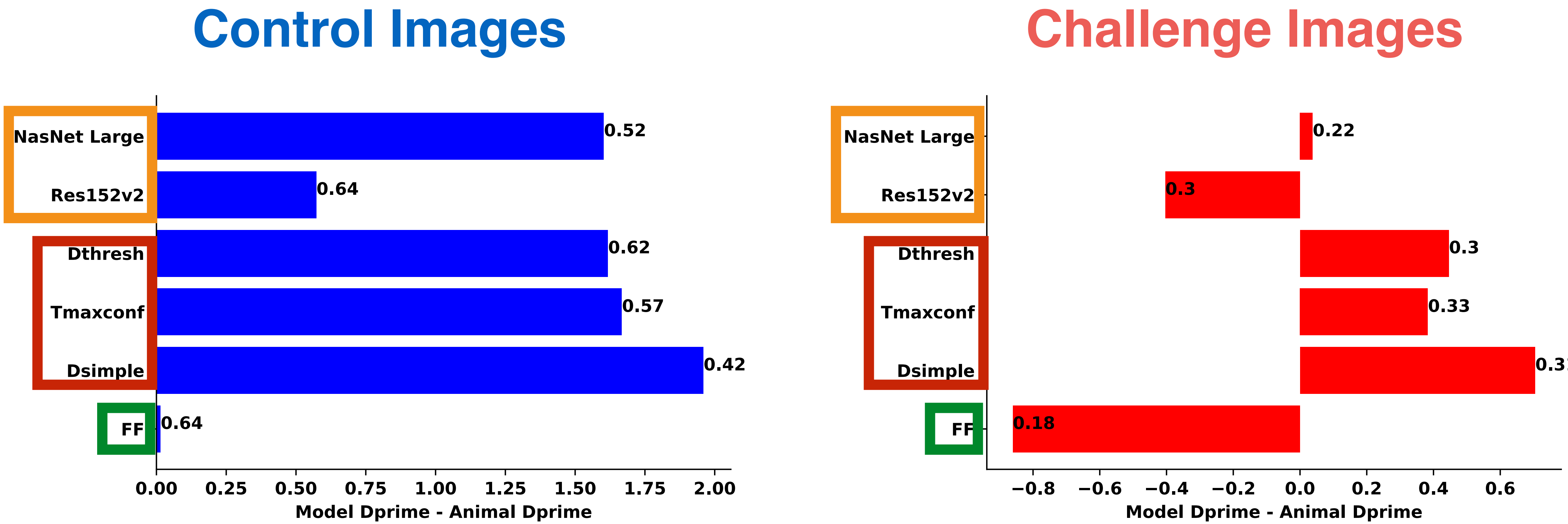


Recurrent models correlate better with animal “challenge” dprimes

Feedforward Base Model

ConvRNN + Decoder

Deeper Feedforward Models



Takeaways

AI/Performance **Neuroscience**

Takeaways

AI/Performance **Neuroscience**

- At scale, one has to be careful with the choice of local recurrent architecture introduced into CNNs to improve performance

Takeaways **AI/Performance** **Neuroscience**

- At scale, one has to be careful with the choice of local recurrent architecture introduced into CNNs to improve performance
  - *With proper local recurrence in place, specific patterns of long-range feedback connections further improve performance*

Takeaways

AI/Performance **Neuroscience**

- At scale, one has to be careful with the choice of local recurrent architecture introduced into CNNs to improve performance
  - *With proper local recurrence in place, specific patterns of long-range feedback connections further improve performance*

- These performance-optimized dynamics provide strong estimates of neural and behavioral benchmarks in the primate ventral stream over feedforward & other recurrent models

Takeaways **AI/Performance** **Neuroscience**

- At scale, one has to be careful with the choice of local recurrent architecture introduced into CNNs to improve performance
  - *With proper local recurrence in place, specific patterns of long-range feedback connections further improve performance*

- These performance-optimized dynamics provide strong estimates of neural and behavioral benchmarks in the primate ventral stream over feedforward & other recurrent models

- These new convolutional recurrent architectures can be applied to many computer vision tasks (segmentation, movie prediction) without much modification

Takeaways **AI/Performance** **Neuroscience**

- At scale, one has to be careful with the choice of local recurrent architecture introduced into CNNs to improve performance
  - *With proper local recurrence in place, specific patterns of long-range feedback connections further improve performance*

- These performance-optimized dynamics provide strong estimates of neural and behavioral benchmarks in the primate ventral stream over feedforward & other recurrent models

- These new convolutional recurrent architectures can be applied to many computer vision tasks (segmentation, movie prediction) without much modification

- We can use these models to explore a variety of normative questions across the entirety of the ventral stream (V1, V4, dorsal stream)

Acknowledgements

Thanks!

Daniel Yamins

Daniel Bear

Jonas Kubilius

Kohitij Kar

Surya Ganguli

Javier Sagastuy

David Sussillo

Jim DiCarlo

Contact:
anayebi@stanford.edu

Funding:

Neurosciences PhD Program

Stanford Mind, Brain, Computation and Technology Training Program,
Wu Tsai Neurosciences Institute