Learning Representations: Machine and Human

Sumeet Agarwal

IIT Delhi (sumeet@iitd.ac.in)

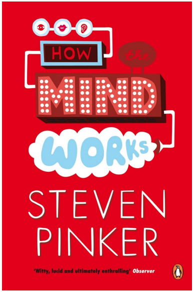
July 9, 2018

Neural network models

Neural networks as computational systems

- The classic mathematical model of the neuron is McCulloch-Pitts (1943)

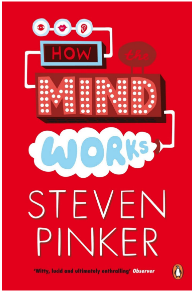
[Pinker 1999]



Neural networks as computational systems

- The classic mathematical model of the neuron is McCulloch-Pitts (1943)
- Sees neurons as switch-like, either ON (1) or OFF (0)

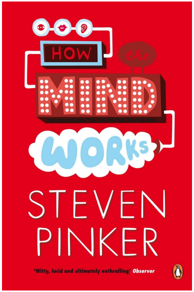
[Pinker 1999]



Neural networks as computational systems

- The classic mathematical model of the neuron is McCulloch-Pitts (1943)
- Sees neurons as switch-like, either ON (1) or OFF (0)
- Each neuron takes a weighted sum of inputs and applies a threshold to it, to decide whether to fire or not

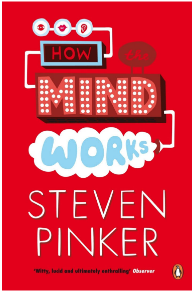
[Pinker 1999]



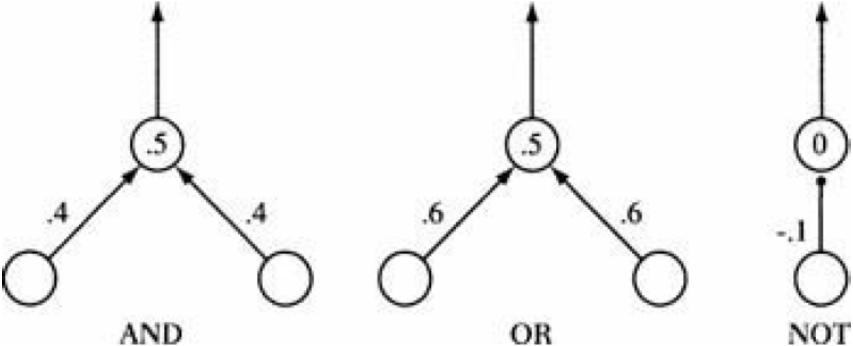
Neural networks as computational systems

- The classic mathematical model of the neuron is McCulloch-Pitts (1943)
- Sees neurons as switch-like, either ON (1) or OFF (0)
- Each neuron takes a weighted sum of inputs and applies a threshold to it, to decide whether to fire or not
- They can thus encode more abstract logical operations

[Pinker 1999]



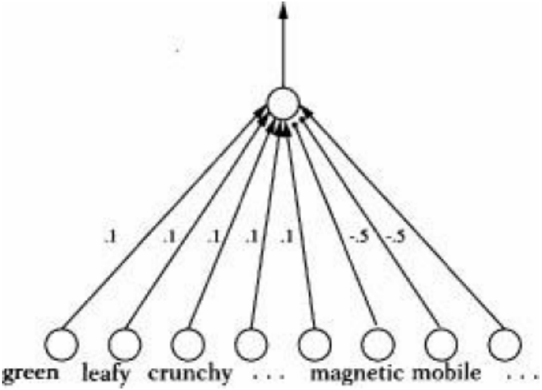
Neural networks as computational systems



[Pinker 1999]

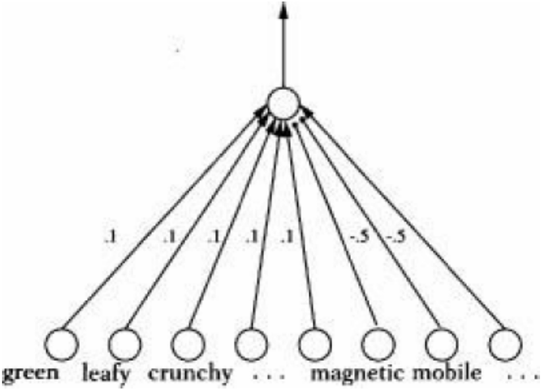
Neural networks as computational systems

Vegetable detection:

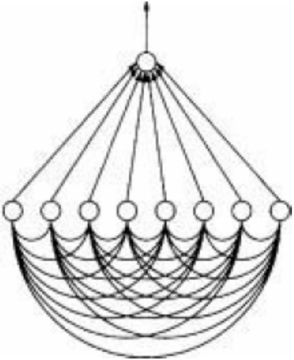


Neural networks as computational systems

Vegetable detection:



Auto-association:



[Pinker 1999]

Cognition as pattern recognition

- **A vheclie epxledod at a plocie cehckipont near the UN haduqertares in Bagahdd on Mnoday kilinlg the bmober and an Irqai polcie offceir** [Matt Davis, MRC Cognition and Brain Sciences Unit, Cambridge]

Cognition as pattern recognition

- **A vheclie epxledod at a plocie cehckipont near the UN haduqertares in Bagahdd on Mnoday kilinlg the bmober and an Irqai polcie offceir** [Matt Davis, MRC Cognition and Brain Sciences Unit, Cambridge]
-  [Pinker 1999]

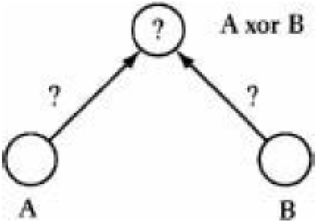
Cognition as pattern recognition

- **A vheclie epxledod at a plocie cehckipont near the UN haduqertares in Bagahdd on Mnoday kilinlg the bmober and an Irqai polcie offceir** [Matt Davis, MRC Cognition and Brain Sciences Unit, Cambridge]
-  [Pinker 1999]
- Robustness to noise and missing information; inference to fill in missing details

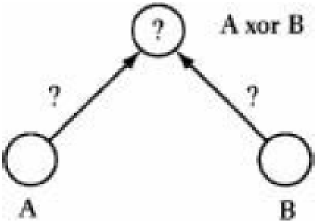
Cognition as pattern recognition

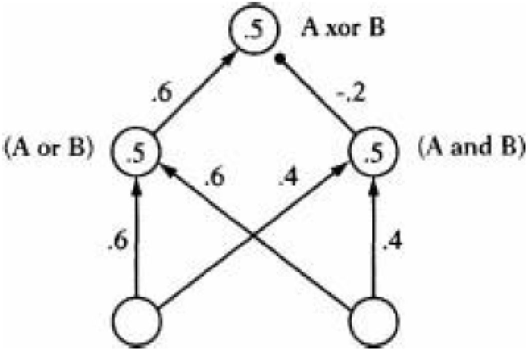
- **A vheclie epxledod at a plocie cehckipont near the UN haduqertares in Bagahdd on Mnoday kilinlg the bmober and an Irqai polcie offceir** [Matt Davis, MRC Cognition and Brain Sciences Unit, Cambridge]
-  [Pinker 1999]
- Robustness to noise and missing information; inference to fill in missing details
- Fits with computational neural network models; hard to explain with purely rule-based models

The XOR problem



The XOR problem

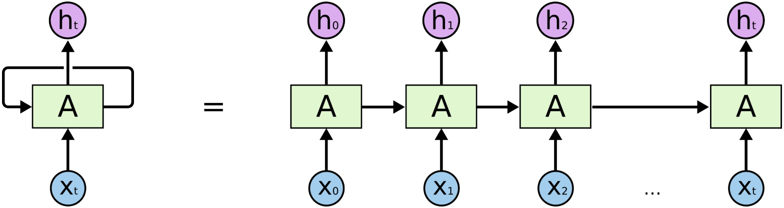




[Pinker 1999]

Recurrent neural networks (RNNs)

Rather than just feed-forward connections, RNNs also allow for recurrent or feedback connections, thus allowing a 'memory' of previous states to be retained. This is useful for processing sequential or temporal data.



[http://colah.github.io/posts/2015-08-Understanding-LSTMs]

Long-range dependencies

- One key challenge in language processing is dealing with long-range dependencies

Long-range dependencies

- One key challenge in language processing is dealing with long-range dependencies
- Consider the sentence *I looked up to see a cloudy* ___. Here just the context of a single preceding word predicts the next with high confidence: can even be done by a bigram model

Long-range dependencies

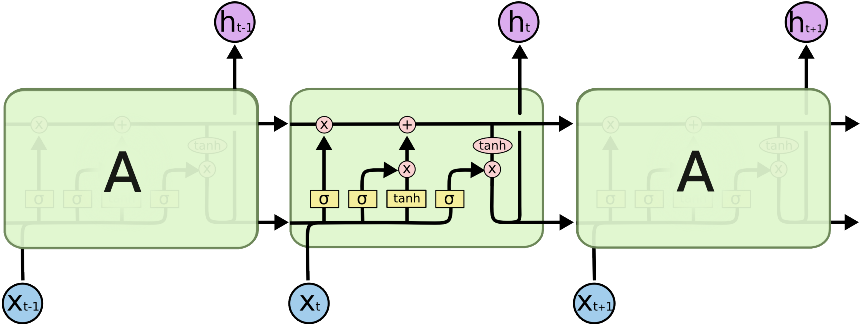
- One key challenge in language processing is dealing with long-range dependencies
- Consider the sentence *I looked up to see a cloudy* ___. Here just the context of a single preceding word predicts the next with high confidence: can even be done by a bigram model
- However, consider *I was born in Paris and spent my childhood there, so I speak fluent* _____. Here a bigram model would predict the next word to be the name of a language; but to predict which language, you need information from much further back in the sentence

Long-range dependencies

- One key challenge in language processing is dealing with long-range dependencies
- Consider the sentence *I looked up to see a cloudy* ___. Here just the context of a single preceding word predicts the next with high confidence: can even be done by a bigram model
- However, consider *I was born in Paris and spent my childhood there, so I speak fluent* _____. Here a bigram model would predict the next word to be the name of a language; but to predict which language, you need information from much further back in the sentence
- RNNs can in principle learn such long-range dependencies, but it is difficult for vanilla RNNs; a specific variety, called LSTMs, are much more powerful at this

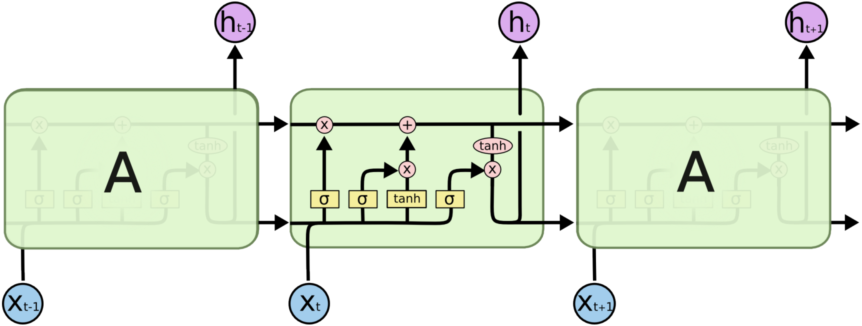
Long Short-Term Memory (LSTM) models

These have a much more sophisticated, multi-layered repeating module:



[http://colah.github.io/posts/2015-08-Understanding-LSTMs]

Long Short-Term Memory (LSTM) models



Very crudely, these essentially work via the repeating module largely passing on information (the 'cell state') from the previous time step as is (the horizontal line along the top). But necessary changes/updates to this state can be made via carefully regulated 'gates'.

RNN applications

- RNNs (mainly LSTMs) have been extremely successful for a range of linguistic tasks (The Unreasonable Effectiveness of Recurrent Neural Networks), and the ability to model the maintenance of long-range dependencies in short-term or working memory seems key to this success

RNN applications

- RNNs (mainly LSTMs) have been extremely successful for a range of linguistic tasks (The Unreasonable Effectiveness of Recurrent Neural Networks), and the ability to model the maintenance of long-range dependencies in short-term or working memory seems key to this success
- Hence these models are clearly of interest from a psycholinguistic perspective, even though so far they have been more prominent in the NLP literature

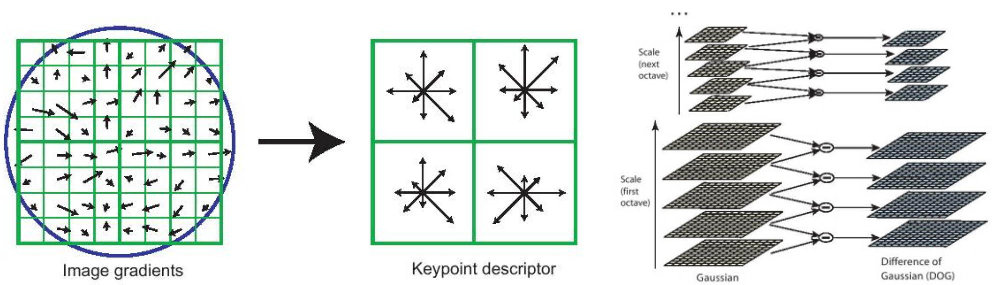
Neural Network and Deep Learning Approaches to Computer Vision

Sumeet Agarwal
Department of Electrical Engineering
IIT Delhi

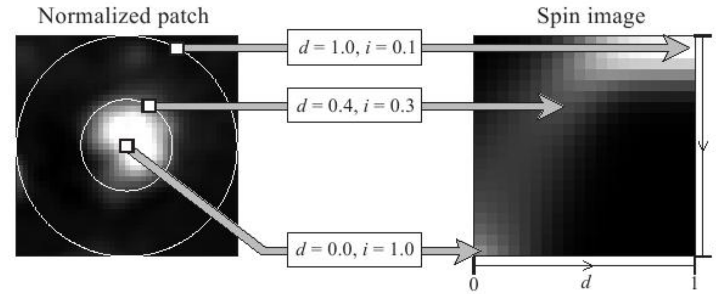
What is the key challenge in vision?

- Arguably, extracting meaningful *features* from images
- How do we construct increasingly complex/abstract representations, starting with raw pixels?
- These representations can be handcoded; but can they also be *learnt* automatically from data?
- Does the learning of such representations have to be guided/supervised, or can it also be achieved in an *unsupervised* fashion?

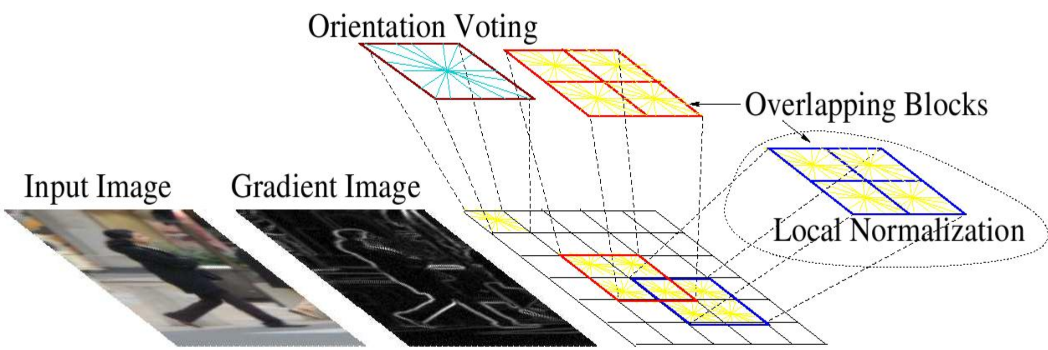
Computer vision features



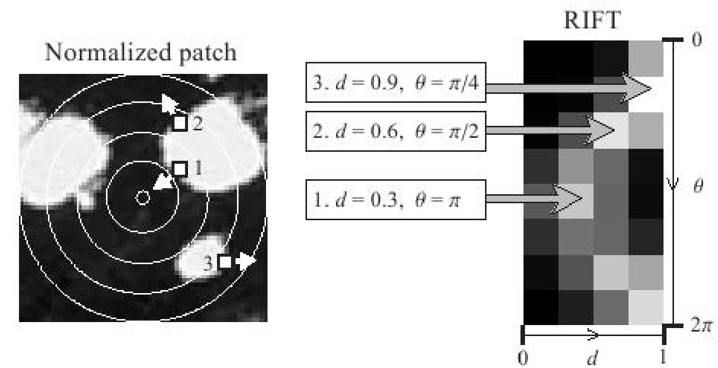
SIFT



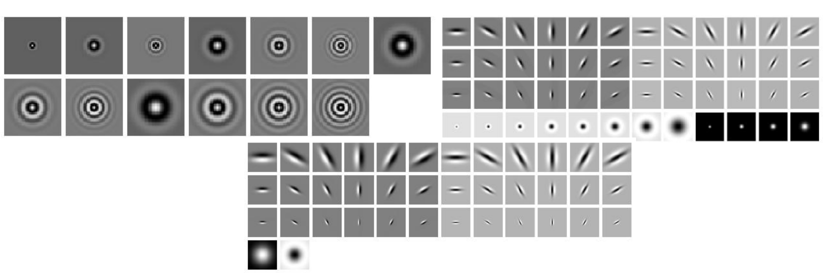
Spin image



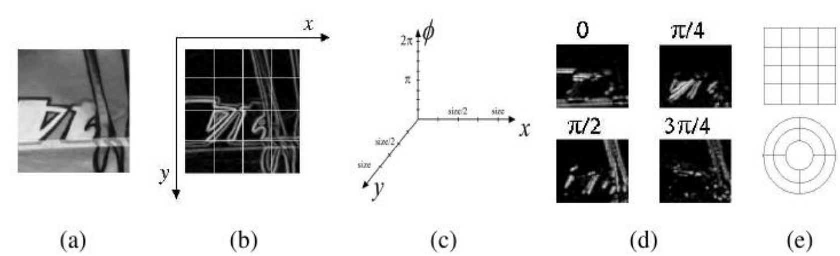
HoG



RIFT



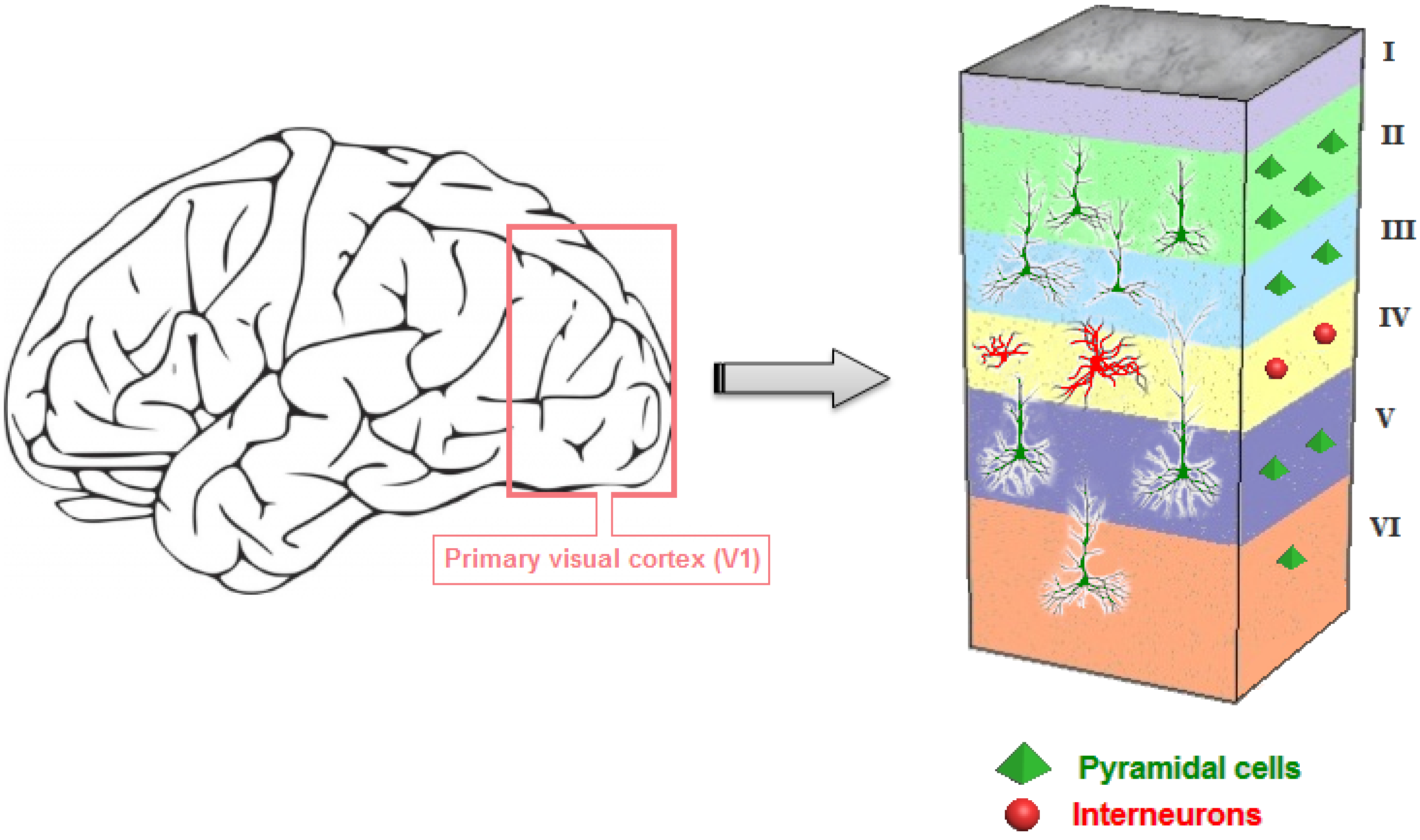
Textons



GLOH

[Andrew Ng]

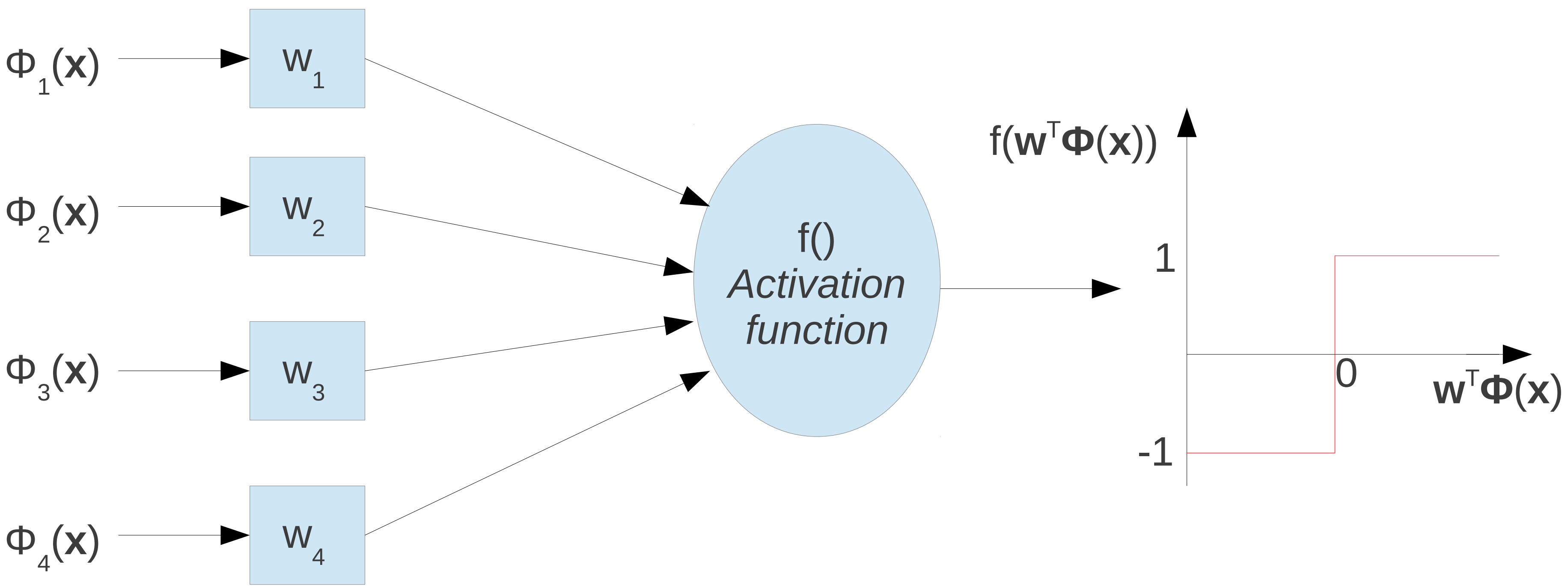
Human vision



[Bachatene *et al.*, 2012]

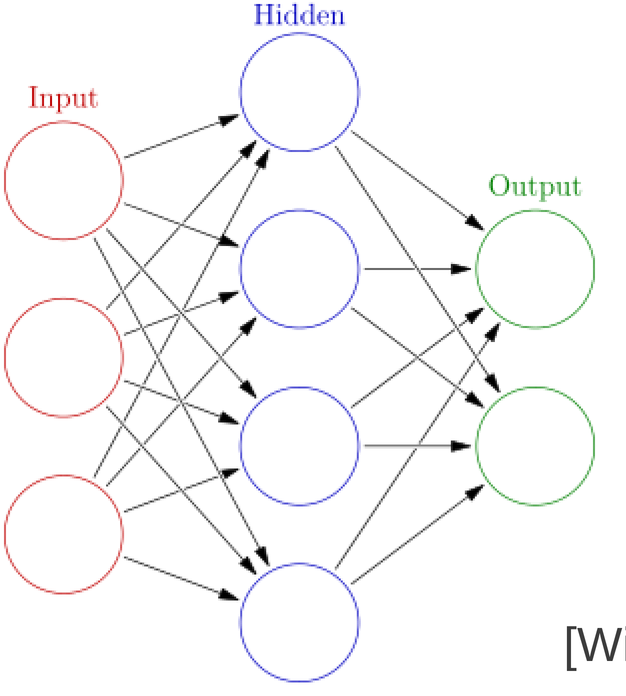
Neuronal networks build up a hierarchy of increasingly complex representations.

Learning models: The Perceptron



A non-linear transformation in the form of a step function is applied to the weighted sum of the input features. This is inspired by the way neurons appear to function, mimicking the *action potential*.

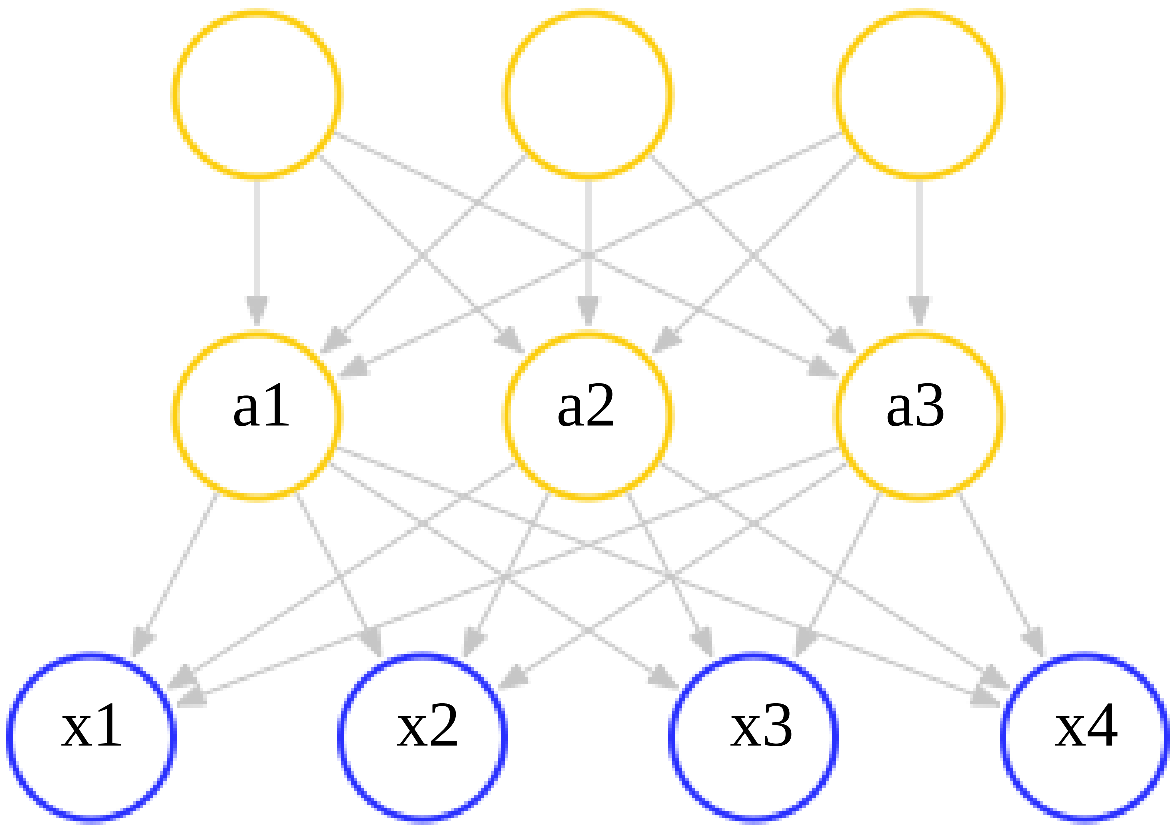
Neural Networks: *Multilayer Perceptrons*



[Wikipedia]

Usually, the non-linear activation function used is a logistic sigmoid: $y = f(\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x})) = \sigma(\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}))$, where $\sigma(a) = 1/(1+e^{-a})$. This makes *y* a differentiable function of the input **x**; each unit/neuron can now be thought of as simply a logistic regression classifier.

Learning feature hierarchies



Higher layer
(Combinations of edges;
cf. V2)

“Sparse coding”
(edges; cf. V1)

Input image (pixels)

[Technical details: Sparse autoencoder or sparse version of Hinton’s DBN.]

[Lee, Ranganath & Ng, 2007]

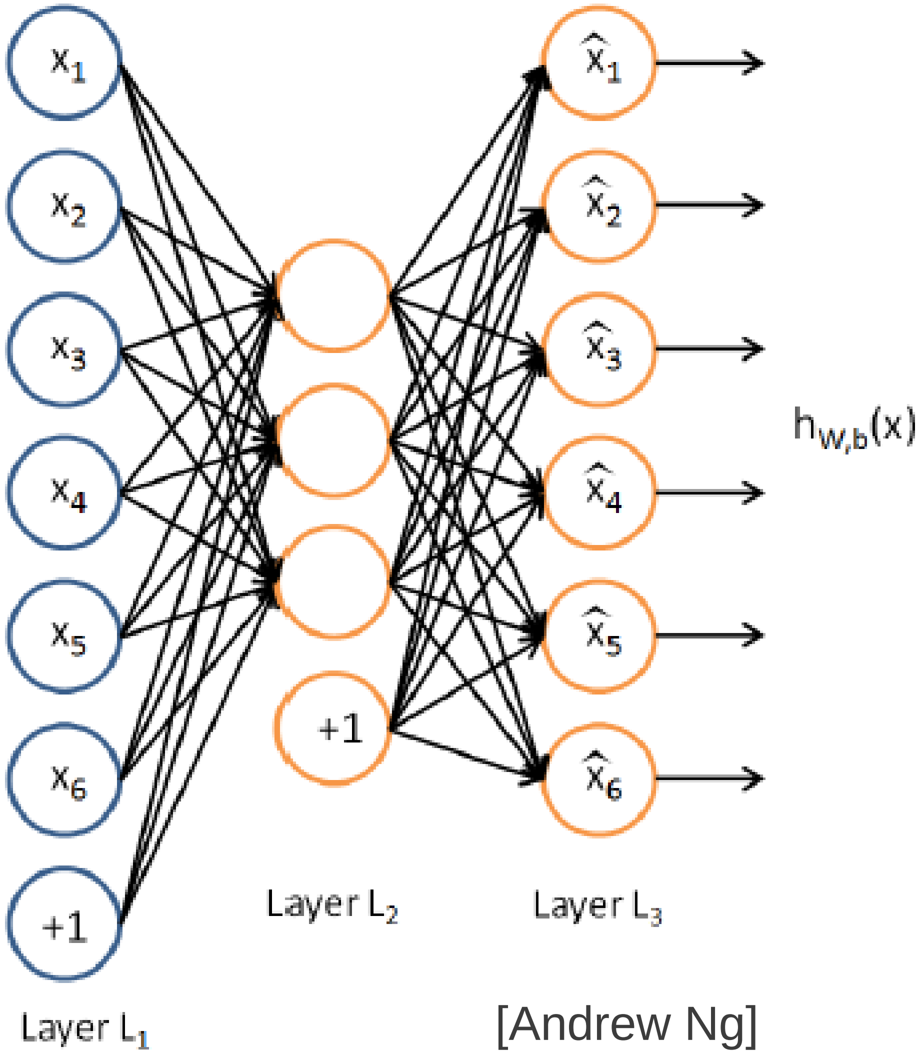
Supervised learning with neural nets

- Target values for the output(s) can be provided as categorical or continuous values, corresponding to classification and regression settings
- An appropriate error function is defined and minimised with respect to the network weights
- Typically done using gradient descent; the gradient of the error function can be computed via *backpropagation*

'Deep' learning

- Is just a fashionable term for the use of neural networks with many hidden layers
- The aim is for hidden neurons to be able to capture a hierarchy of representations, similar to the visual cortex
- Labelled training data may be limited; can useful representations also be learnt in an unsupervised fashion?

Sparse autoencoders



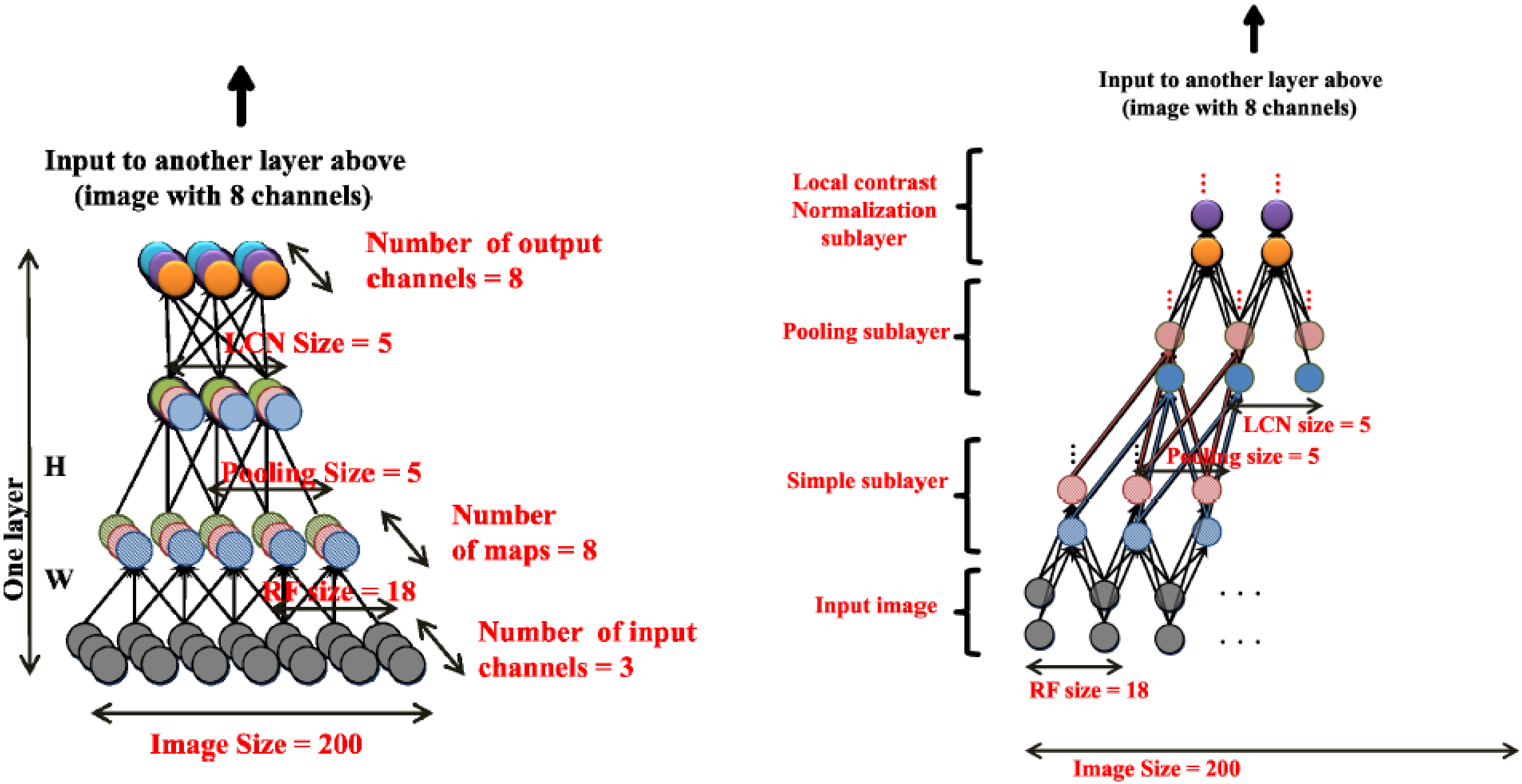
[Andrew Ng]

- A neural net with as many outputs as inputs
- The idea is to reproduce the input as closely as possible (minimise reconstruction error)
- How can this be done whilst retaining a relatively small number of features in the hidden layers, i.e., enforcing *sparsity*?
- Represents a form of *dimensionality reduction*

Large-scale deep learning

- Made possible by massive computing power and parallelisation
- *Google Brain*: A deep neural network with 9 layers and ~10^9 connections
- Still only one-millionth the size of a 3-year-old human brain!
- Important for demonstrating that complex concepts like faces can be discovered in an entirely *unsupervised* fashion

Visual input network architectures



[Le *et al*. 2012]

Conclusions

- Classical neural networks provide a biologically-inspired approach to the problem of learning appropriate visual representations
- Recent advances in technology have made it possible to train 'deep' networks, with millions or billions of connections
- Unsupervised learning by minimising reconstruction error whilst enforcing sparsity can be a powerful tool for feature/concept discovery

References

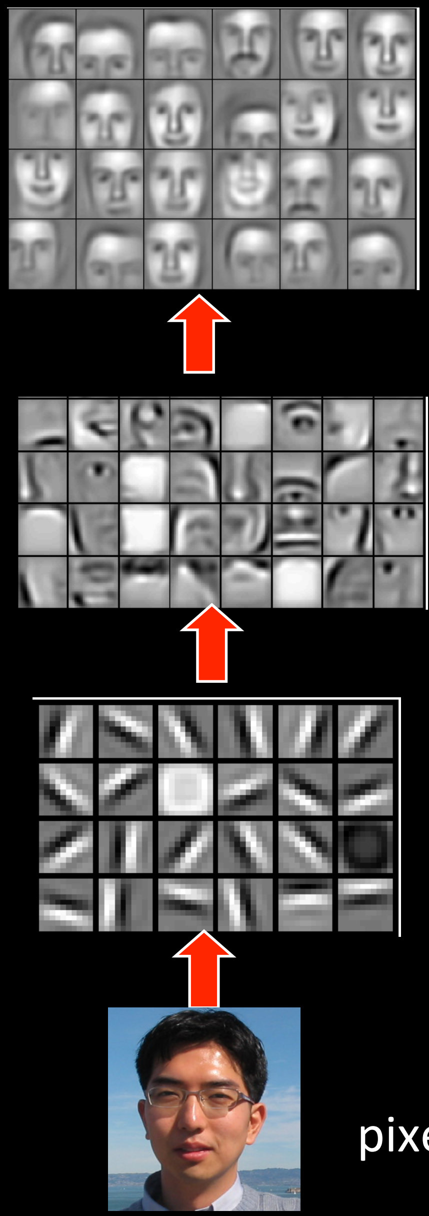
- Chris Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- Quoc V. Le *et al*. Building High-level Features Using Large Scale Unsupervised Learning. *ICML* 2012.
- Quoc V. Le *et al*. ICA with Reconstruction Cost for Efficient Overcomplete Feature Learning. *NIPS* 2011.
- Andrew Ng. Sparse autoencoder. *CS294A Lecture notes*, Stanford University, 2011.
- Lyes Bachatene *et al*. Adaptation and Neuronal Network in Visual Cortex. *Visual Cortex - Current Status and Perspectives*. InTech, 2012.

Building high-level features using large scale unsupervised learning

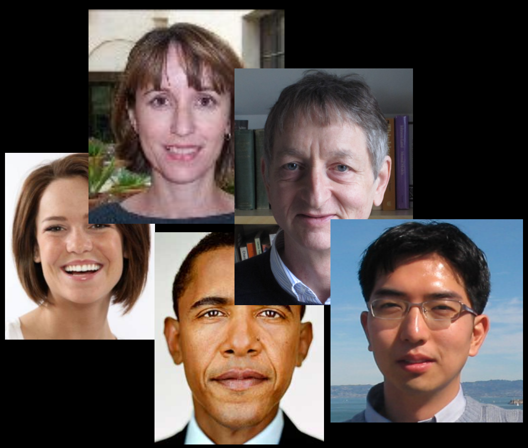
Quoc V. Le
Stanford University and Google

Joint work with: Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg Corrado, Jeff Dean, Andrew Y. Ng

Hierarchy of feature representations



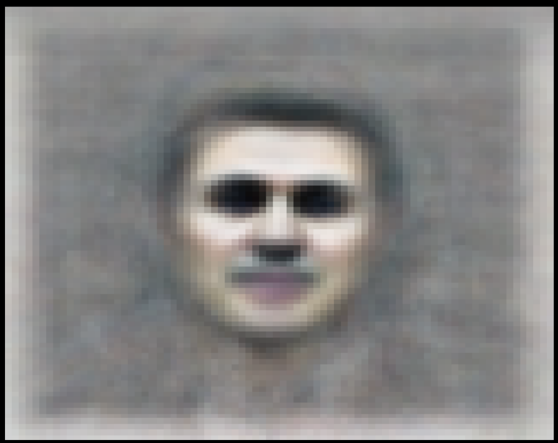
Lee et al, 2009. Sparse DBNs.



Faces

Random images from the Internet

Key results

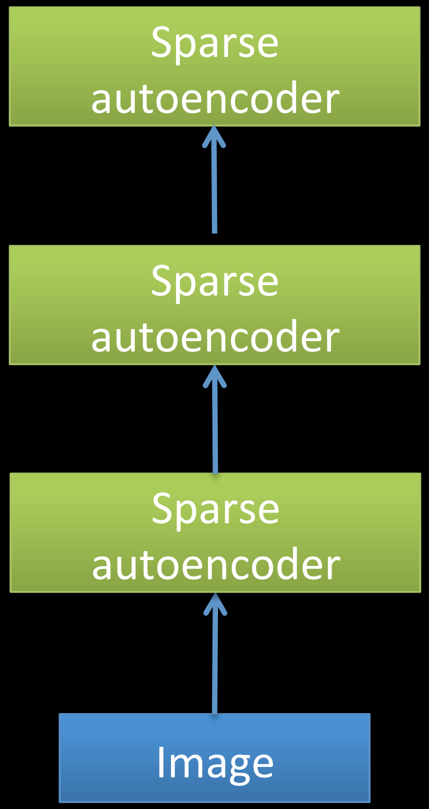


Face detector

Human body detector

Cat detector

Algorithm

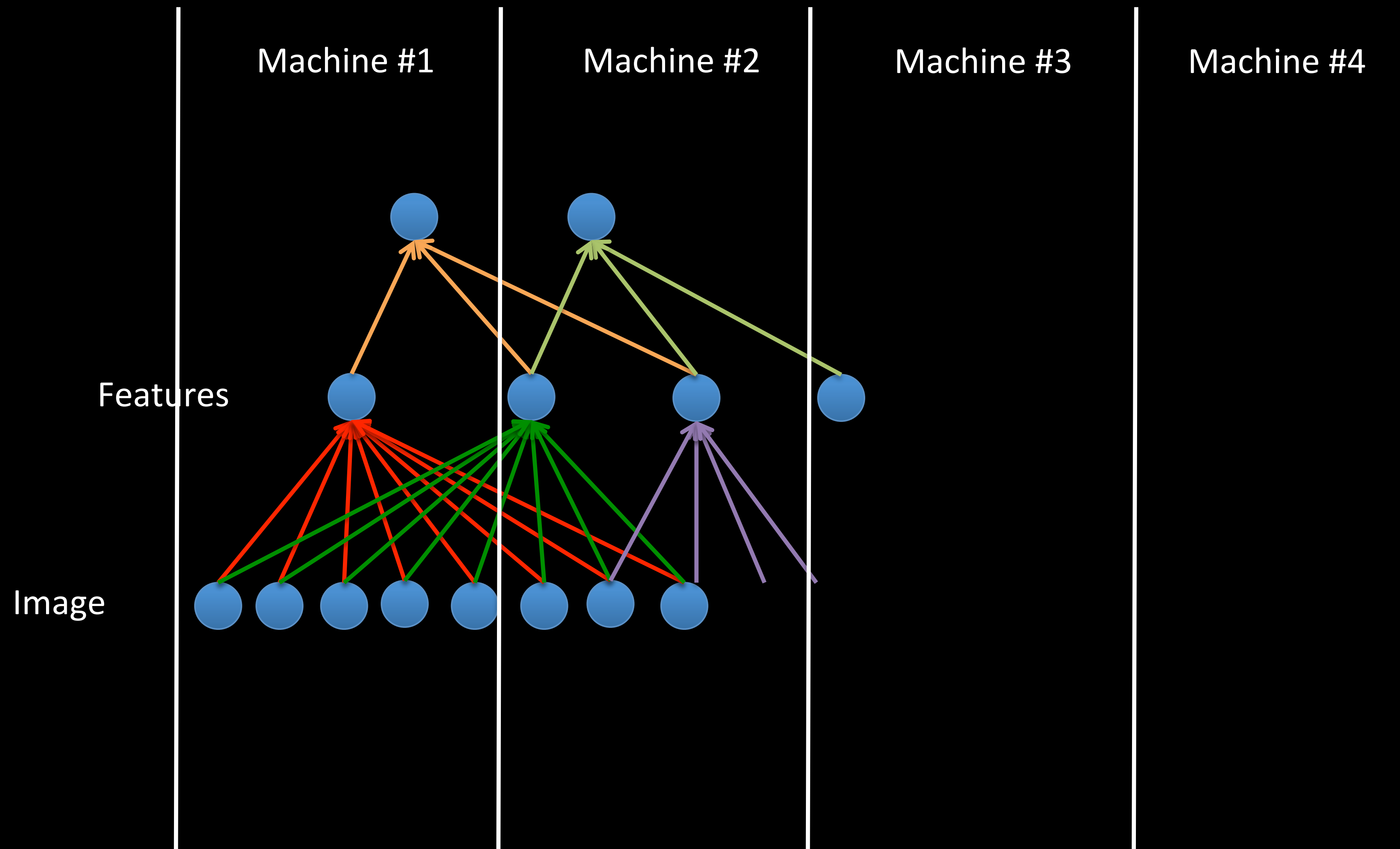


Each RICA layer = 1 filtering layer + pooling layer + local contrast normalization layer

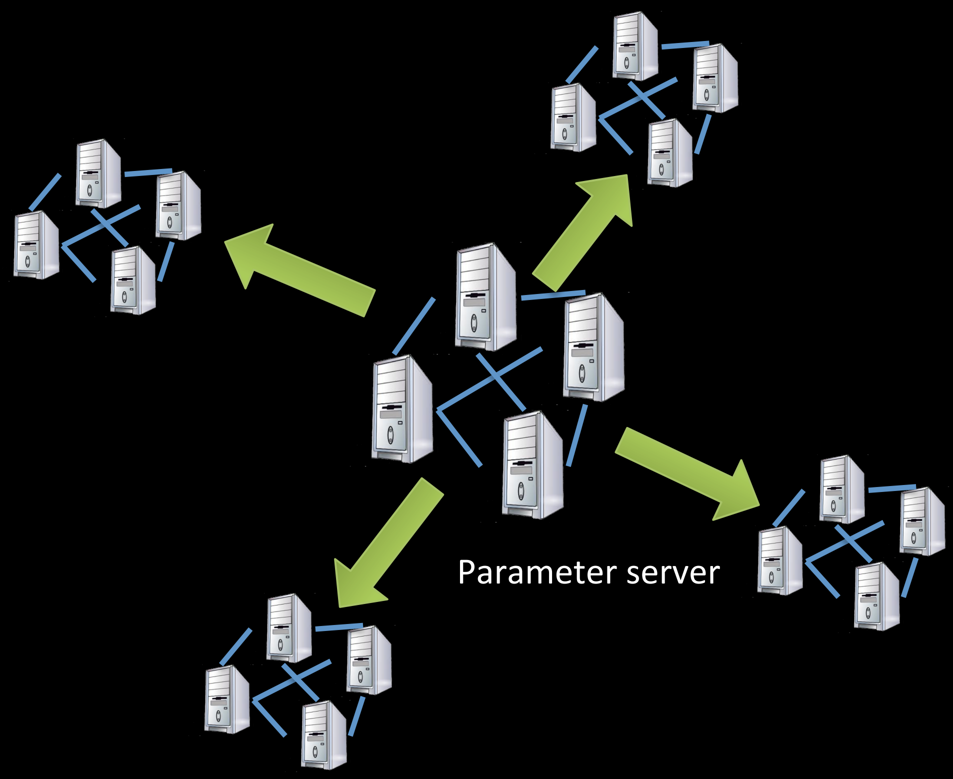
See Le et al, NIPS 11 and Le et al, CVPR 11 for applications on action recognition, object recognition, biomedical imaging

Very large model -> Cannot fit in a single machine
-> Model parallelism, Data parallelism

Local receptive field networks

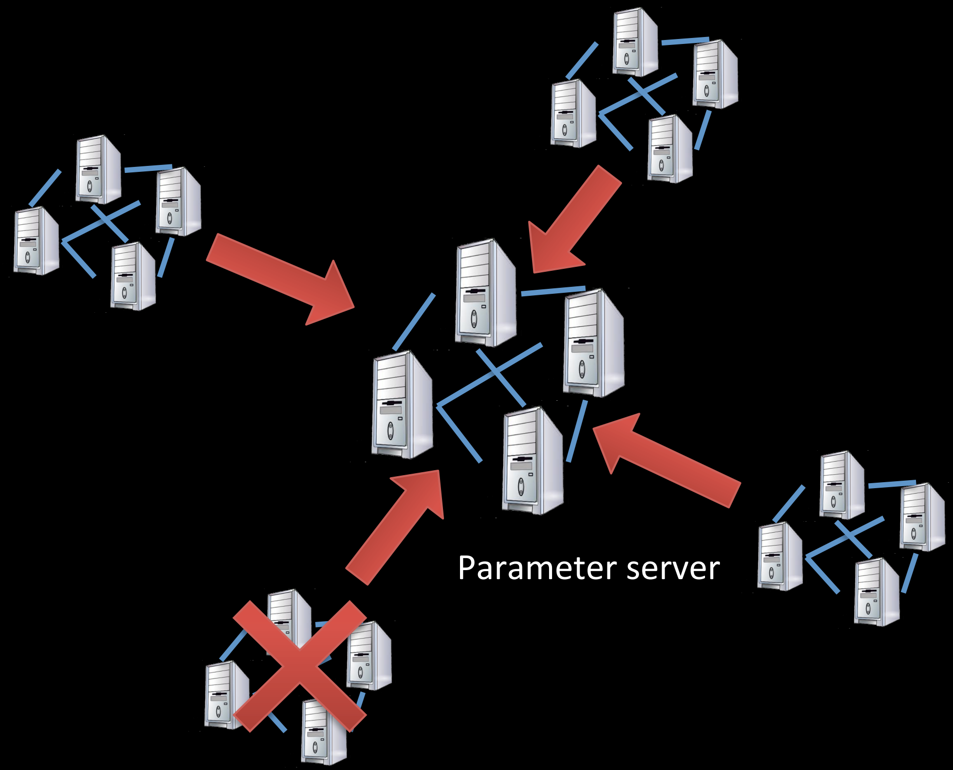


Asynchronous Parallel SGDs



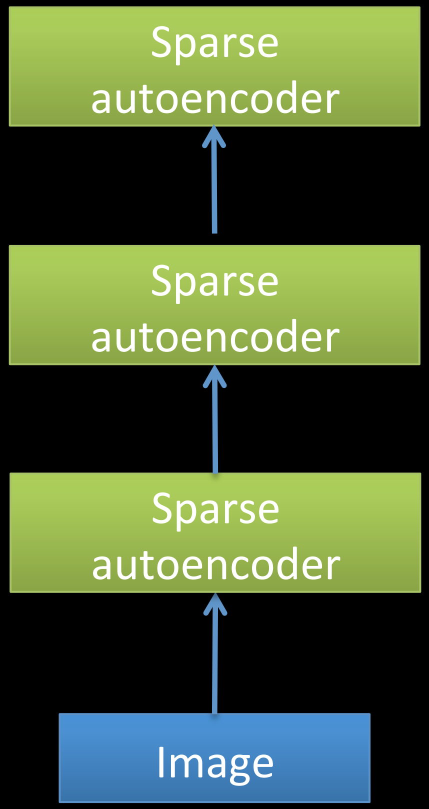
Quoc V. Le

Asynchronous Parallel SGDs



Quoc V. Le

Training



Dataset: 10 million 200x200 unlabeled images from YouTube/Web

Train on 1000 machines (16000 cores) for 1 week

1.15 billion parameters

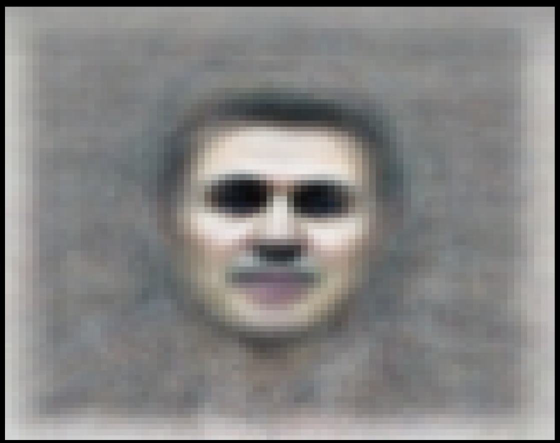
- 100x larger than previously reported
- Small compared to visual cortex

Top stimuli from the test set

Quoc V. Le

Optimal stimulus via optimization

Quoc V. Le

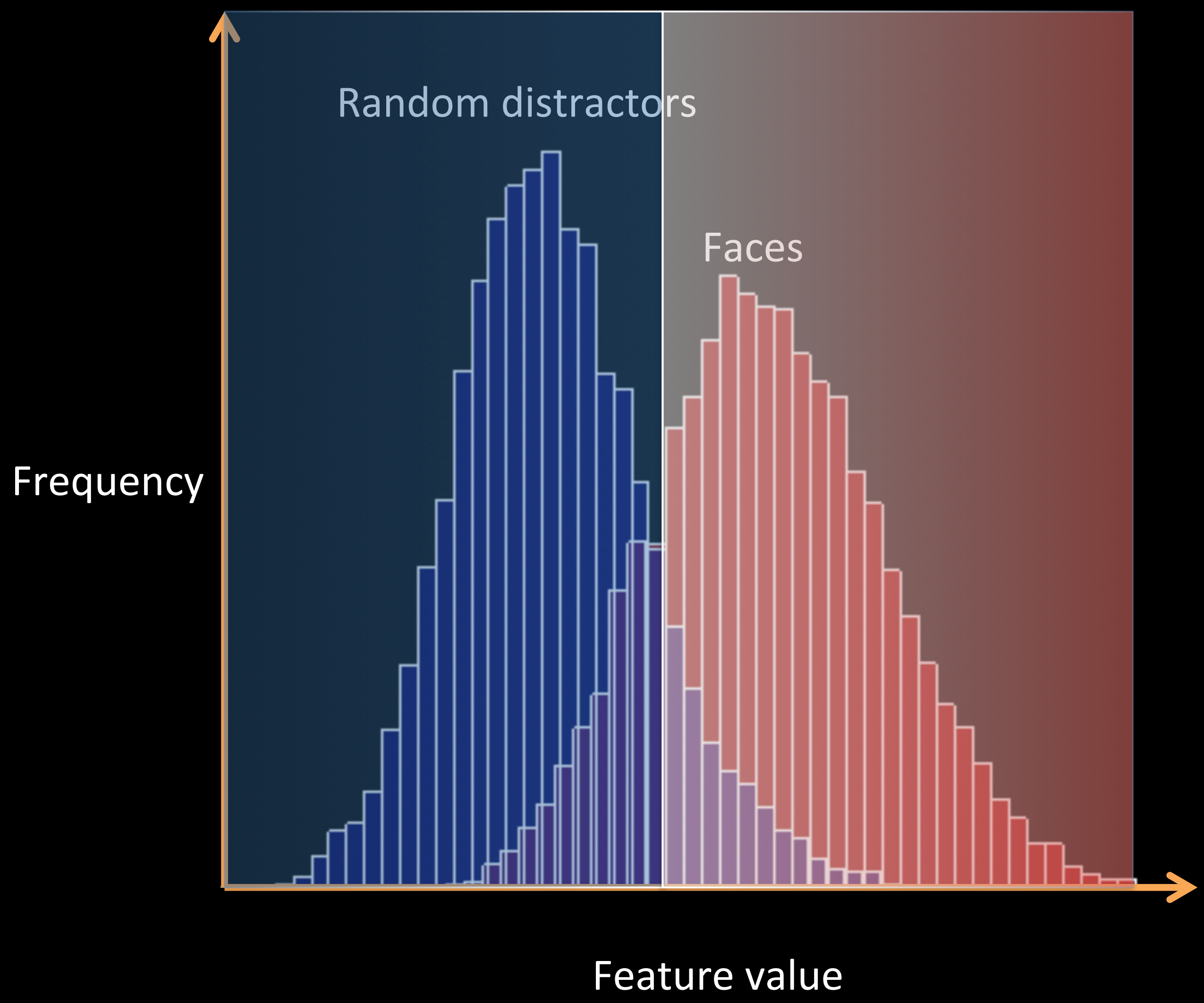


Face detector

Human body detector

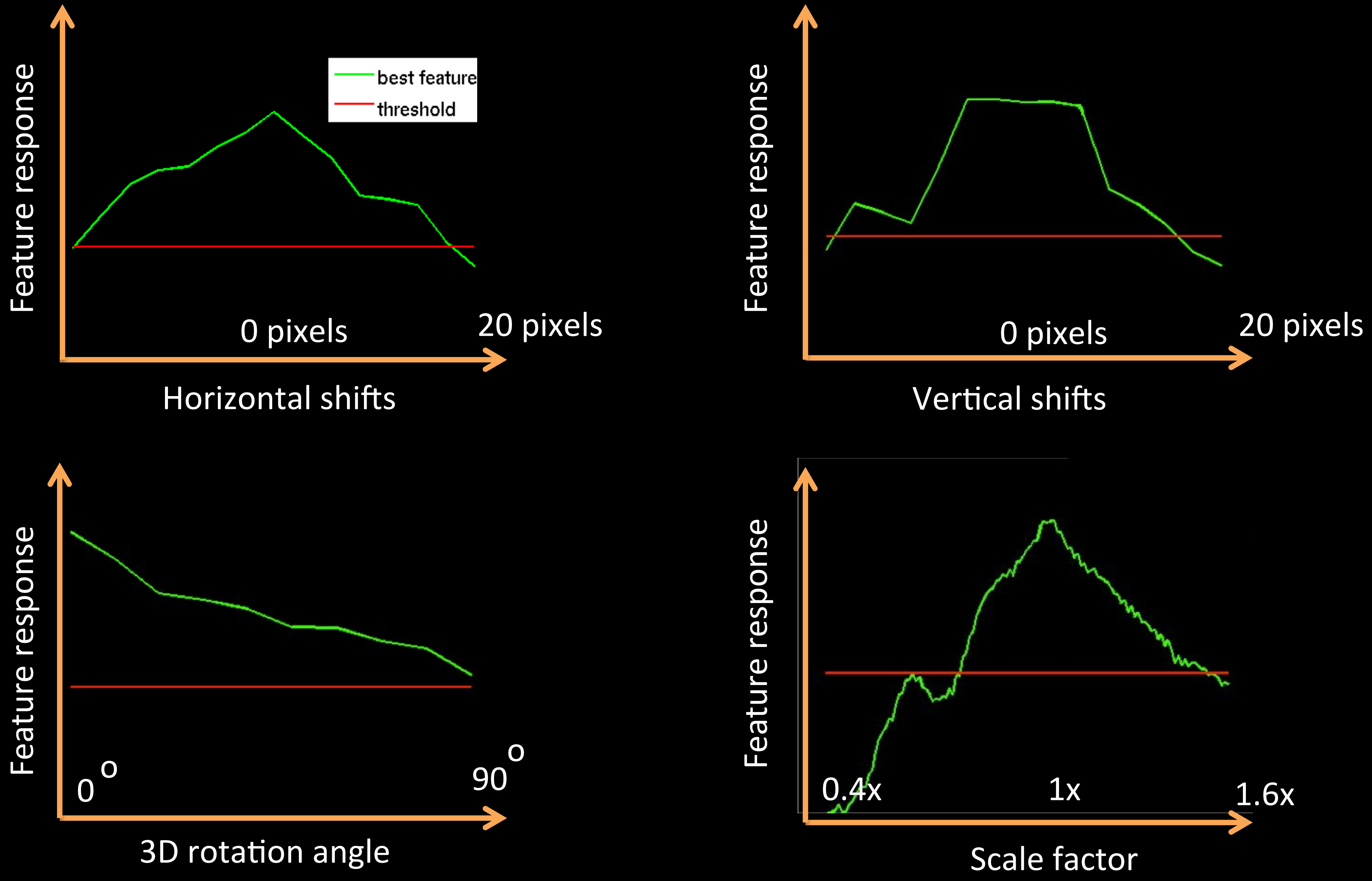
Cat detector

Quoc V. Le



Quoc V. Le

Invariance properties



Quoc V. Le

ImageNet classification

20,000 categories, 16,000,000 images

Hand-engineered features (SIFT, HOG, LBP), Spatial pyramid, SparseCoding/Compression, Kernel SVMs

20,000 is a lot of categories…

…
smoothhound, smoothhound shark, Mustelus mustelus
American smooth dogfish, Mustelus canis
Florida smoothhound, Mustelus norrisi
whitetip shark, reef whitetip shark, Triaenodon obseus
Atlantic spiny dogfish, Squalus acanthias
Pacific spiny dogfish, Squalus suckleyi
hammerhead, hammerhead shark
smooth hammerhead, Sphyrna zygaena
smalleye hammerhead, Sphyrna tudes
shovelhead, bonnethead, bonnet shark, Sphyrna tiburo
angel shark, angelfish, Squatina squatina, monkfish
electric ray, crampfish, numbfish, torpedo
smalltooth sawfish, Pristis pectinatus
guitarfish
roughtail stingray, Dasyatis centroura
butterfly ray
eagle ray
spotted eagle ray, spotted ray, Aetobatus narinari
cownose ray, cow-nosed ray, Rhinoptera bonasus
manta, manta ray, devilfish
Atlantic manta, Manta birostris
devil ray, Mobula hypostoma
grey skate, gray skate, Raja batis
little skate, Raja erinacea
…

Stingray

Mantaray

0.005%	9.5%	?
Random guess	State-of-the-art (Weston, Bengio '11)	Feature learning From raw pixels

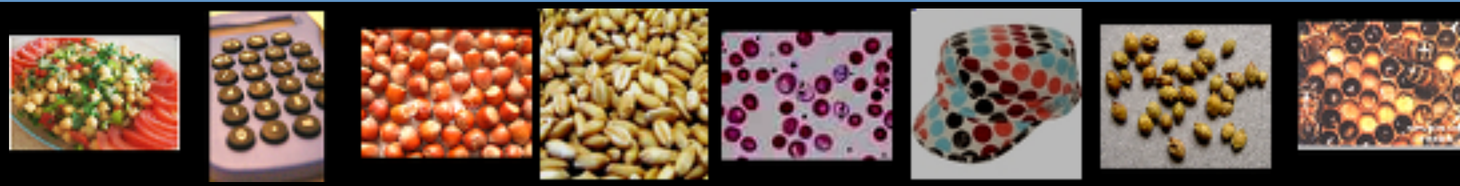
0.005%	9.5%	15.8%
Random guess	State-of-the-art (Weston, Bengio '11)	Feature learning From raw pixels

ImageNet 2009 (10k categories): Best published result: 17%
(Sanchez & Perronnin '11),
Our method: 19%

Using only 1000 categories, our method > 50%

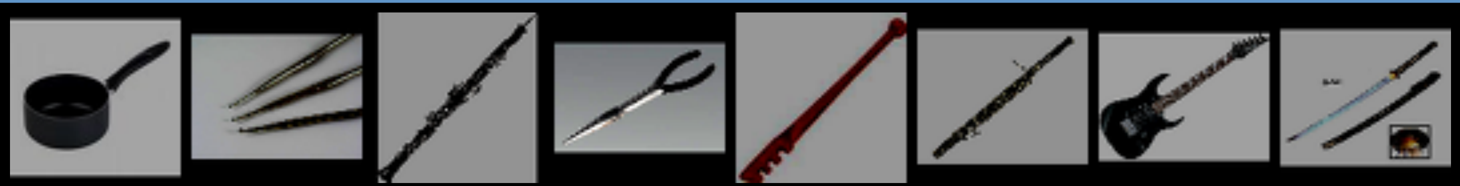
Feature 1

Feature 2



Feature 3

Feature 4



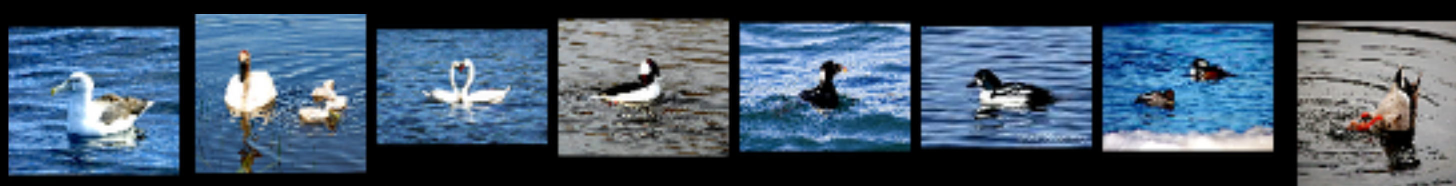
Feature 5

Feature 6

Feature 7

Feature 8

Feature 9



Quoc V. Le

Feature 10

Feature 11

Feature 12

Feature 13

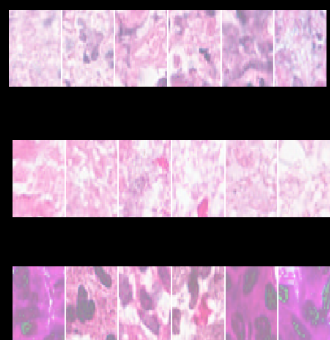
Quoc V. Le

Conclusions

- RICA learns invariant features
- Face neuron with totally unlabeled data with enough training and data
- State-of-the-art performances on
  - Action Recognition
  - Cancer image classification
  - ImageNet

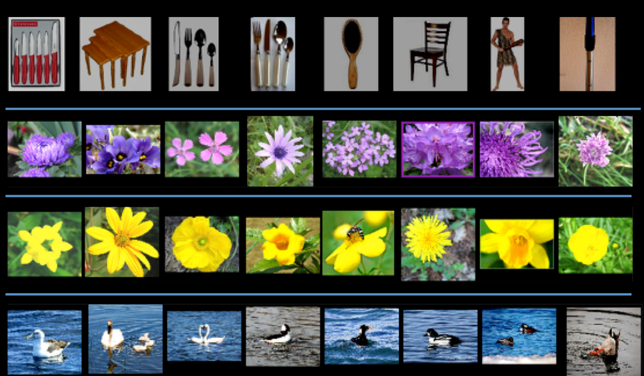
ImageNet

0.005%	9.5%	15.8%
Random guess	Best published result	Our method

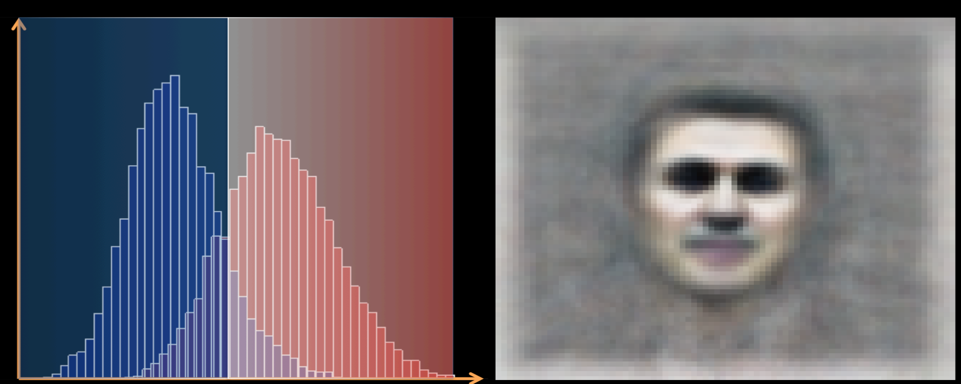


Cancer classification

Action recognition

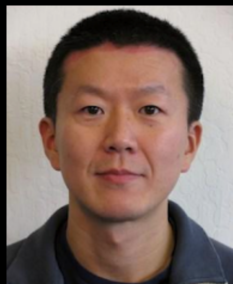


Feature visualization



Face neuron

Joint work with


Kai Chen


Greg Corrado


Jeff Dean


Matthieu Devin


Rajat Monga


Andrew Ng


Marc' Aurelio Ranzato


Paul Tucker


Ke Yang

<u>Additional Thanks:</u>

Samy Bengio, Zhenghao Chen, Tom Dean, Pangwei Koh, Mark Mao, Jiquan Ngiam, Patrick Nguyen, Andrew Saxe, Mark Segal, Jon Shlens, Vincent Vanhouke, Xiaoyun Wu, Peng Xe, Serena Yeung, Will Zou

References

- Q.V. Le, M.A. Ranzato, R. Monga, M. Devin, G. Corrado, K. Chen, J. Dean, A.Y. Ng. **Building high-level features using large-scale unsupervised learning**. *ICML*, 2012.
- Q.V. Le, J. Ngiam, Z. Chen, D. Chia, P. Koh, A.Y. Ng. **Tiled Convolutional Neural Networks**. *NIPS*, 2010.
- Q.V. Le, W.Y. Zou, S.Y. Yeung, A.Y. Ng. **Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis**. *CVPR*, 2011.
- Q.V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, A.Y. Ng. **On optimization methods for deep learning**. *ICML*, 2011.
- Q.V. Le, A. Karpenko, J. Ngiam, A.Y. Ng. **ICA with Reconstruction Cost for Efficient Overcomplete Feature Learning**. *NIPS*, 2011.
- Q.V. Le, J. Han, J. Gray, P. Spellman, A. Borowsky, B. Parvin. **Learning Invariant Features for Tumor Signatures**. *ISBI*, 2012.
- I.J. Goodfellow, Q.V. Le, A.M. Saxe, H. Lee, A.Y. Ng, **Measuring invariances in deep networks**. *NIPS*, 2009.

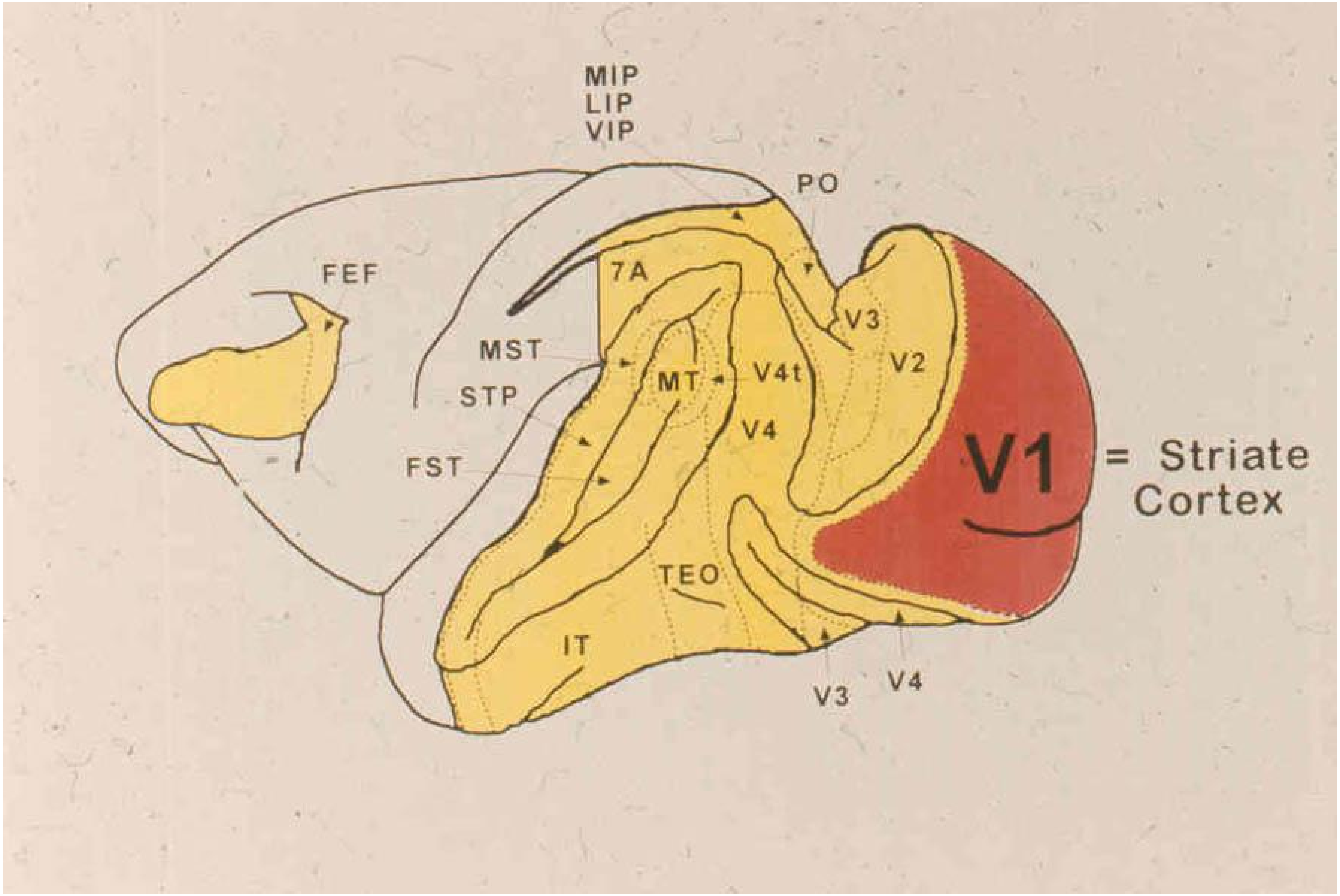
http://ai.stanford.edu/~quocle

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

[Cadieu *et al*., *PLoS Computational Biology* 2014]

(Slides by Sumeet Agarwal)

Left Cerebral Cortex of a Macaque

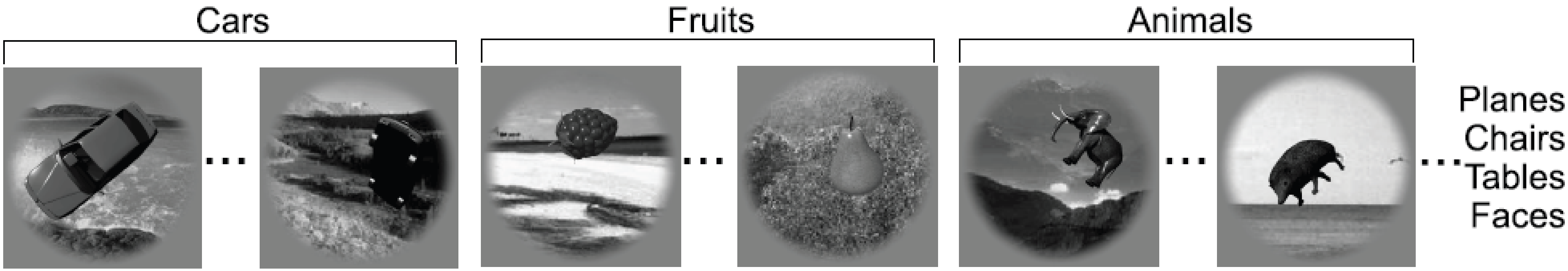


[Scholarpedia]

Visual Object Recognition

- How do we recognise objects despite variation in position, pose, scale, and background?
- Key problem in higher-order visual perception
- Need to create a **representation** (found in IT cortex for primates) that is selective for object identity and robust to variations
- Can computational models like neural networks learn such representations?

Data

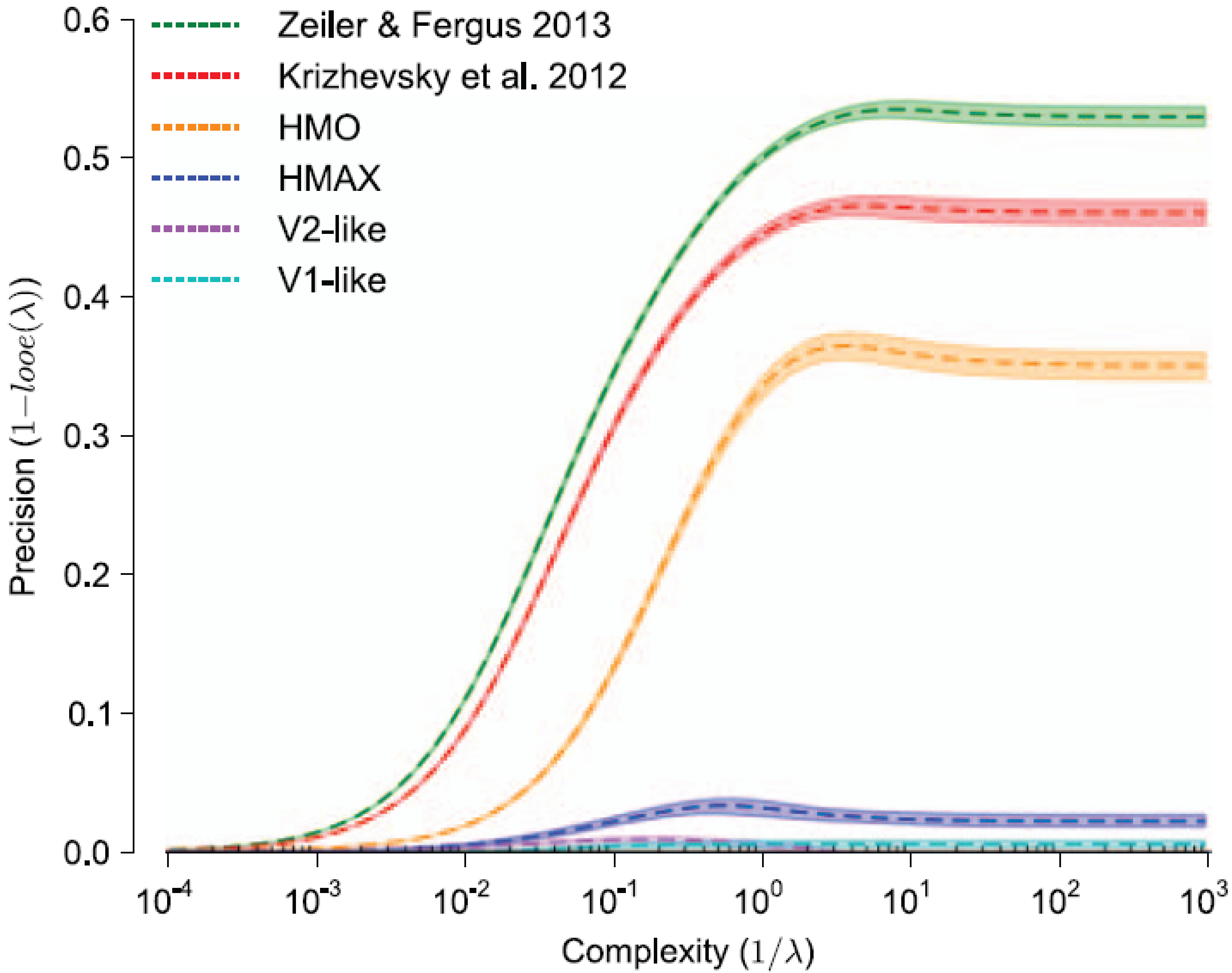


7 categories X 7 exemplars X 40 instances (varying position, scale, rotation/pose, and background) = 1960 images

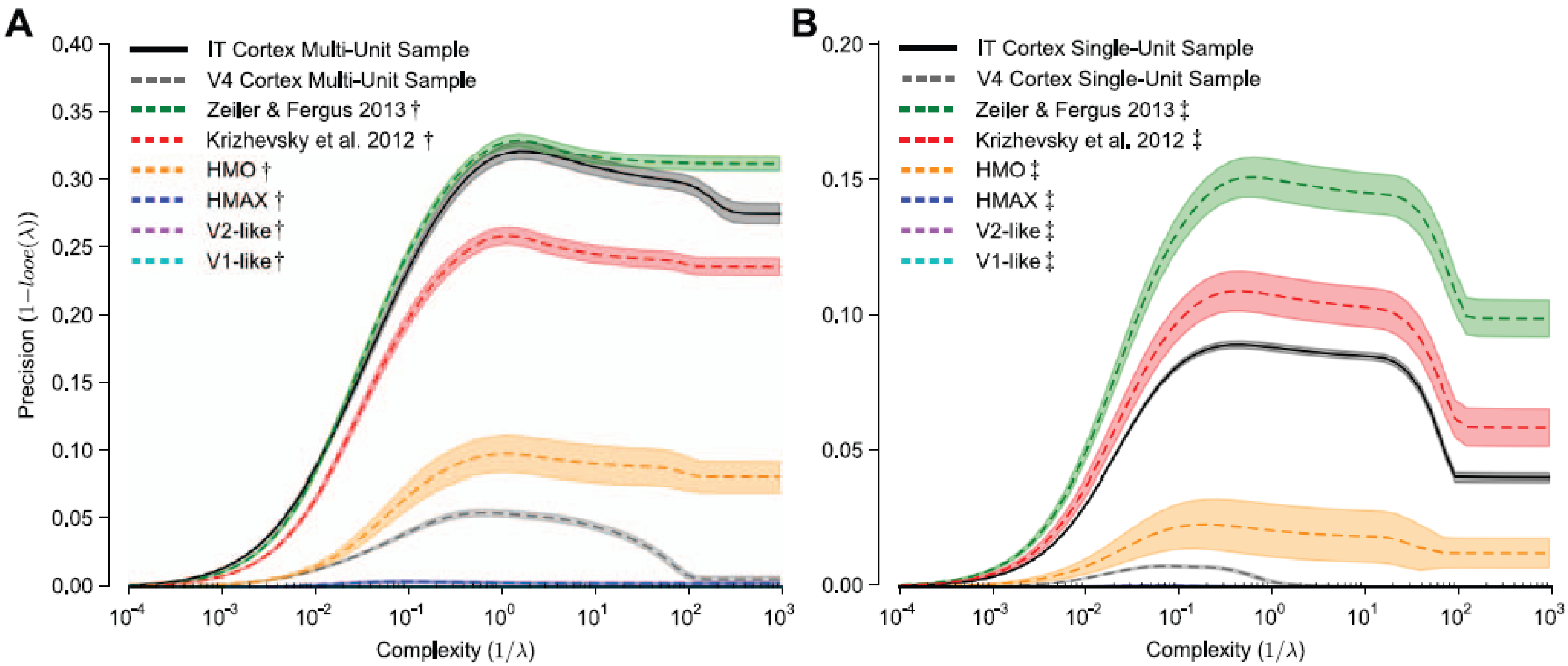
Approach

- Objective is to compare deep neural net representations with actual neural representations
- For actual representations, images shown to macaque monkeys and multi-unit and single-unit recordings (in IT cortex and V4 cortex) taken via a multi-electrode array
- Kernel analysis used to compare the performance of different representations for the object classification task (after equalising for noise and subsampling)

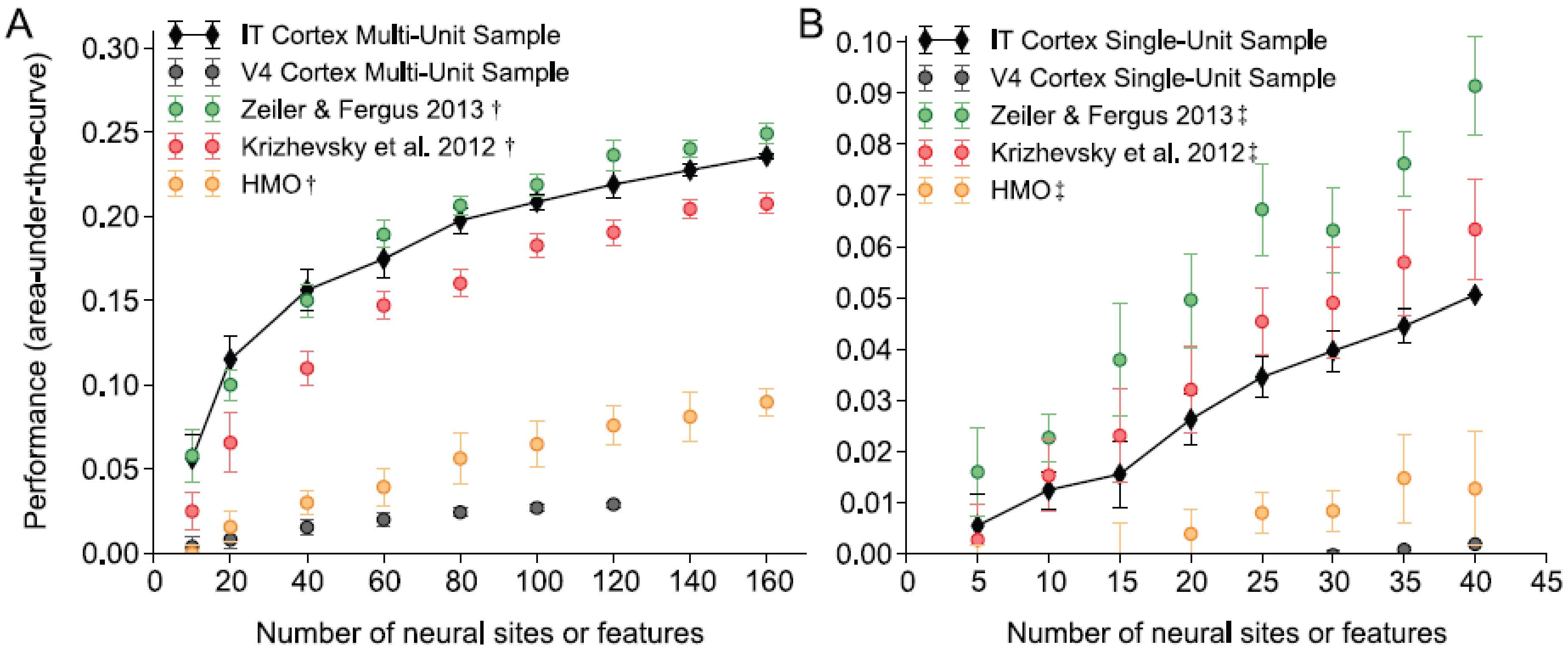
Kernel analysis curves of model representations



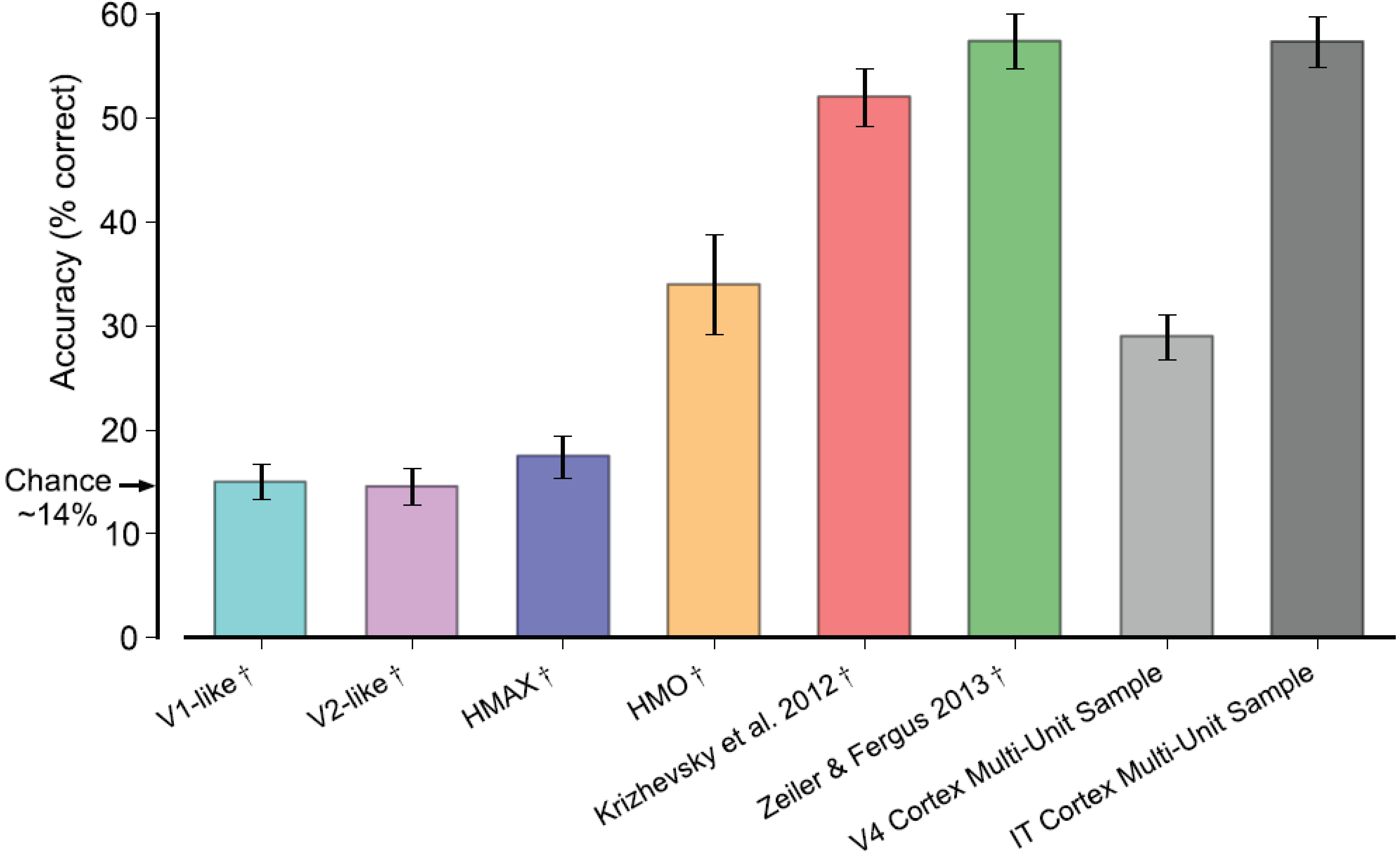
Kernel analysis comparison of model and neural representations



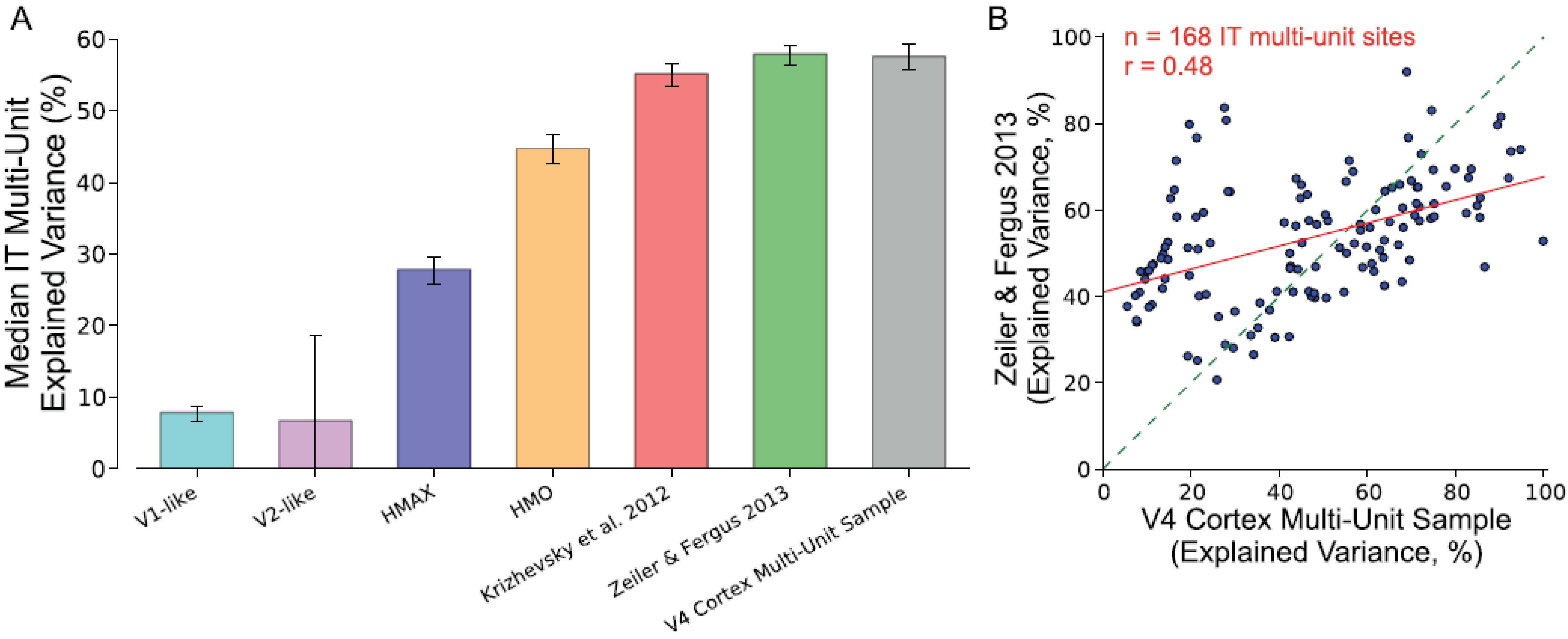
Sampling effects



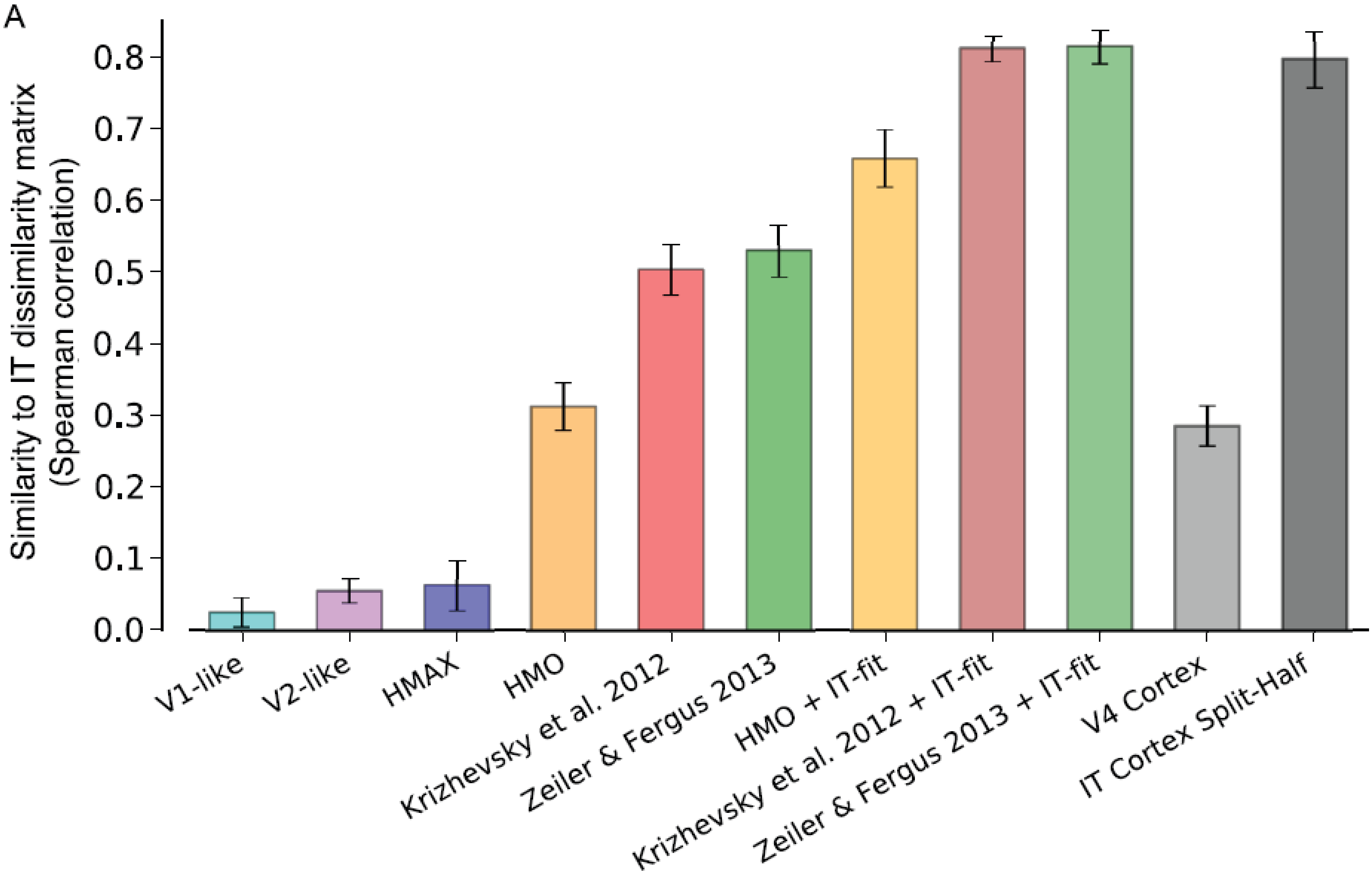
SVM classification performance

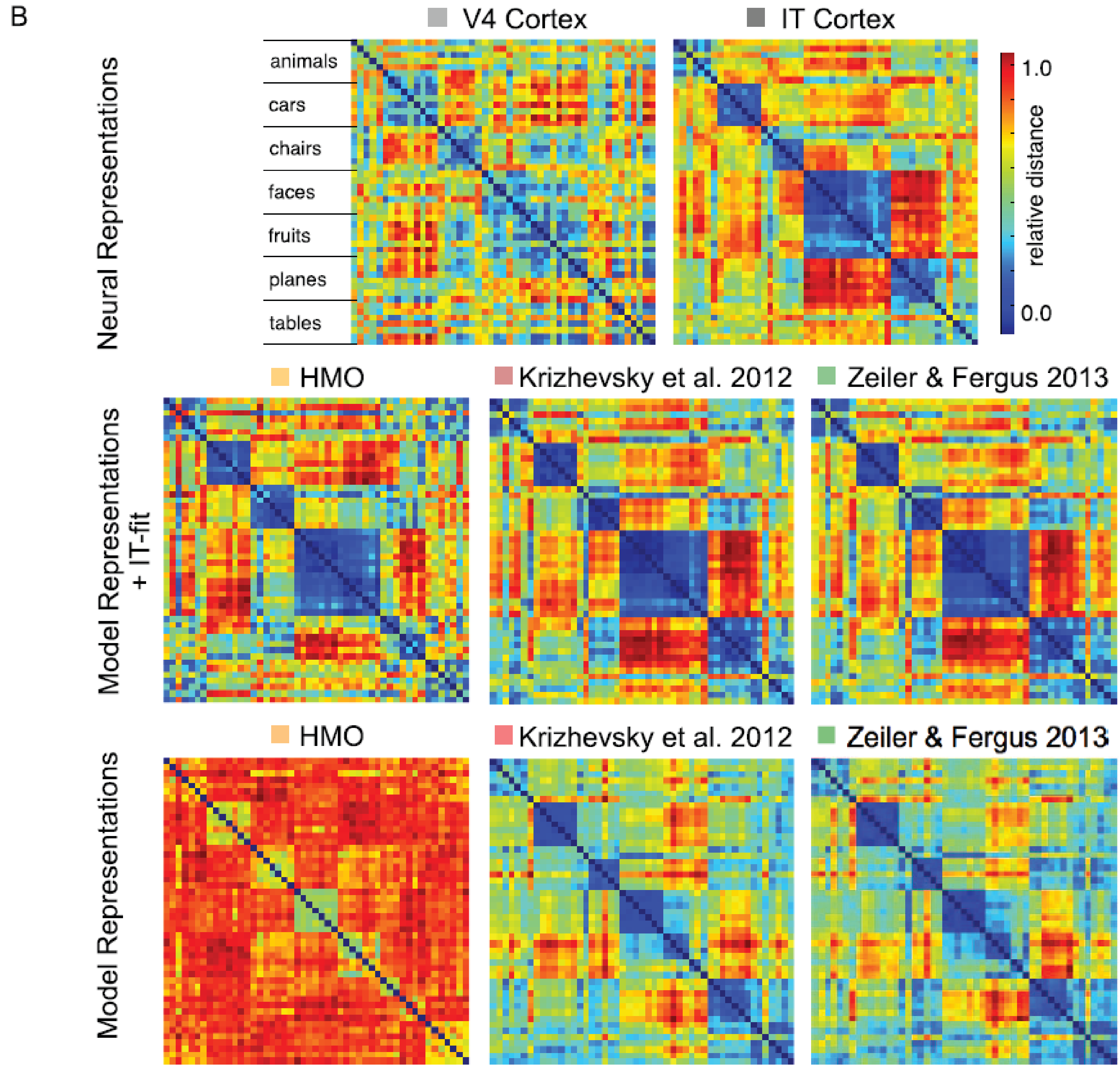


Predictability of IT cortex representations



Object-level representational similarity analysis




B
V4 Cortex
IT Cortex
Neural Representations
animals
cars
chairs
faces
fruits
planes
tables
1.0
relative distance
0.0
HMO
Krizhevsky et al. 2012
Zeiler & Fergus 2013
Model Representations + IT-fit
HMO
Krizhevsky et al. 2012
Zeiler & Fergus 2013
Model Representations

Our evaluations show that, unlike previous bio-inspired models, the latest DNNs rival the representational performance of IT cortex on this visual object recognition task. Furthermore, we show that models that perform well on measures of representational performance also perform well on measures of representational similarity to IT, and on measures of predicting individual IT multi-unit responses. Whether these DNNs rely on computational mechanisms similar to the primate visual system is yet to be determined, but, unlike all previous bioinspired models, that possibility cannot be ruled out merely on representational performance grounds.

Questions, thoughts, ideas, project positions in cognitive science (incl. PhD and post-doctoral fellowships):

- Sumeet Agarwal, EE IIT Delhi (sumeet@iitd.ac.in)
- Rajakrishnan Rajkumar, HSS IISER Bhopal (rajak@iiserb.ac.in)