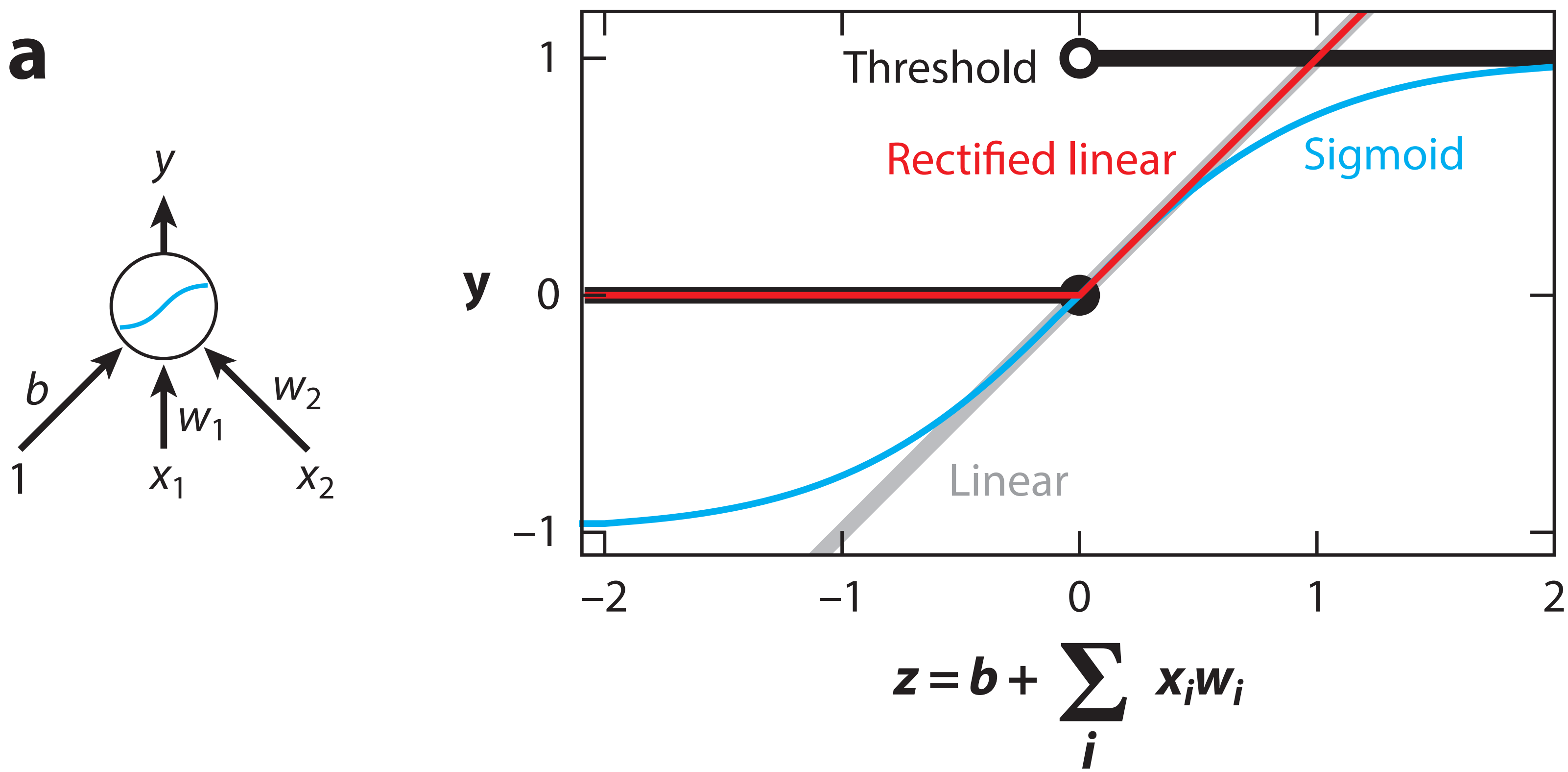
Discussion:

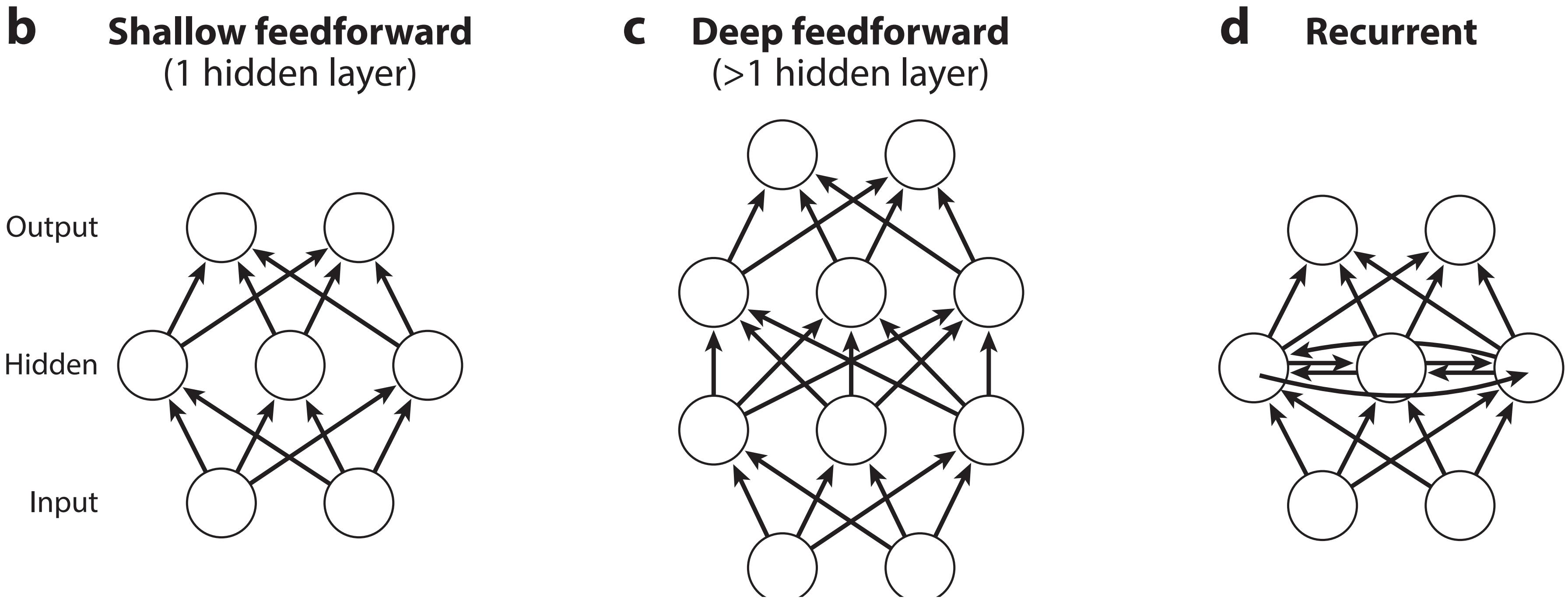
Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing

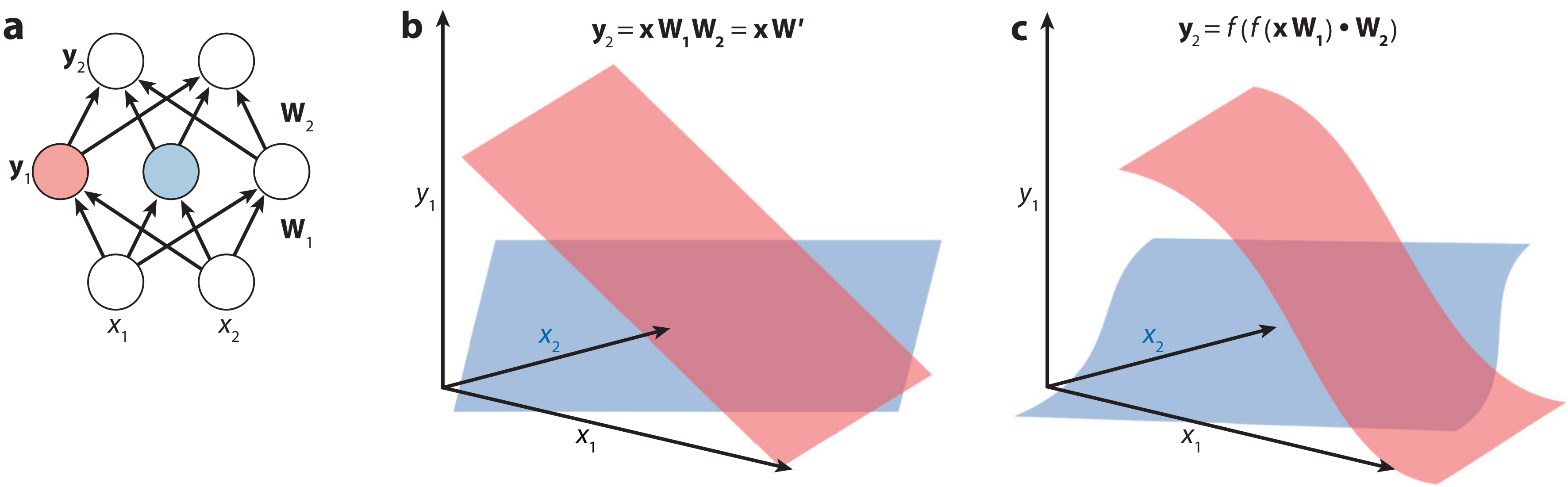
Nikolaus Kriegeskorte

Table 1 Historical progress toward understanding how the brain works

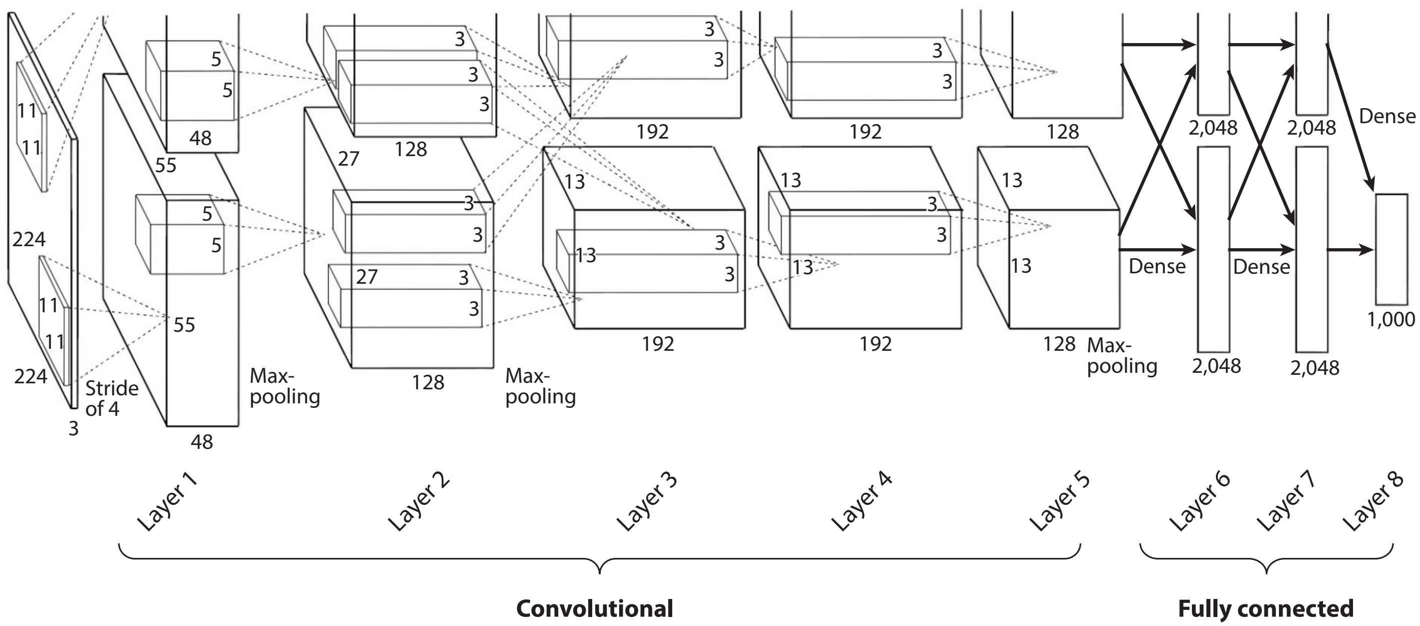
Elements required for understanding how the brain works		Behaviorism	Cognitive psychology	Cognitive science	Cognitive neuroscience	Classical computational neuroscience	Future cognitive computational neuroscience
Data	Behavioral	✓	✓	✓	✓	✓	✓
	Neurophysiological				✓	✓	✓
Theory	Cognitive		✓	✓	✓		✓
	Fully computationally explicit			✓		✓	✓
	Neurally plausible			✓		✓	✓
Explanation of real-world tasks requiring rich knowledge and complex computations			✓		✓		✓
Explanation of how high-level neuronal populations represent and compute							✓


a
y
b
w_1
w_2
1
x_1
x_2
Threshold
Rectified linear
Sigmoid
Linear
y
1
0
−1
−2
−1
0
1
2
$z = b + \sum_i x_i w_i$


b
Shallow feedforward
(1 hidden layer)
Output
Hidden
Input
c
Deep feedforward
(>1 hidden layer)
d
Recurrent



a
$\mathbf{y}_2$
$\mathbf{W}_2$
$\mathbf{y}_1$
$\mathbf{W}_1$
x_1
x_2
b
$\mathbf{y}_2 = \mathbf{x}\,\mathbf{W}_1\mathbf{W}_2 = \mathbf{x}\,\mathbf{W}'$
y_1
x_2
x_1
c
$\mathbf{y}_2 = f\,(f\,(\mathbf{x}\,\mathbf{W}_1) \bullet \mathbf{W}_2)$
y_1
x_2
x_1



11
11
224
224
3
Stride
of 4
5
5
48
55
55
48
Max-
pooling
3
3
3
27
128
27
3
128
Max-
pooling
3
3
192
13
13
3
3
192
3
3
192
13
13
3
3
192
128
13
13
128
Max-
pooling
2,048
2,048
2,048
2,048
Dense
Dense
Dense
1,000
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Layer 8
Convolutional
Fully connected

Layer 1

Learn weights and bias

One RF: 11 x 11 x 3
(3 color axes)

(11 x 11 x 3 weights and 1 bias term)

Layer 1

Learn weights and bias

One RF: 11 x 11 x 3
(3 color axes)

(11 x 11 x 3 weights and 1 bias term)

Total of 96 RFs (each convolved/replicated along all locations)

Layer 1

Learn weights and bias

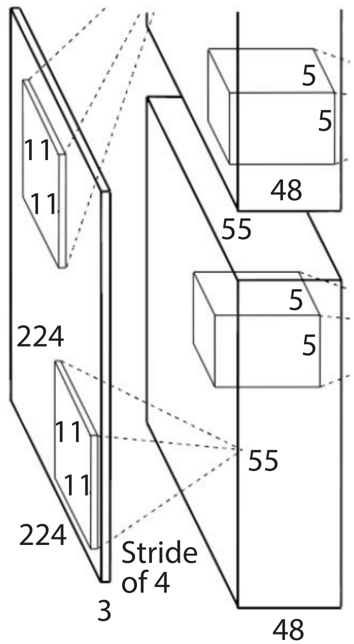
One RF: 11 x 11 x 3
(3 color axes)

(11 x 11 x 3 weights and 1 bias term)

Total of 96 RFs (each convolved/replicated along all locations)

Number parameters = (11*11*3)*96=35k

Layer 1

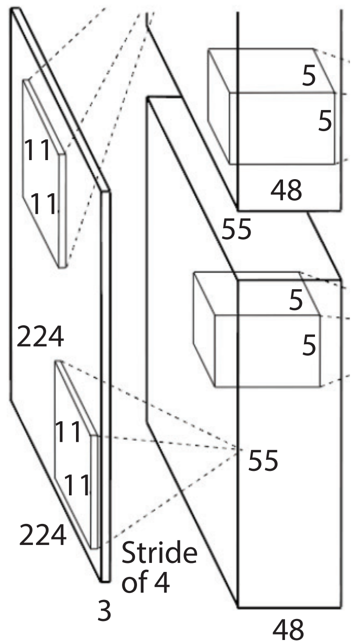


Stride of 4 between each location
(reduces from 227 x 227 to 55 x 55)
Note typo in original paper/figure;
size is 227 and not 224

(227-11)/4 + 1 = 55

RF size = 11
Stride = 4

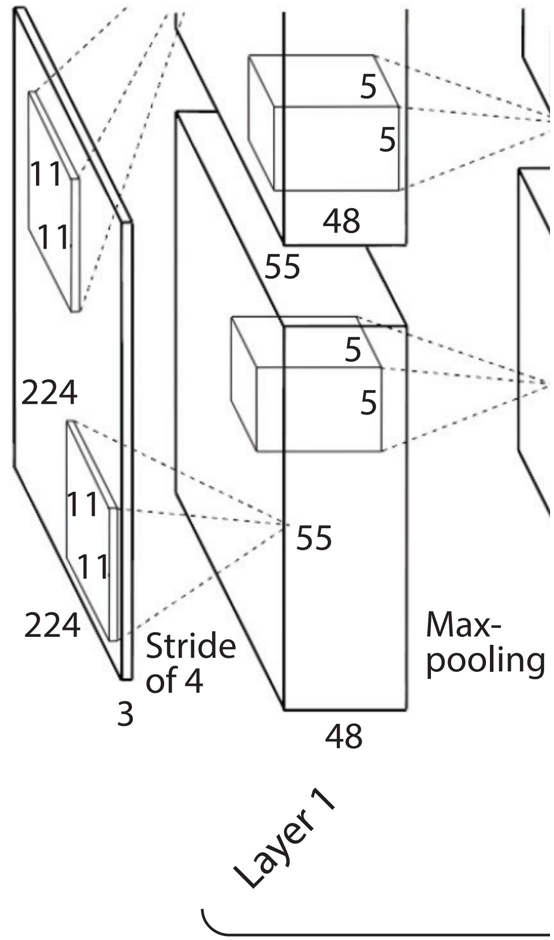
Layer 1



Stride of 4 between each location
(reduces from 227 x 227 to 55 x 55)
Note typo in original paper/figure
(227-11)/4 + 1 = 55

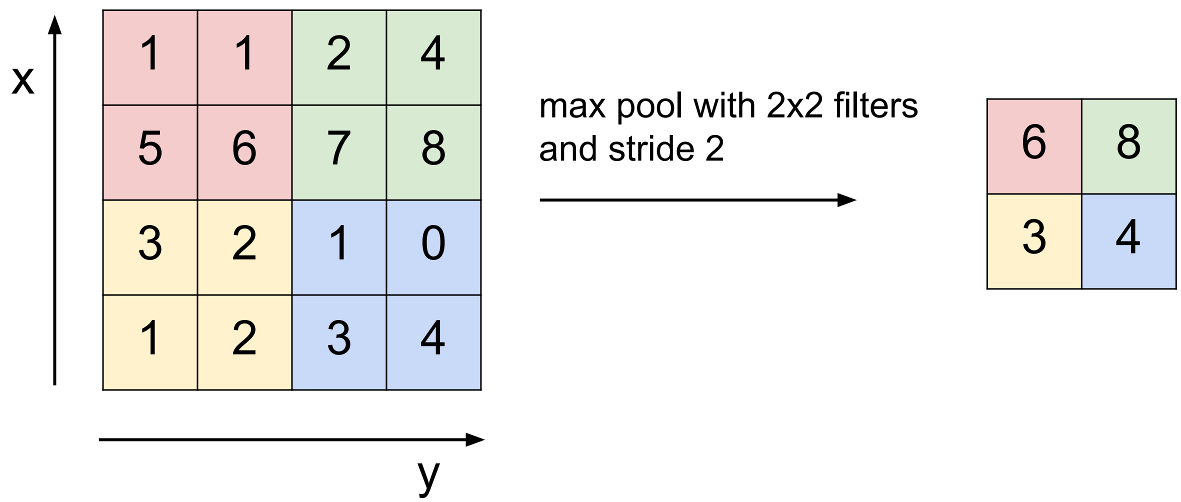
RF size = 11
Stride = 4

Conv 1 layer output: 55 x 55 x 96

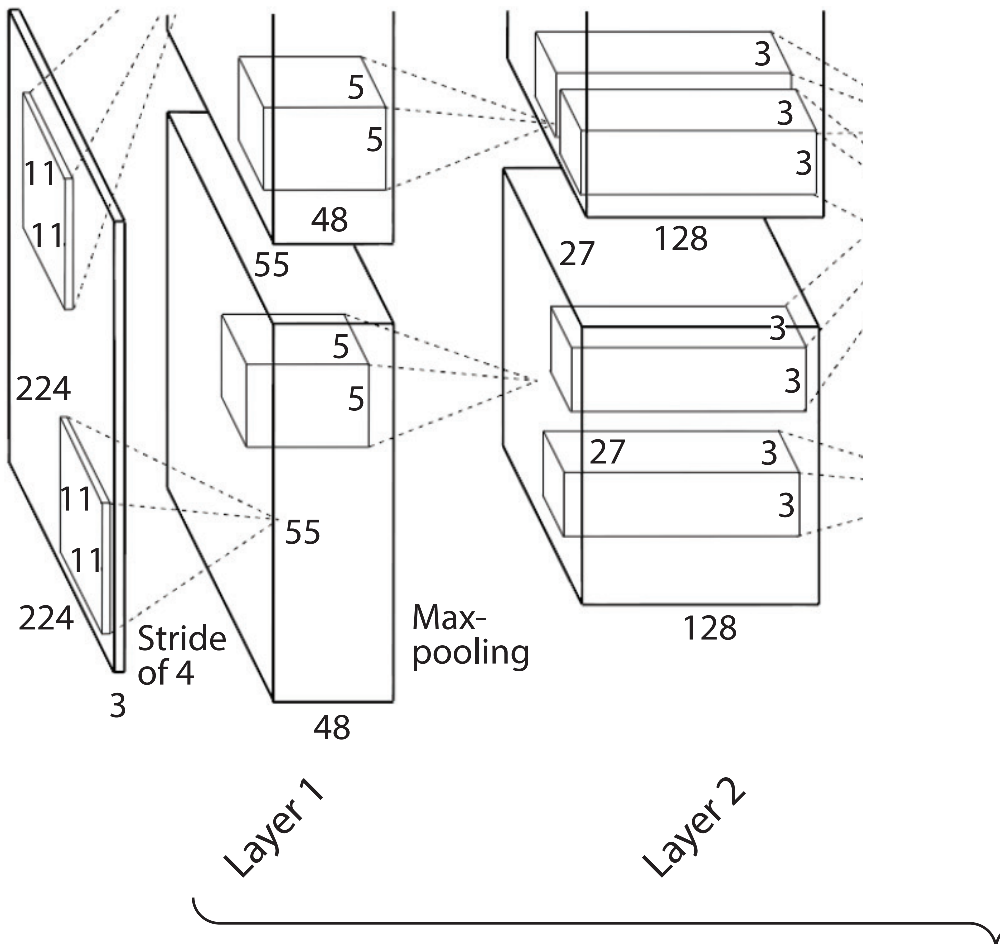


Max pooling (and then local normalization)

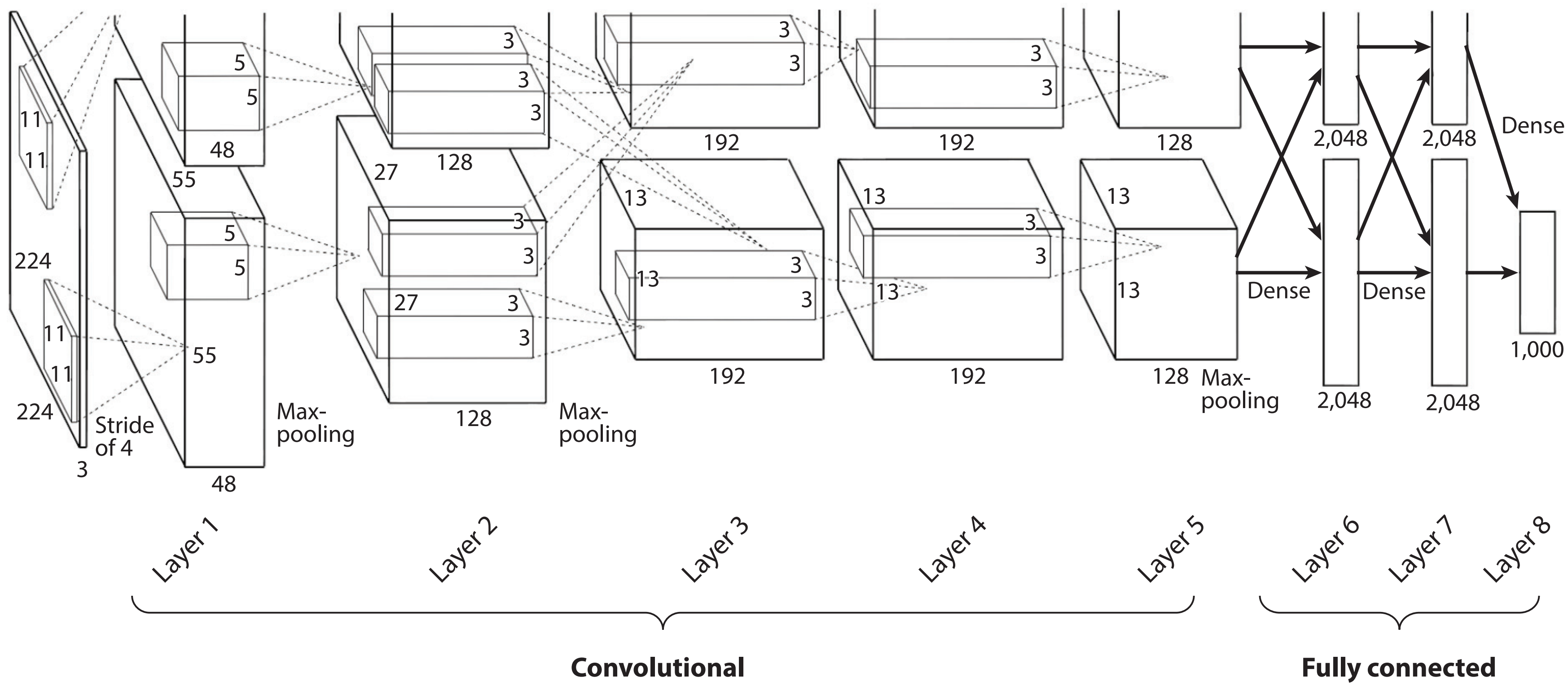
Convolutional Neural Networks: example of max pooling



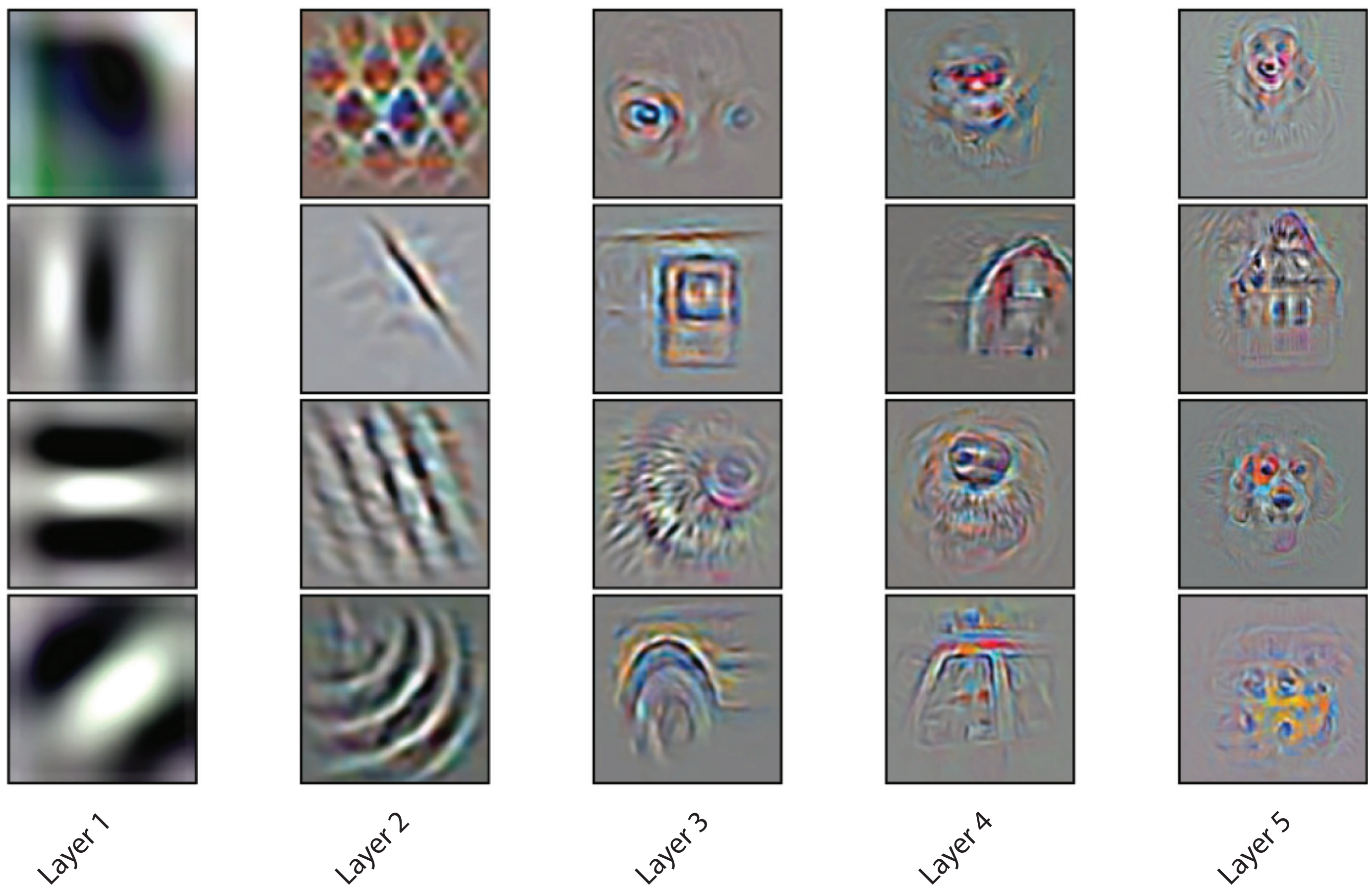
From http://cs231n.github.io/convolutional-networks/
Fei Fei, Karpathy, Johnson



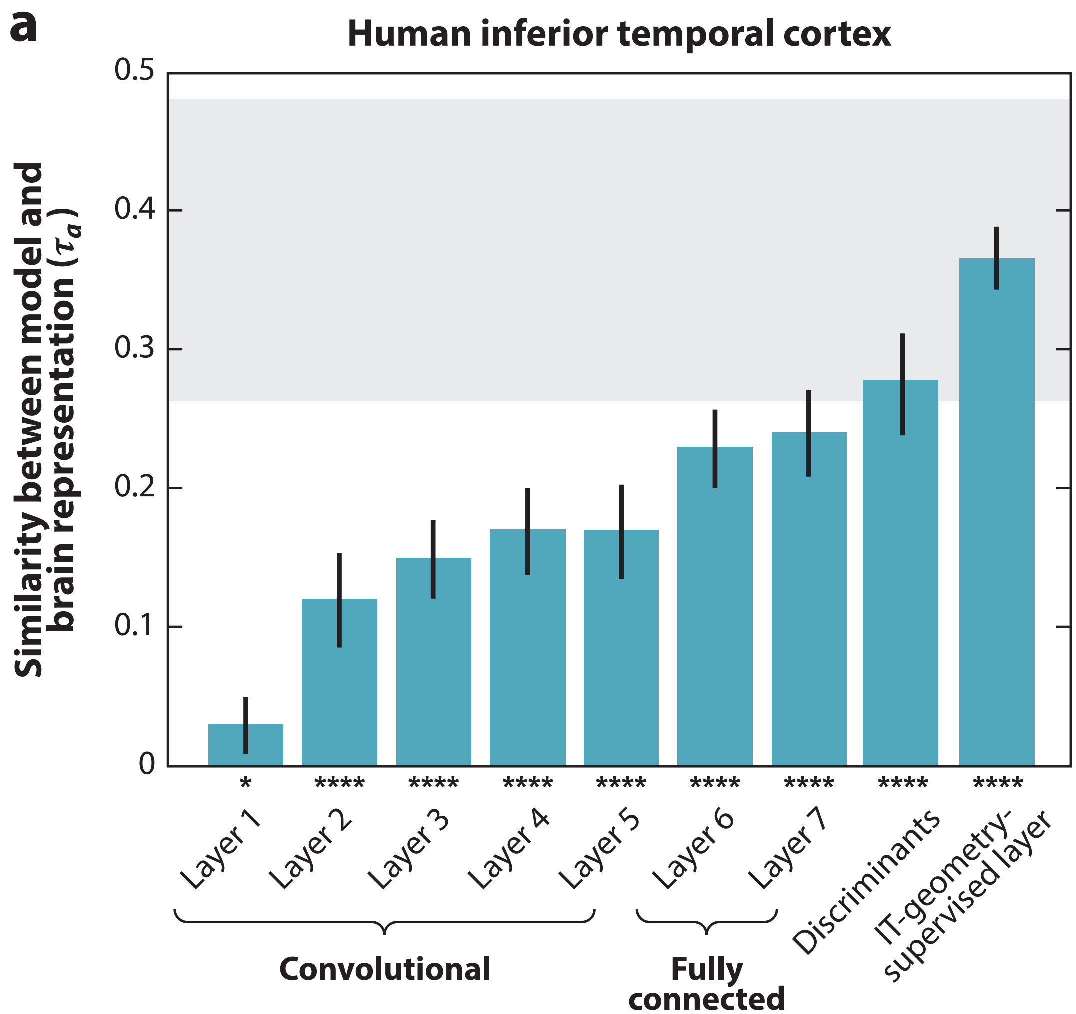
Max pooling:
3x3 filters
Stride 2
Output size:
(55-3)/2+1 = 27

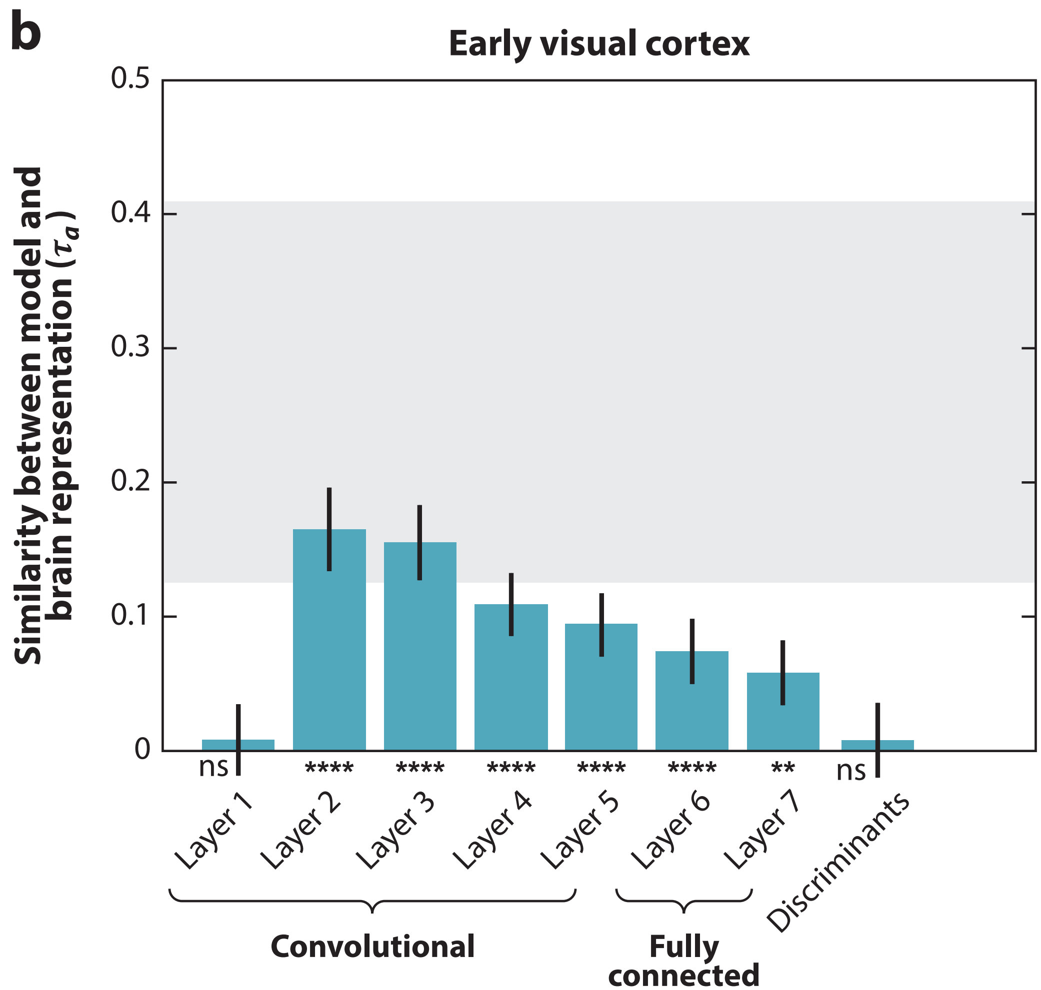


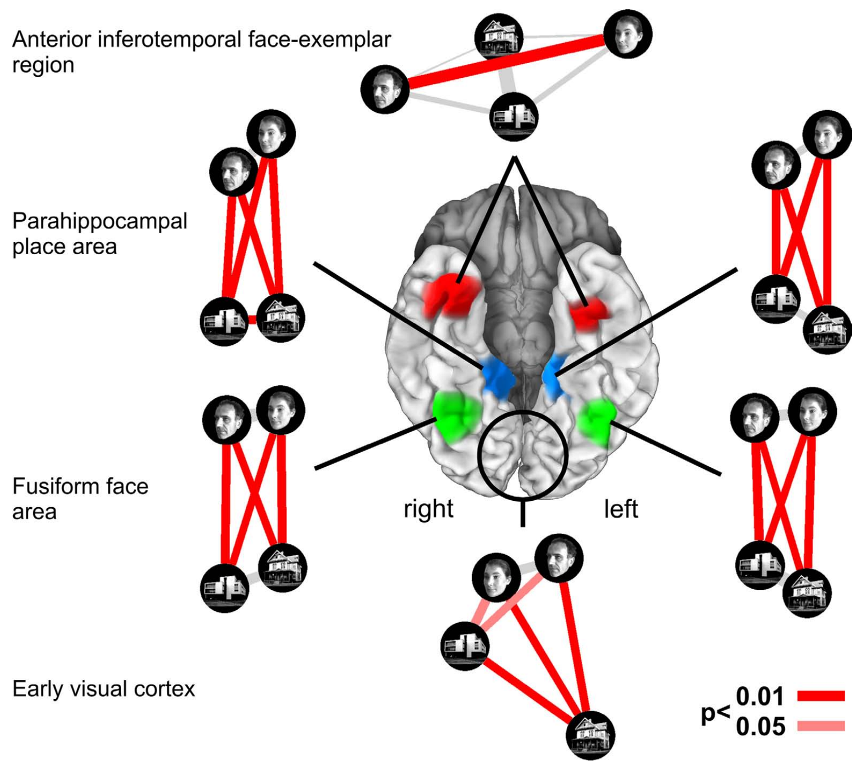
11
11
224
224
11
11
3
Stride of 4
5
5
48
55
5
5
55
48
Max-pooling
3
3
3
128
27
3
3
27
3
3
128
Max-pooling
3
3
192
13
3
13
3
192
3
3
192
13
3
13
3
192
128
13
13
128
Max-pooling
Dense
2,048
2,048
Dense
2,048
Dense
2,048
Dense
1,000
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Layer 8
Convolutional
Fully connected



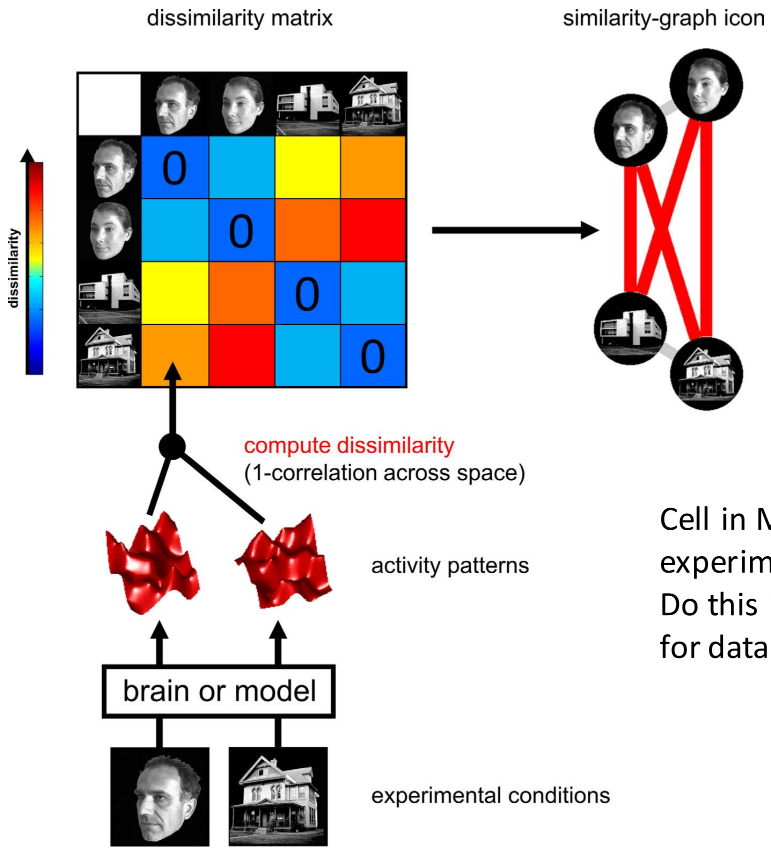
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5


a
Human inferior temporal cortex
Similarity between model and brain representation (τ_a)
0.5
0.4
0.3
0.2
0.1
0
*
****
****
****
****
****
****
****
****
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Discriminants
IT-geometry-supervised layer
Convolutional
Fully connected


b
Early visual cortex
Similarity between model and brain representation (τ_a)
0.5
0.4
0.3
0.2
0.1
0
ns
****
****
****
****
****
**
ns
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Discriminants
Convolutional
Fully connected

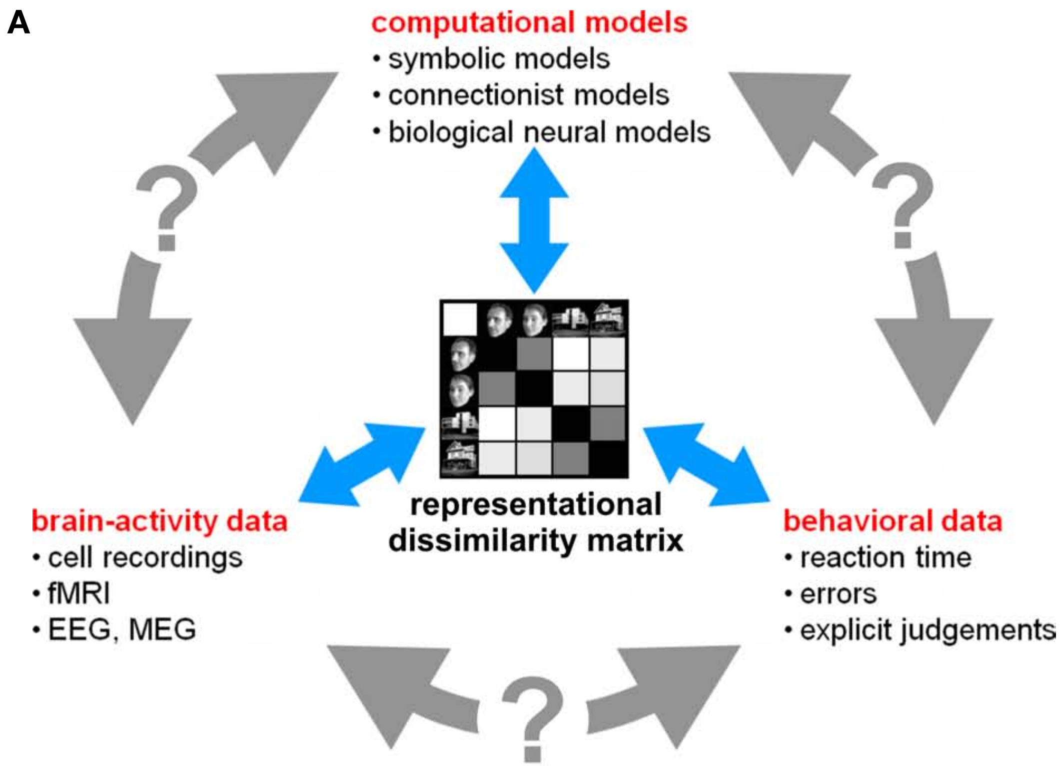


Images placed close together elicit similar response patterns; red line for significance (Kriegeskorte, 2008)

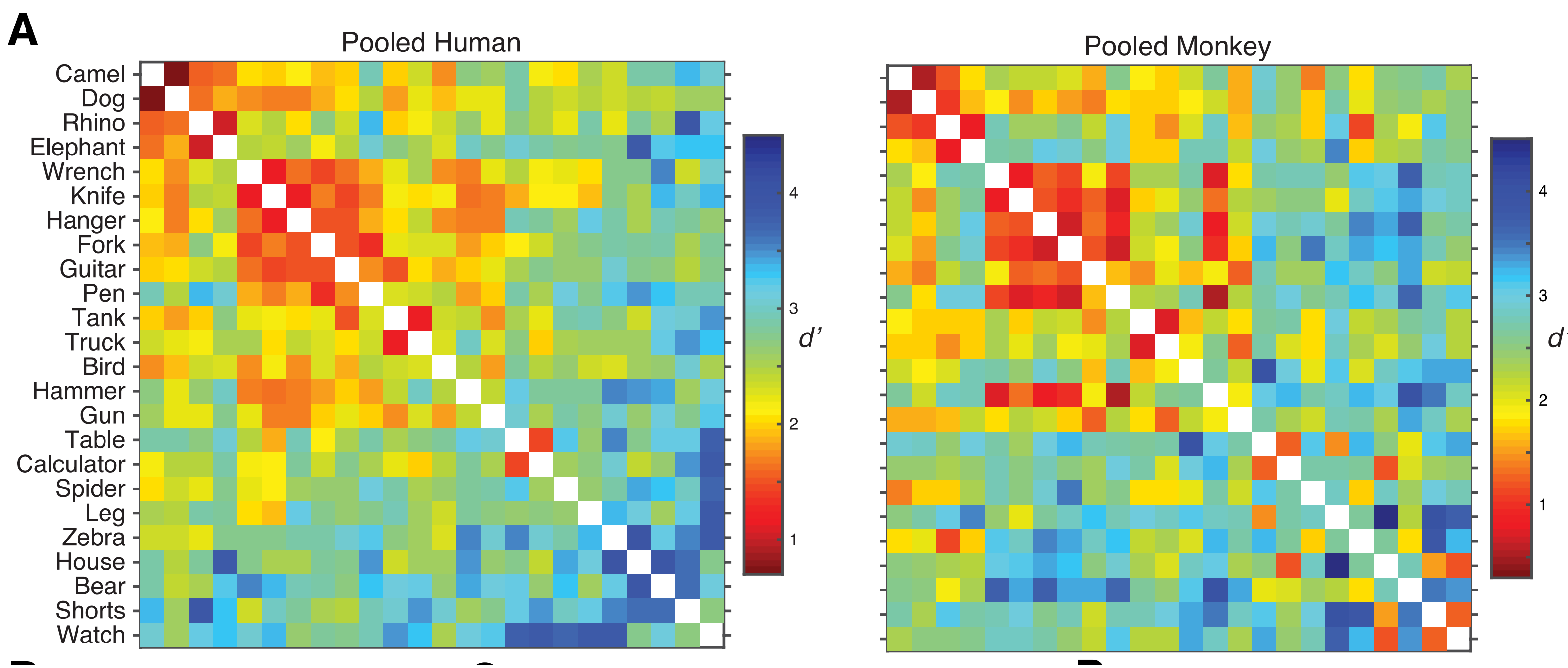


Cell in Matrix for each pair of experimental conditions.
Do this in model, and separately for data – and compare matrices

Associated activity patterns for a given image compared by spatial correlation (Kriegeskorte, 2008)



Kriegeskorte, 2008



Comparison of Object Recognition Behavior in Human and Monkey (IT): Rajalingham, Schmidt, and DiCarlo 2015

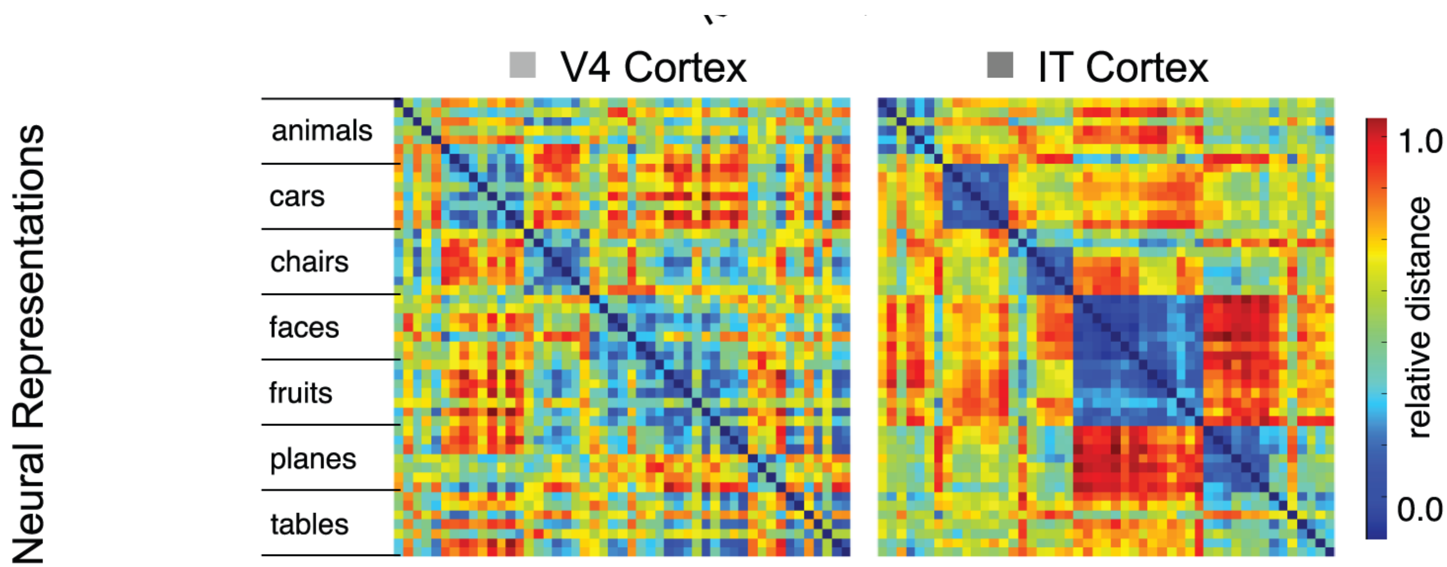


Figure 7. Object-level representational similarity analysis comparing model and neural representations to the IT multi-unit representation.

Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, et al. (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10(12): e1003963. doi:10.1371/journal.pcbi.1003963

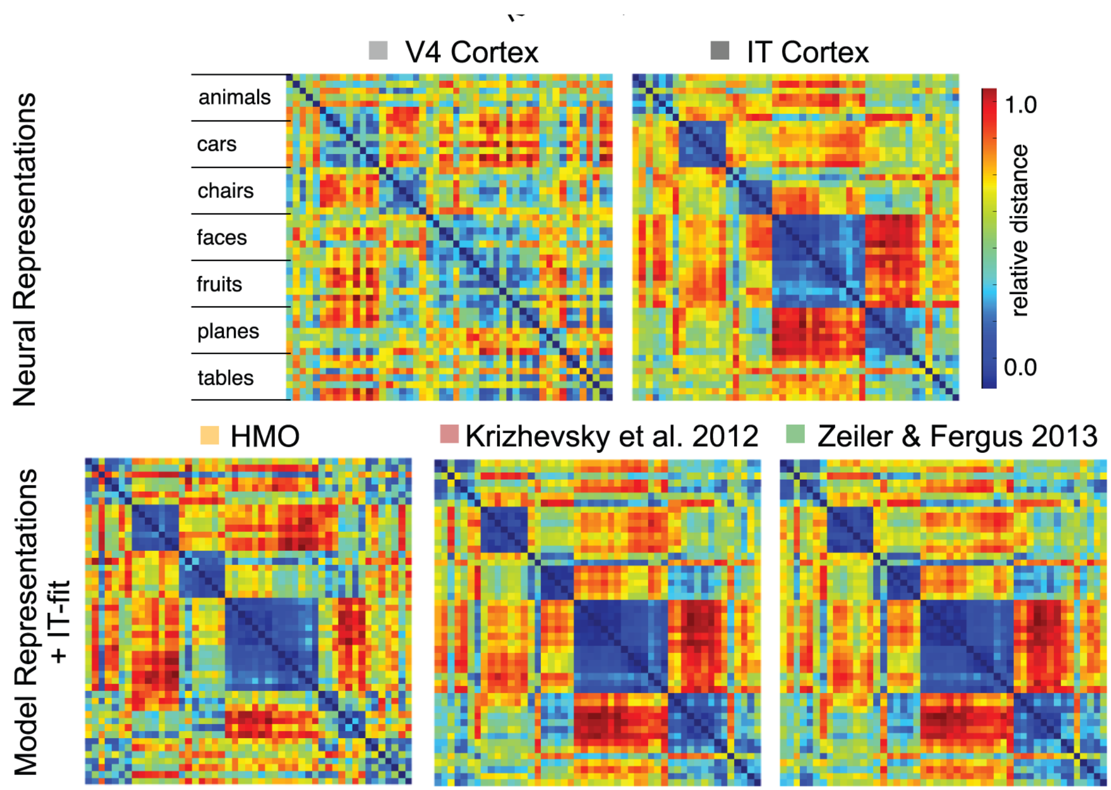


Figure 7. Object-level representational similarity analysis comparing model and neural representations to the IT multi-unit representation.

Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, et al. (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10(12): e1003963. doi:10.1371/journal.pcbi.1003963

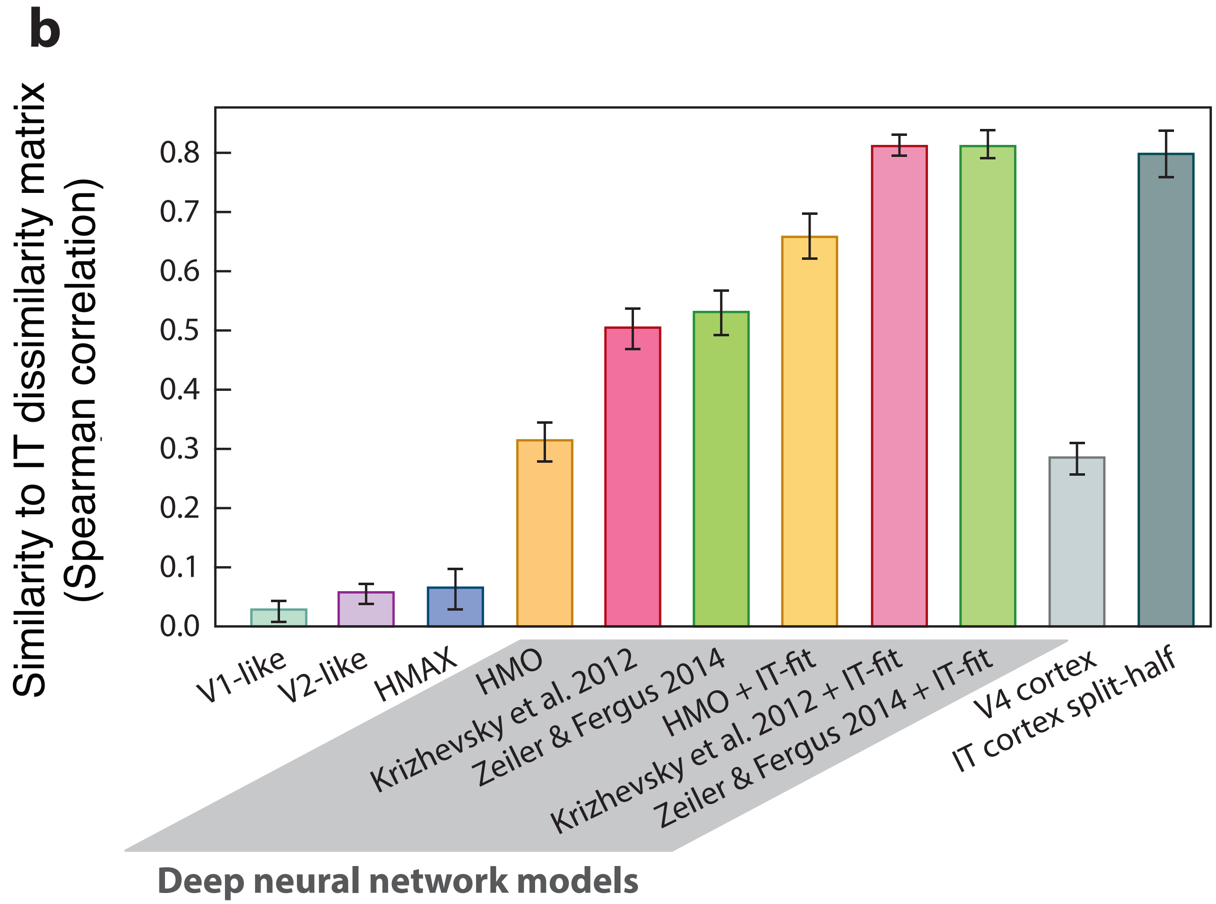
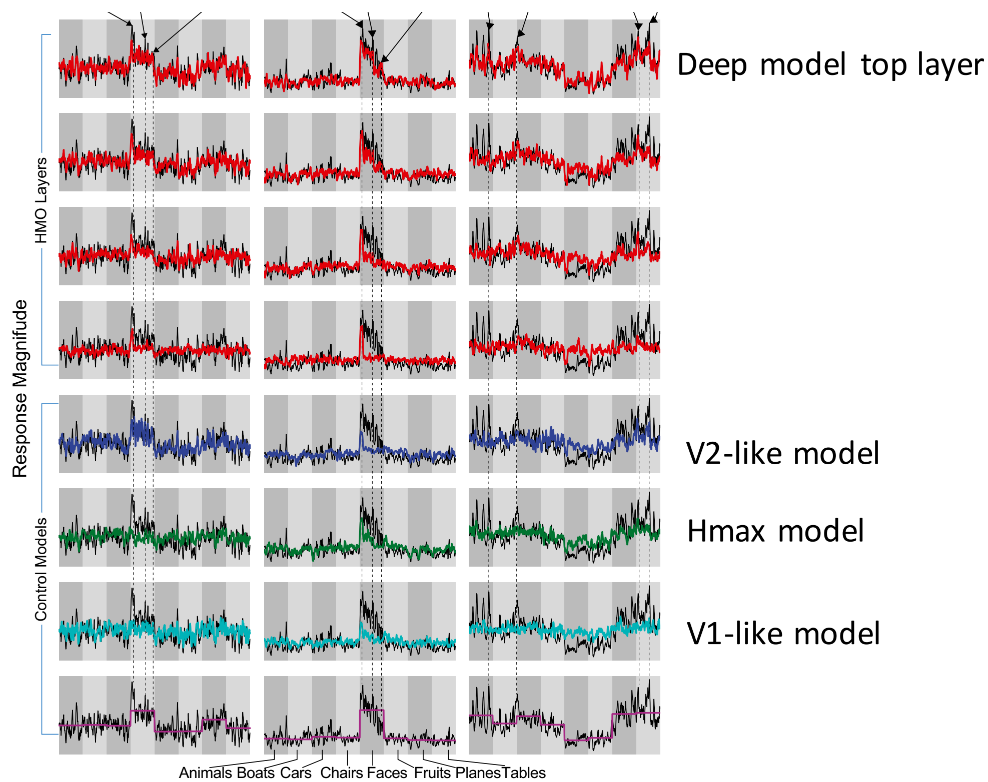


Figure from Cadieu et al. 2014 paper

Fitting neurons in IT



Yamins et al. 2014

Deep learning software

- Berkeley Caffe (visual models) ; now also Caffe2
- Google TensorFlow
- Theano
- Keras on top of TensorFlow, Theano
- Web browser demo:
  http://cs.stanford.edu/people/karpathy/convnetjs/index.html

All have Python interface, Caffe has Python/Matlab interface

Flexibility versus modifying existing frameworks

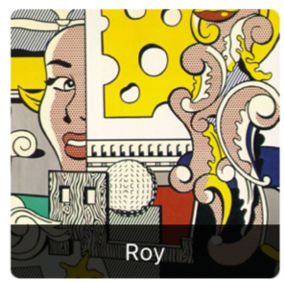
See some comparisons here:

http://deeplearning4j.org/compare-dl4j-torch7-pylearn.html

Deep learning in your phone app

See Gatys et al. 2015:
Separating content and style in a deep network

Deep learning in your phone app



See Gatys et al. 2015:
Separating content and style in a deep network

SUMMARY POINTS

1. Neural networks are brain-inspired computational models that now dominate computer vision and other AI applications.
2. Neural networks consist of interconnected units that compute nonlinear functions of their input. Units typically compute weighted combinations of their inputs followed by a static nonlinearity.
3. Feedforward neural networks are universal function approximators.
4. Recurrent neural networks are universal approximators of dynamical systems.
5. Deep neural networks stack multiple layers of nonlinear transformations and can concisely represent complex functions such as those needed for vision.
6. Convolutional neural networks constrain the input connections of units in early layers to local receptive fields with weight templates that are replicated across spatial positions. The restriction and sharing of weights greatly reduce the number of parameters that need to be learned.
7. Deep convolutional feedforward networks for object recognition are not biologically detailed and rely on nonlinearities and learning algorithms that may differ from those of biological brains. Nevertheless they learn internal representations that are highly similar to representations in human and nonhuman primate IT cortex.
8. Neural networks now scale to real-world AI tasks, providing an exciting technological framework for building more biologically faithful models of complex feats of brain information processing.

FUTURE ISSUES

1. We will build neural net models that engage complex real-world tasks and simultaneously explain biological brain-activity patterns and behavioral performance.
2. The models will have greater biological fidelity in terms of architectural parameters, nonlinear representational transformations, and learning algorithms.
3. Network layers should match the areas of the visual hierarchy in their response characteristics and representational geometries.
4. Models should predict a rich array of behavioral measurements, such as reaction times for particular stimuli in different tasks, similarity judgments, task errors, and detailed motor trajectories in continuous interactive tasks.
5. New supervised learning techniques will drive neural networks into alignment with measured functional and anatomical brain data and with behavioral data.
6. Recurrent neural network models will explain the representational dynamics of biological brains.
7. Recurrent neural network models will explain how feedforward, lateral, and feedback information flow interact to implement probabilistic inference on generative models of image formation.