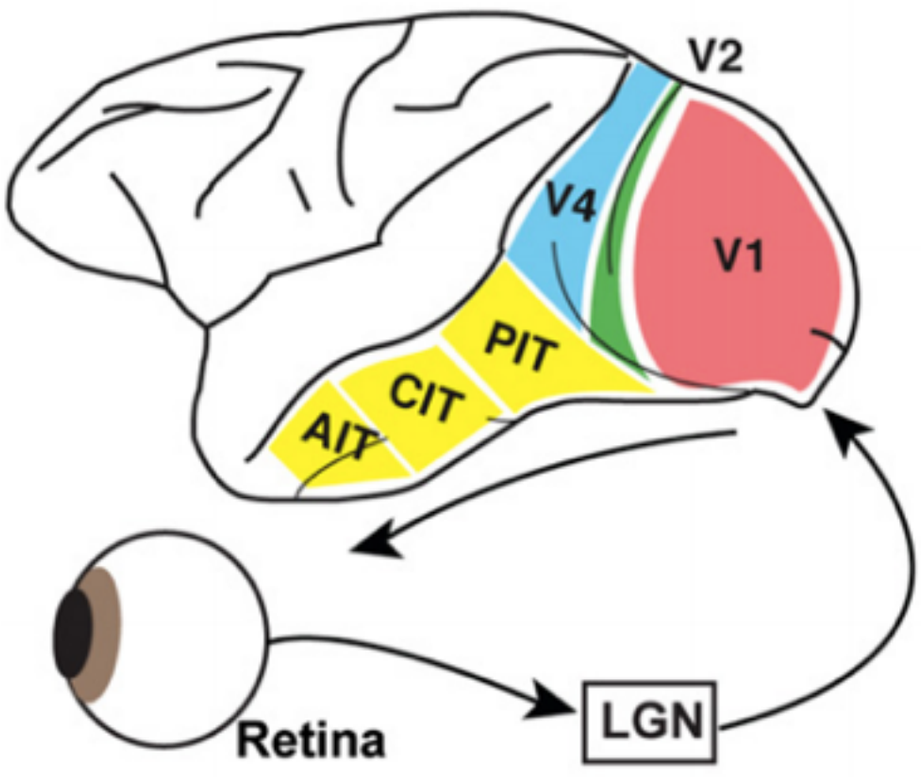
Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition

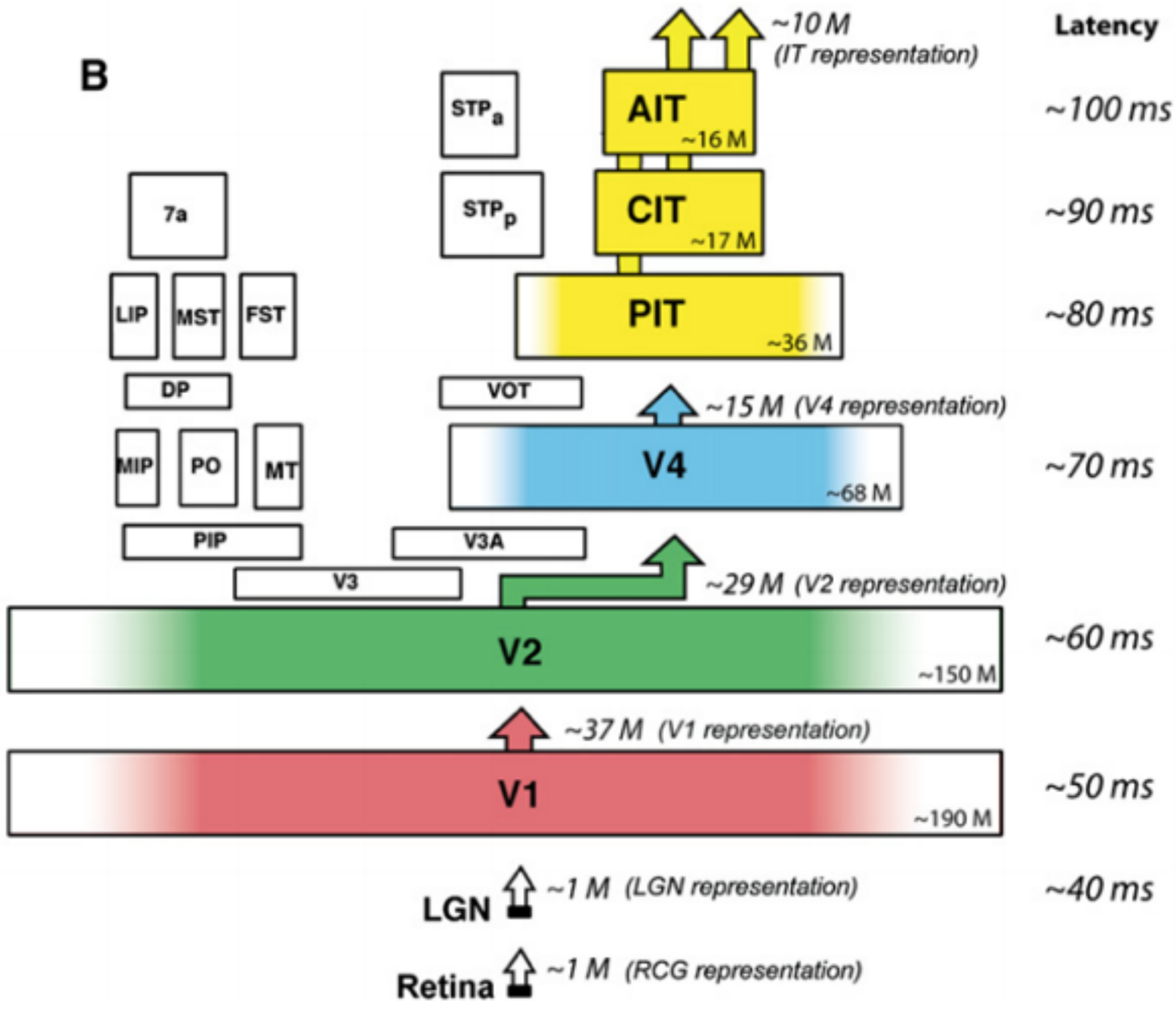
Charles Cadieu, Ha Hong, Dan Yamins, Nicolas Pinto, Diego Ardilla,
Ethan Solomon, Najib Majaj, and James DiCarlo

McGovern Institute for Brain Research
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

RE.WORK Deep Learning Summit, January 2015

Rapid
∧Object Recognition in the Macaque Visual System





The Task: Vision in a Glance

Behavior: can be performed by both human and macaque,

Neurons: visual cortex “solves” this task,

Algorithms: relatively difficult for machine algorithms

Practical: object recognition

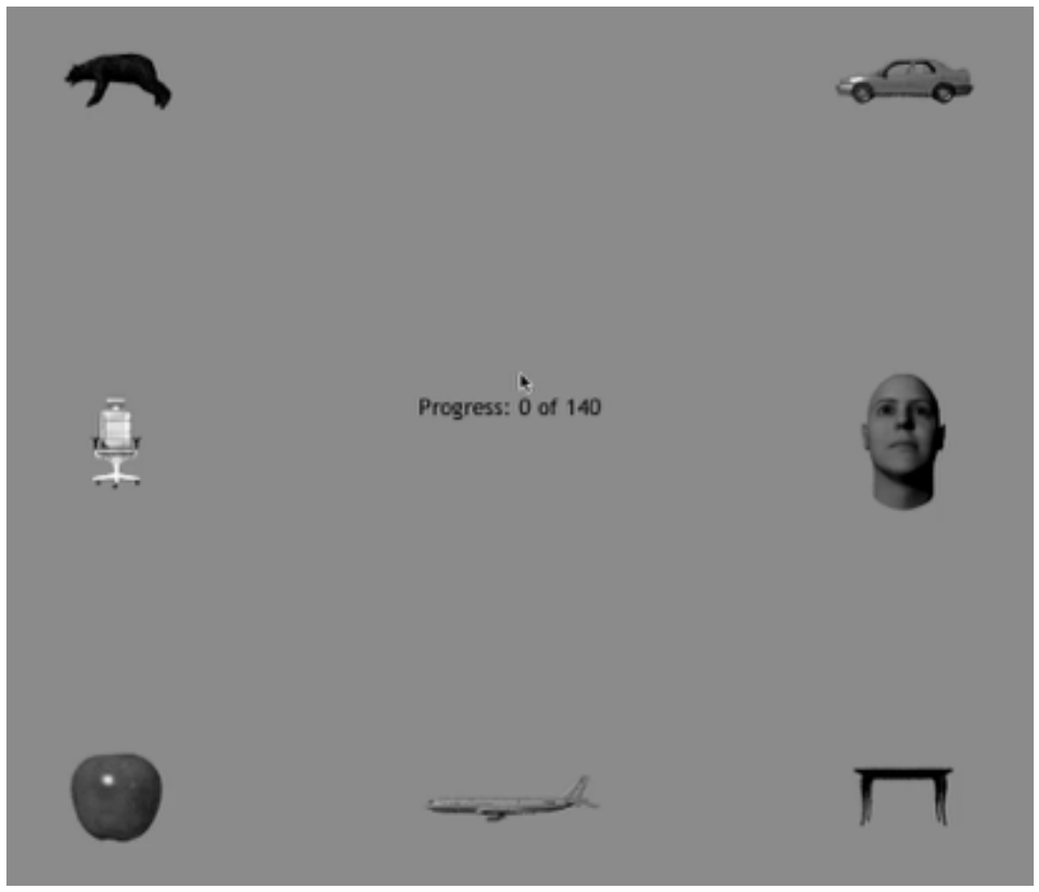
The Task: Vision in a Glance

Behavior: can be performed by both human and macaque,

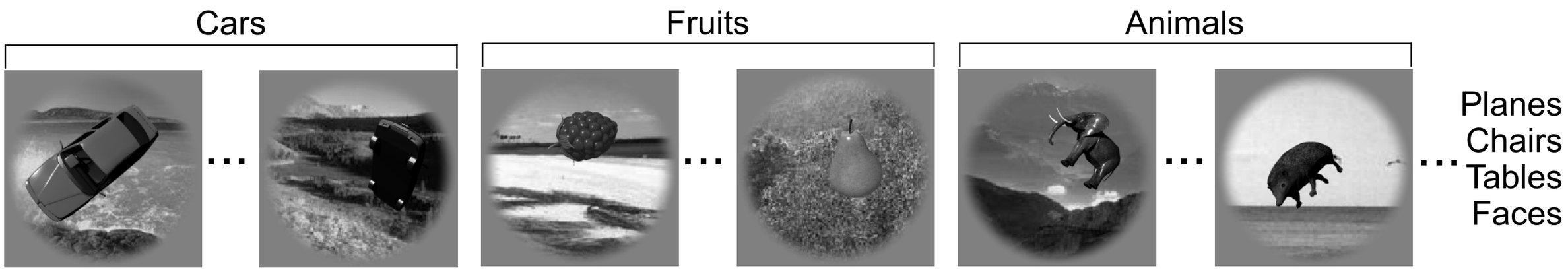
Neurons: visual cortex “solves” this task,

Algorithms: relatively difficult for machine algorithms

Practical: object recognition



The Task: Category-level Object Recognition



Variation due to:

- object exemplar (pear, raspberry)
- pose of object (facing left)
- position (on the right)
- scale (close, far)
- background

What is a “Representation”?

How should we represent Numbers?

Arabic Numerals (1, 2, 3, 4, …)

Roman Numerals (I, II, III, IV, …)

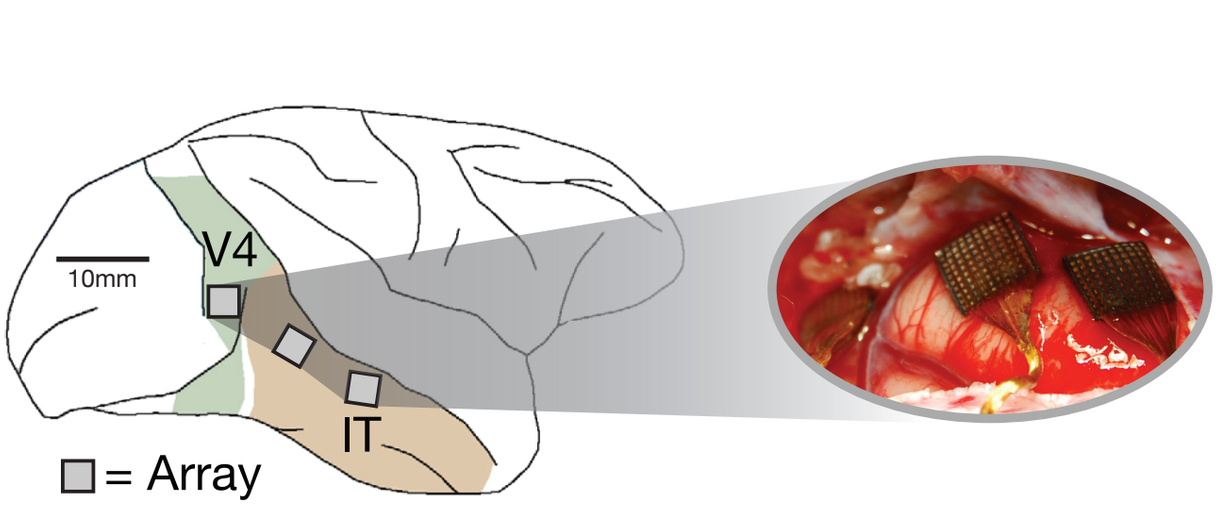
Binary Numbers (001, 010, 011, 100, …)

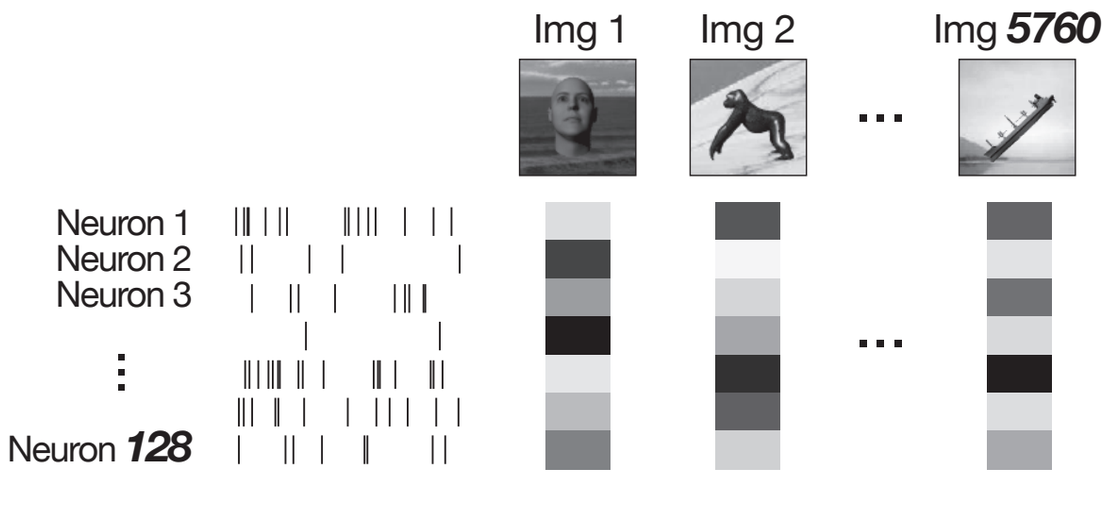
What is a “Representation”?

How should we represent *Images*?

In our Machines?	In the Brain?
Pixels	Retinae
JPEG	Primary Visual Cortex (V1)
Fourier Transform	V2
Bag-of-Words	V4
Neural Network Features	Inferotemporal Cortex (IT)

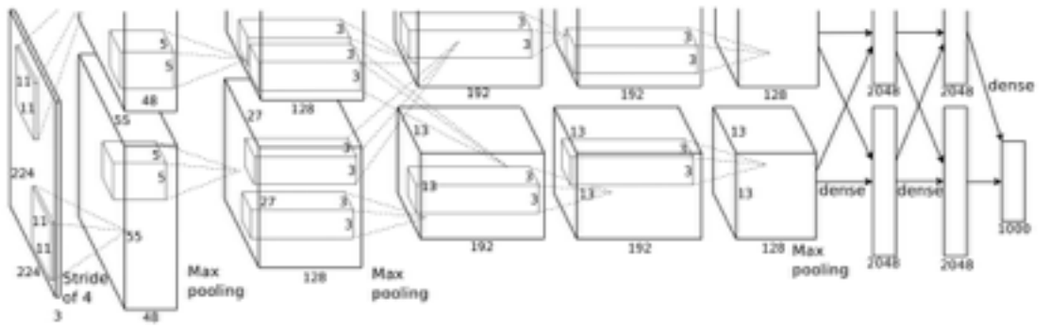
Measuring the Neural Representation



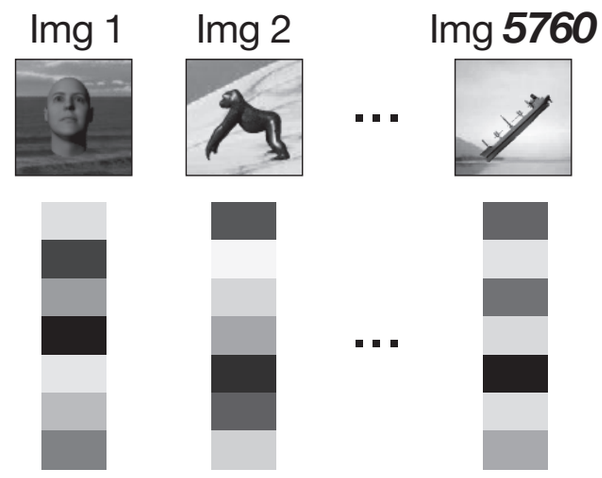


- Rapid serial visual presentation of the stimuli
- Grid electrodes placed in V4 and IT
- Neural feature: Multi-unit spike count, averaged over time window locked to presentation

Measuring the DNN Representation



Artificial neuron 1	0.01
Artificial neuron 2	3.26
Artificial neuron 3	1.17
…	…
Artificial neuron N	0.93



- Sent images to collaborators
- They ran their trained (usually trained on ImageNet) DNNs on our images to compute features
- We received features from different network layers.
- We have also run our own DNNs and verified results

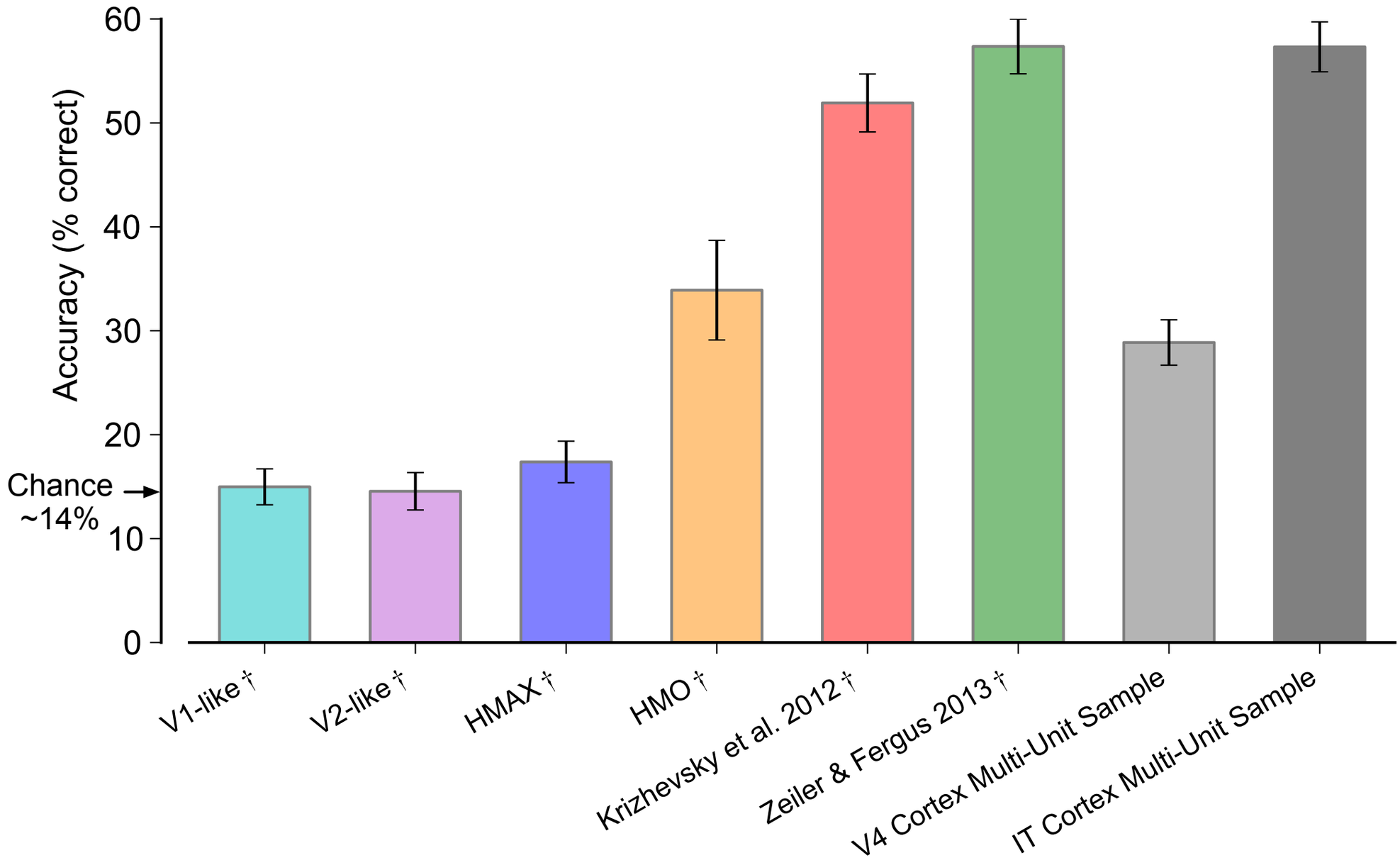
The Test: Effective Representation

An “effective” representation is one that makes the problem simple.

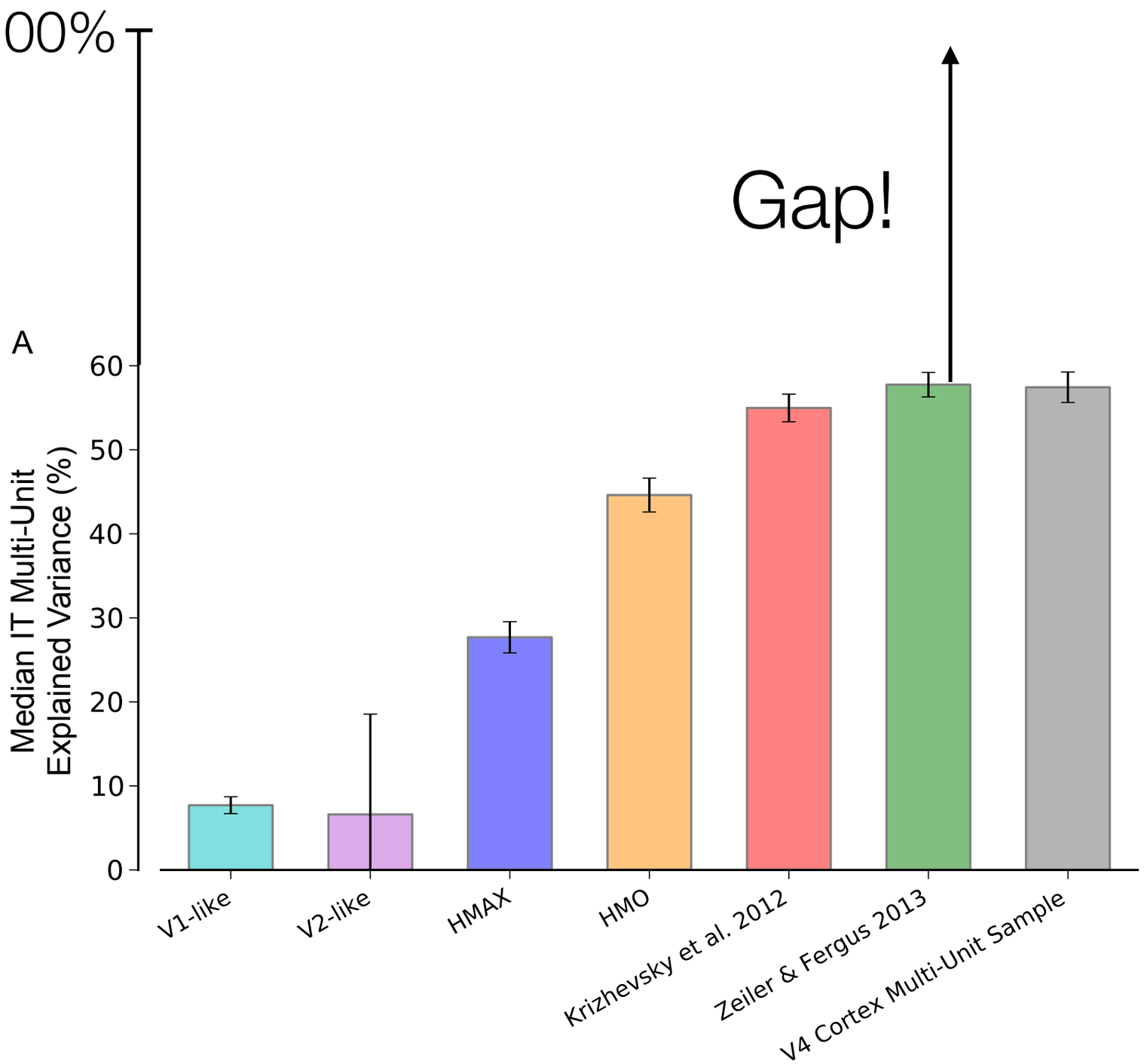
Object recognition is a complex problem when represented at the pixels or retinae.

Effective systems for visual object recognition transform the complex problem present in the pixel representation into a simple problem.

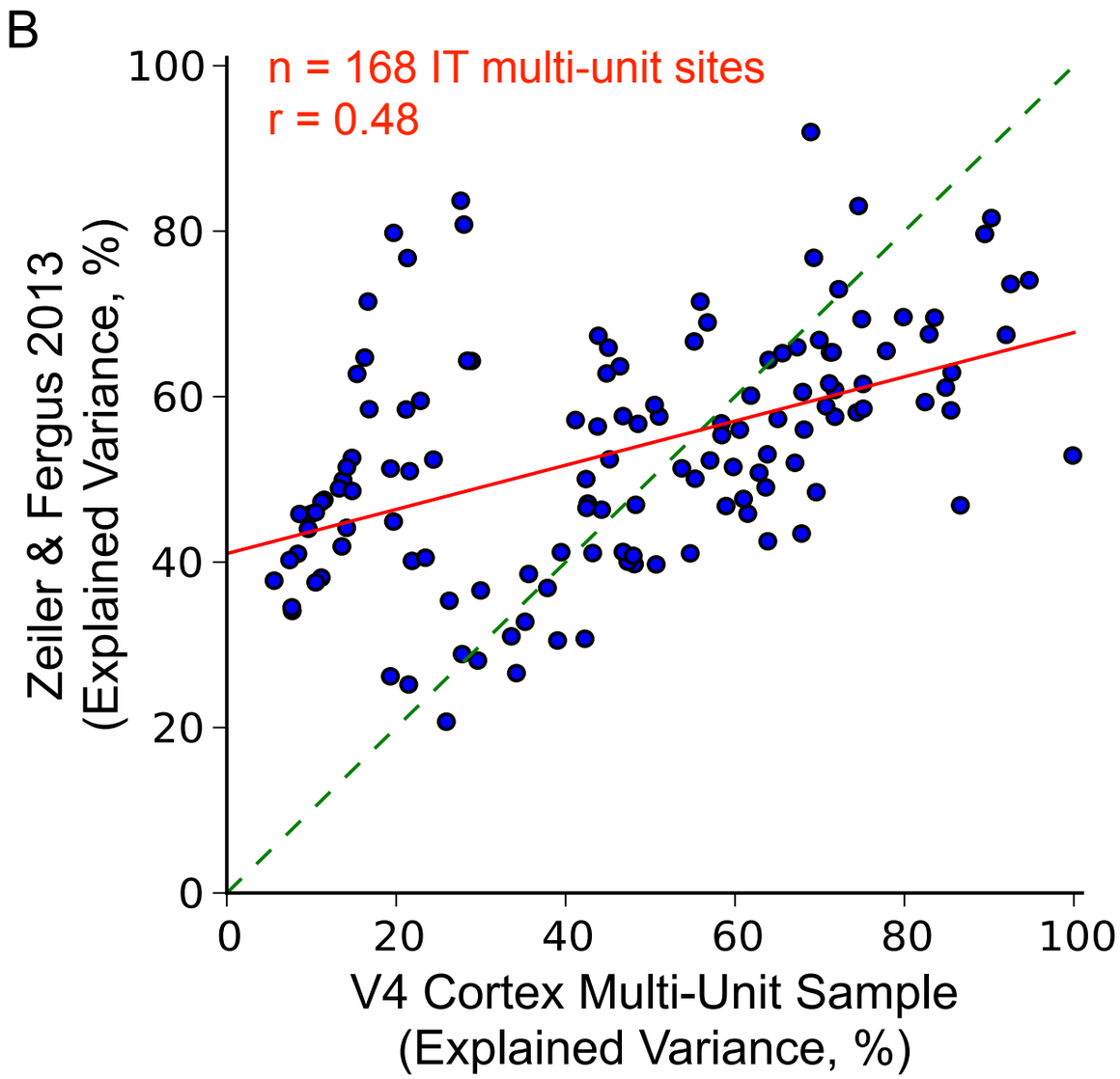
Result: Linear-SVM Analysis



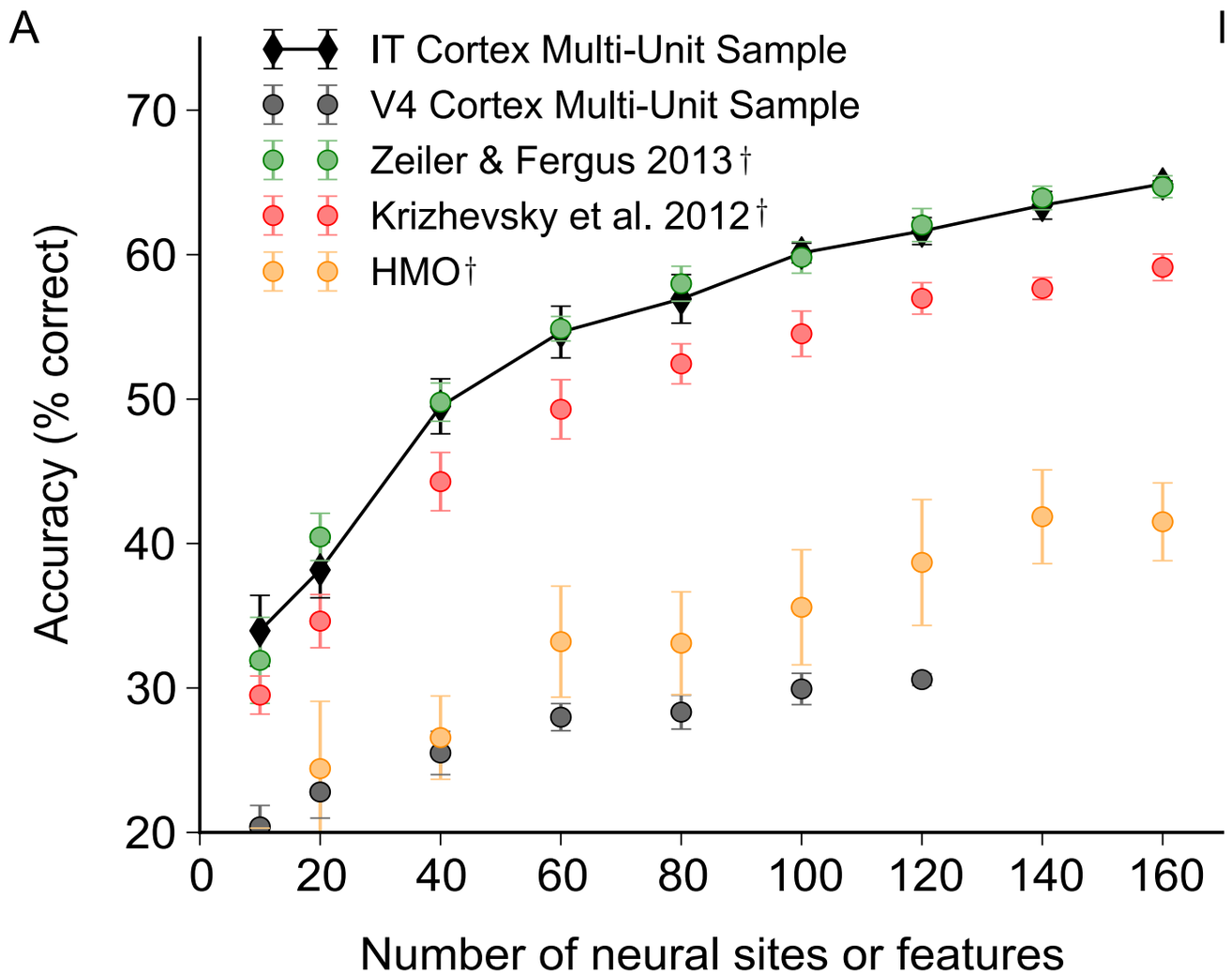
But are the DNNs “Like” the Neurons?



Explanation of Variance



Accuracy vs # Features



Implications

- We now have a better quantitative understanding of feedforward visual processing in the brain (but far from complete!).

- DNNs likely rival and will soon surpass humans on natural rapid visual object recognition.

- DNNs may shortly surpass any human ability in rapid judgement (given enough data).

Caveats

- Only “image in a glance” task.
- Have we measured the “right” neurons?
- Have we used the “right” way to readout the neurons?
- Could attention influence the result?
- Could learning or exposure influence the result?
- Active decision making (animals are awake, but passive viewing).
- Limited Object Categories
- Images without context (but by design!)