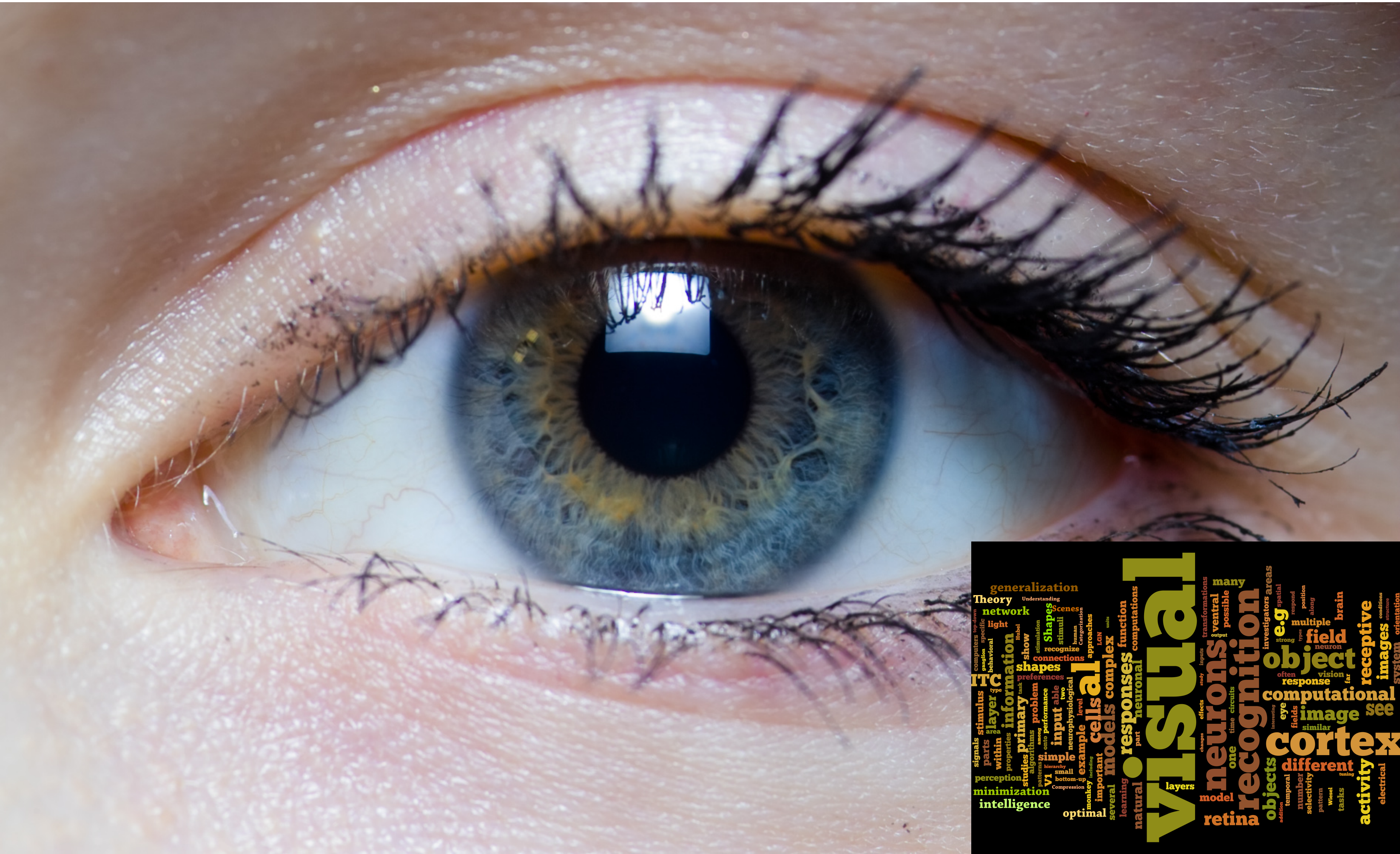
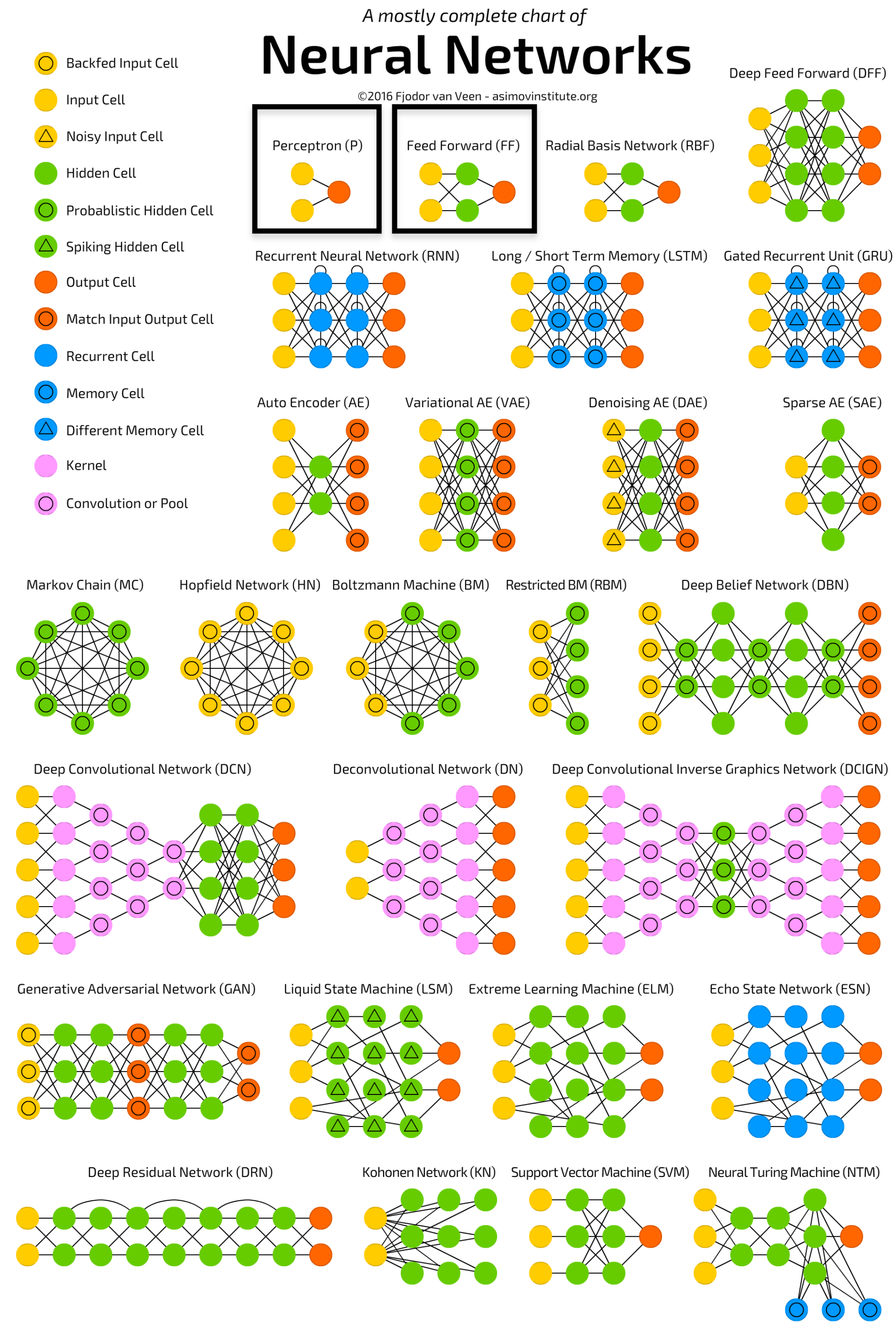
Visual Object Recognition
Computational Models and Neurophysiological Mechanisms
Neuro 130/230. Harvard College/GSAS 78454



Visual Object Recognition
Computational Models and Neurophysiological Mechanisms
Neurobiology 230. Harvard College/GSAS 78454

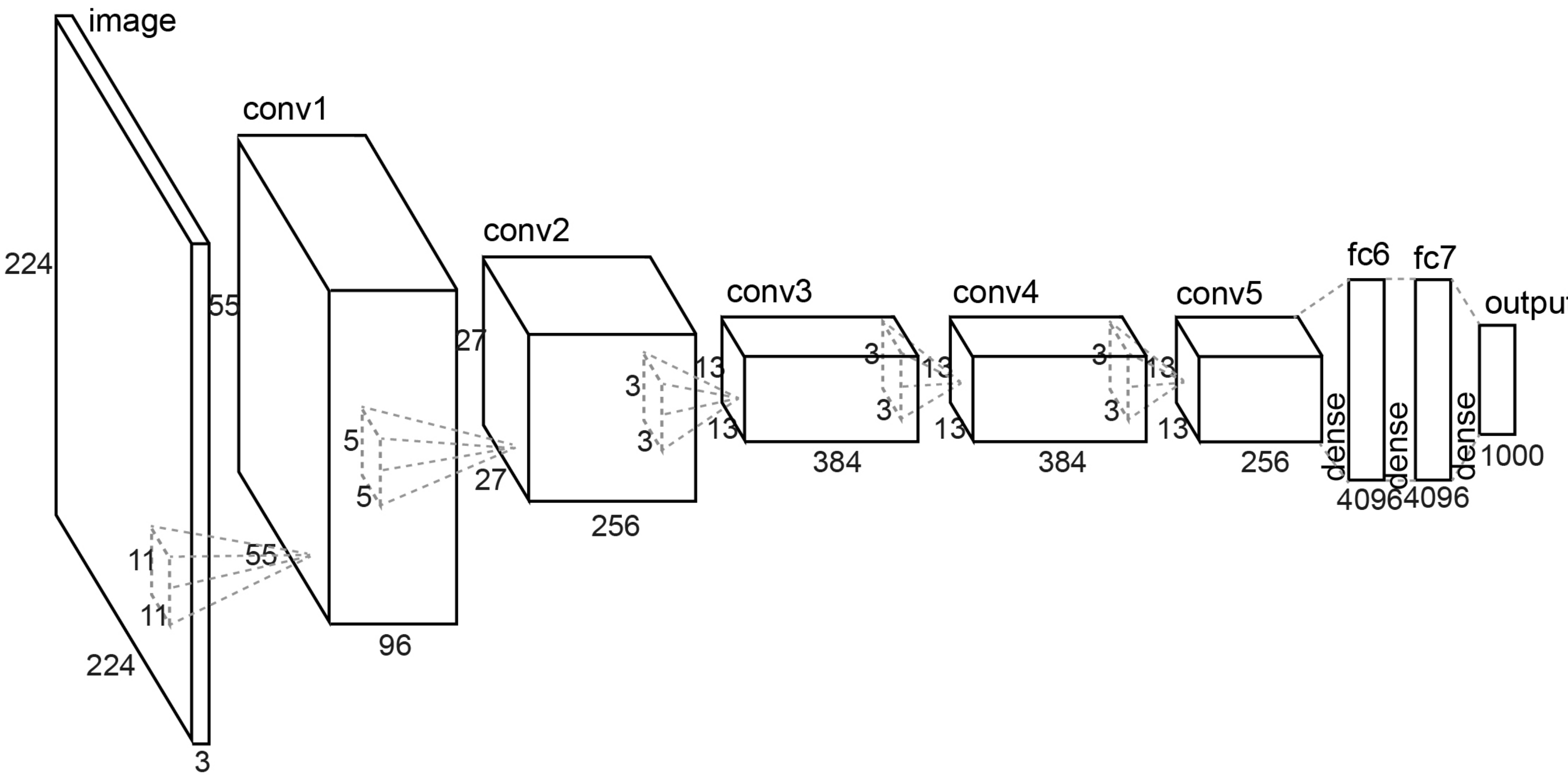
Class 1 [09/02/2020]. Introduction to Vision
Class 2 [09/14/2020]. Natural image statistics and the retina
Class 3 [09/21/2020]. The Phenomenology of Vision
Class 4 [09/28/2020]. Learning from Lesions
Class 5 [10/05/2020]. Primary Visual Cortex
October 12th: University Holiday
Class 6 [10/19/2020]. Adventures into *terra incognita*
Class 7 [10/26/2020]. From the Highest Echelons of Visual Processing to Cognition
Class 8 [11/02/2020]. First Steps into in silico vision
Class 9 [11/09/2020]. Teaching Computers how to see
Class 10 [11/16/2020]. Computer Vision
Class 11 [11/23/2020]. Connecting Vision to the rest of Cognition
Class 12 [11/30/2020]. Visual Consciousness
FINAL EXAM, PAPER DUE 12/14/2020. No extensions.

A big happy family of neural networks



https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

Deep convolutional neural networks: AlexNet



Krizhevsky et al 2012

Formulation of the visual recognition problem

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	38	43	105	255	253	253	253	253	253	174	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	43	139	224	226	252	253	252	252	252	252	252	252	158	14	0	0	0	0	0
0	0	0	0	0	0	0	0	0	178	252	252	252	252	253	252	252	252	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	109	252	252	230	132	133	132	132	189	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4	29	29	24	0	0	0	0	14	226	252	252	172	7	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	243	252	252	144	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88	189	252	252	252	14	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	212	247	252	252	252	204	9	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	32	125	193	193	193	253	252	252	252	238	102	28	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	222	252	252	252	252	253	252	252	252	177	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	223	253	253	253	253	255	253	253	253	253	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	31	123	52	44	44	44	44	143	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	5	75	9	0	0	0	0	0	0	98	242	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	61	183	252	29	0	0	0	0	18	92	239	252	252	243	65	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	147	134	134	134	134	203	253	252	252	188	83	0	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	252	252	252	252	252	252	253	230	153	8	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	49	157	252	252	252	252	252	217	207	146	45	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	7	103	235	252	172	103	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

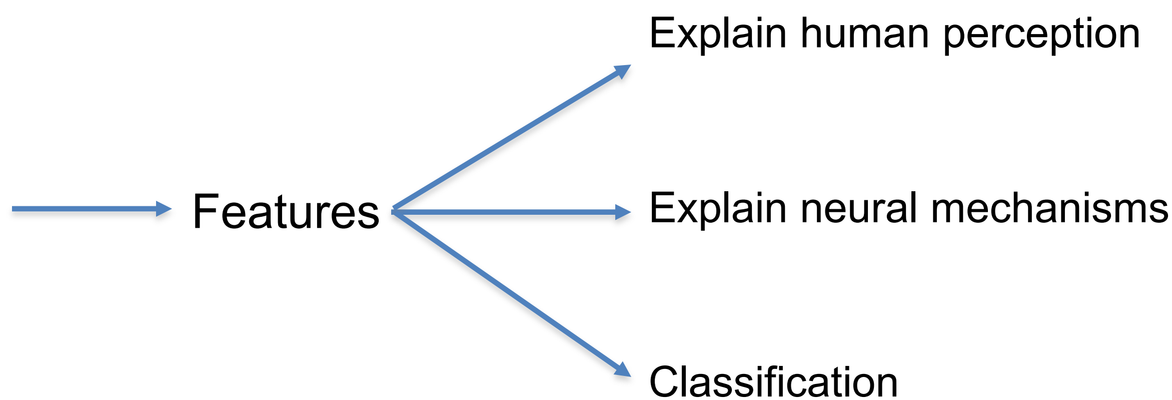
Image (or video)

→ Features → Classification

A more ambitious formulation

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	38	43	105	255	253	253	253	253	253	174	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	43	139	224	226	252	253	252	252	252	252	252	252	158	14	0	0	0	0	0
0	0	0	0	0	0	0	0	0	178	252	252	252	252	253	252	252	252	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	109	252	252	230	132	133	132	132	189	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4	29	29	24	0	0	0	0	14	226	252	252	172	7	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	243	252	252	144	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88	189	252	252	252	14	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	212	247	252	252	252	204	9	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	32	125	193	193	193	253	252	252	252	238	102	28	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	222	252	252	252	252	253	252	252	252	177	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	223	253	253	253	253	255	253	253	253	253	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	31	123	52	44	44	44	44	143	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	5	75	9	0	0	0	0	0	0	98	242	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	61	183	252	29	0	0	0	0	18	92	239	252	252	243	65	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	147	134	134	134	134	203	253	252	252	188	83	0	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	252	252	252	252	252	252	253	230	153	8	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	49	157	252	252	252	252	252	217	207	146	45	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	7	103	235	252	172	103	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Image (or video)



A brief history of computational models

Hubel and Wiesel, simple and complex cells (1950s')

Neocognitron (Fukushima 1980)

HMAX (Poggio 1999), Work on MNIST (LeCun 1998)

Deep convolutional neural networks (circa 2012)

Some of the typical computational operations

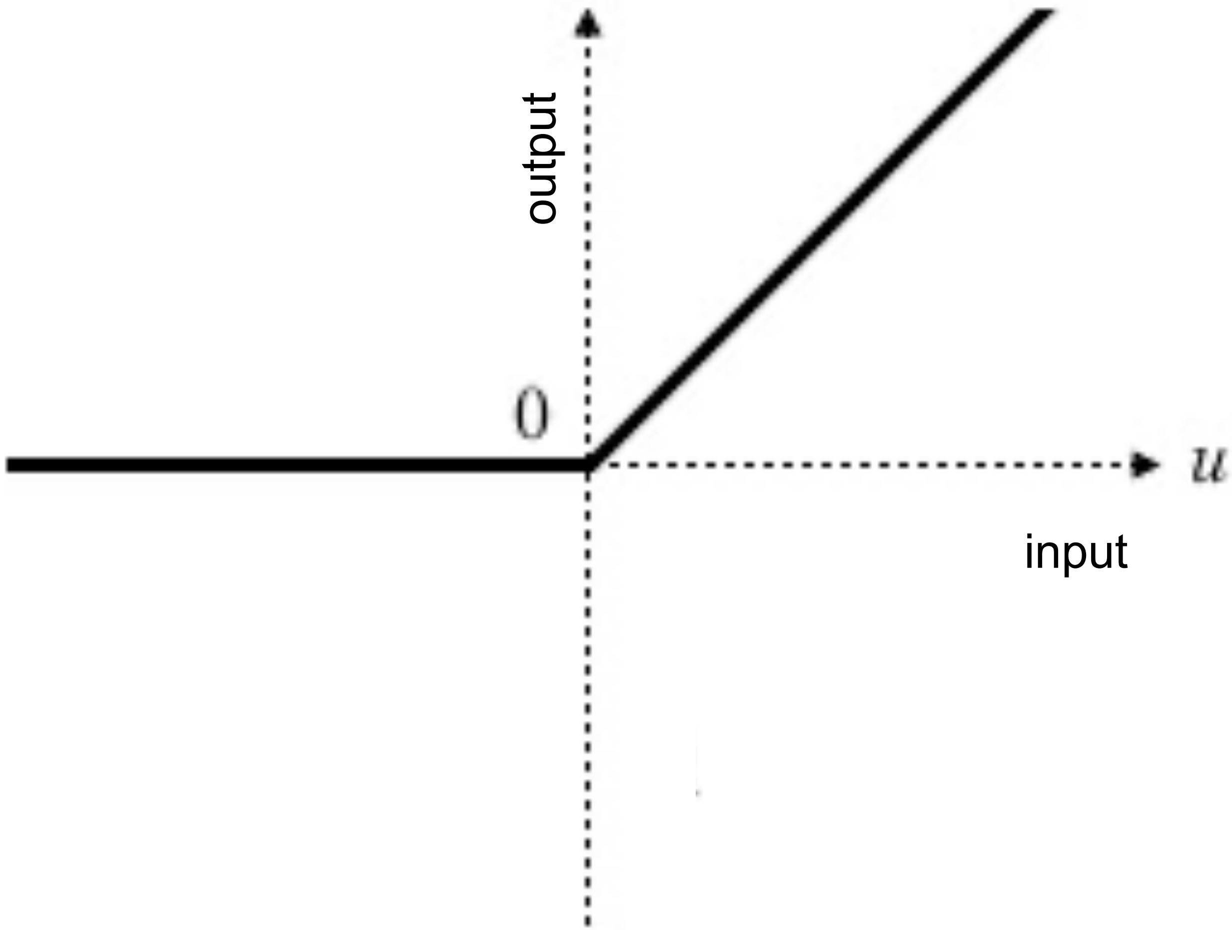
- Convolution
- Normalization
- ReLU
- Pooling

The convolution operation

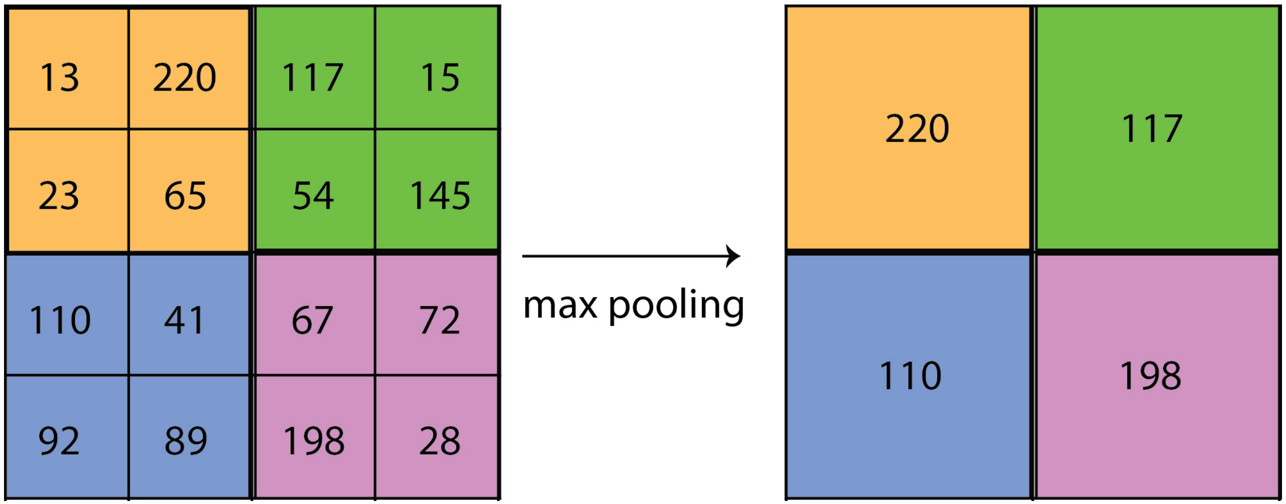
$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau$$

ReLU

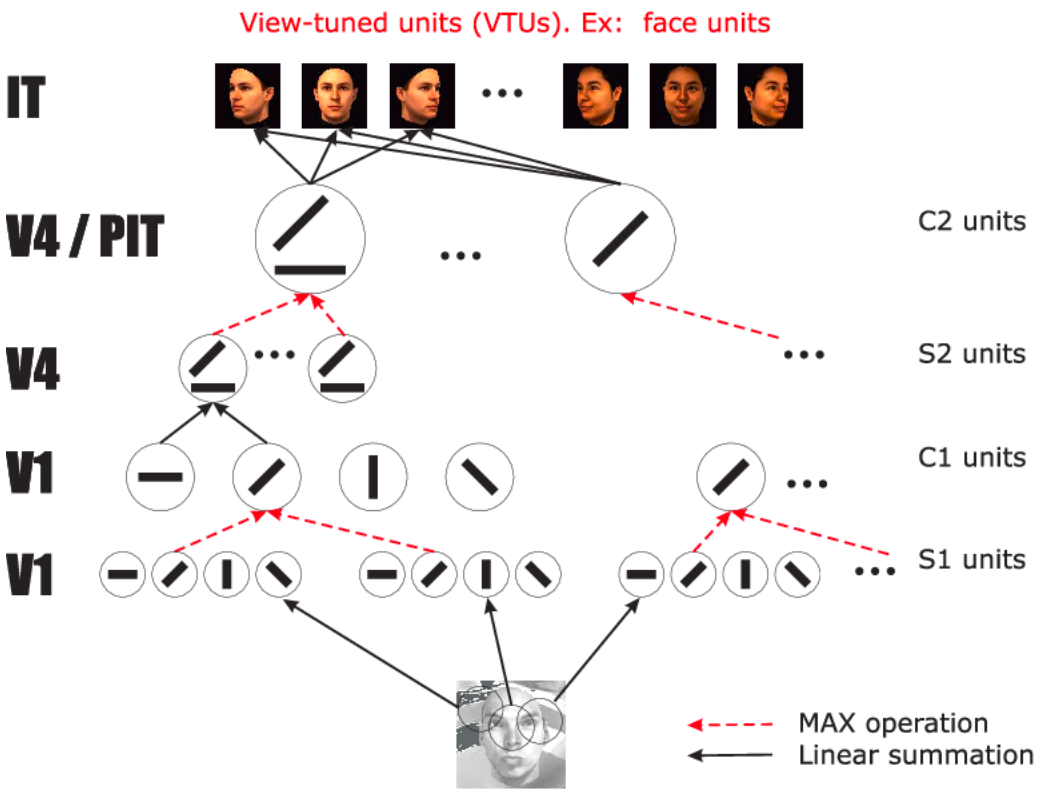
$$f(u) = \max(0, u)$$

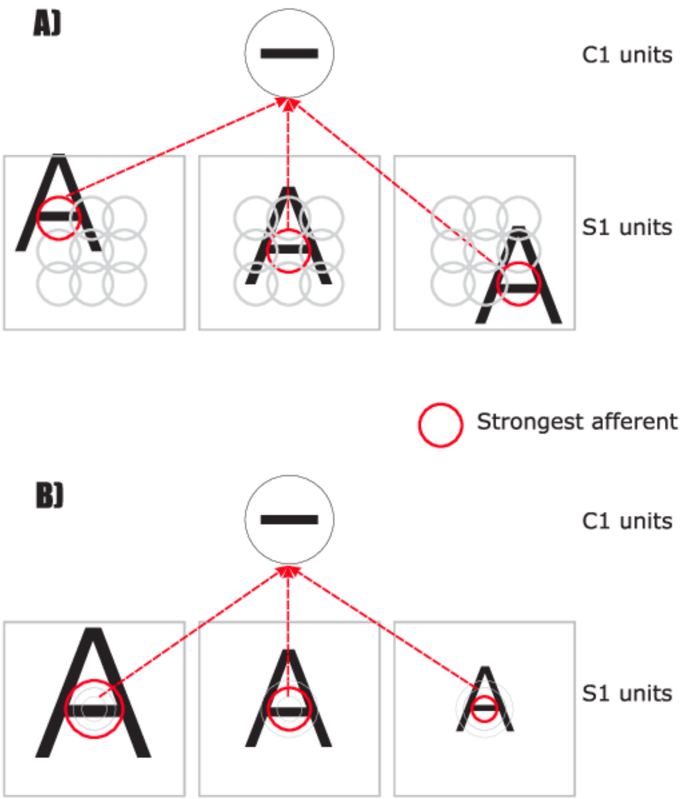


Max pooling


13
220
117
15
23
65
54
145
110
41
67
72
92
89
198
28
max pooling
220
117
110
198

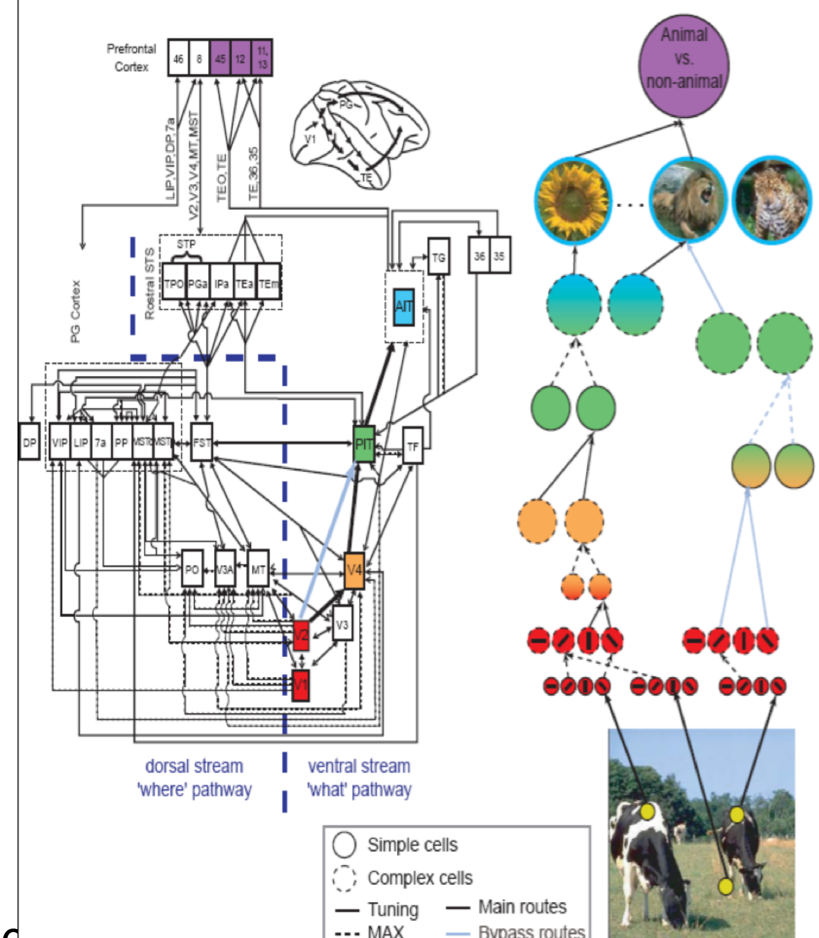
The HMAX model





Riesenhuber and Poggio 1999

The HMAX model



Model layers	Corresponding brain area (tentative)	RF sizes	Number units	
classifier	PFC		$1.0\ 10^{0}$	
S4	AIT	$>4.4^{\circ}$	$1.5\ 10^{2}$	~ 5,000 subunits
C3	PIT - AIT	$>4.4^{\circ}$	$2.5\ 10^{3}$	
C2b	PIT	$>4.4^{\circ}$	$2.5\ 10^{3}$	
S3	PIT	1.2°- 3.2°	$7.4\ 10^{4}$	~ 100 subunits
S2b	V4 - PIT	0.9°- 4.4°	$1.0\ 10^{7}$	~ 100 subunits
C2	V4	1.1°- 3.0°	$2.8\ 10^{5}$	
S2	V2 - V4	0.6°- 2.4°	$1.0\ 10^{7}$	~ 10 subunits
C1	V1 - V2	0.4°- 1.6°	$1.2\ 10^{4}$	
S1	V1 - V2	0.2°- 1.1°	$1.6\ 10^{6}$	

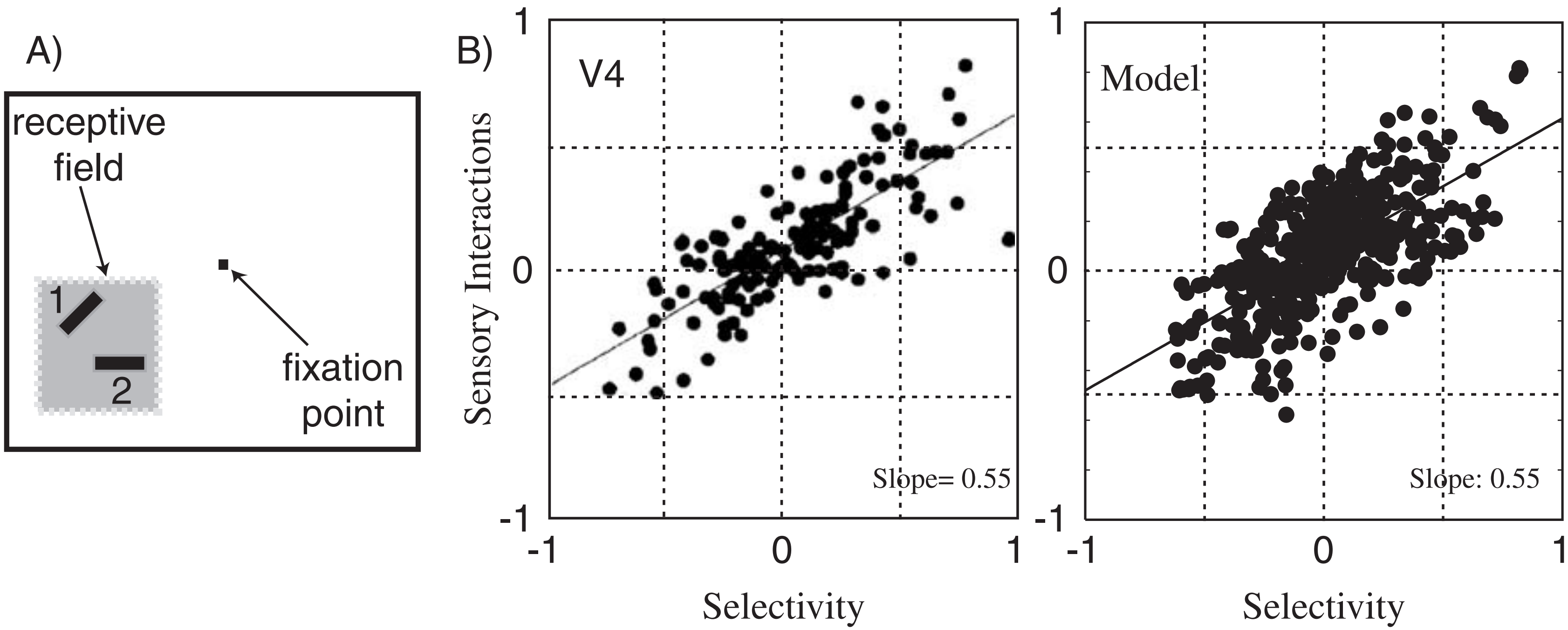
Supervised task-dependent learning

Unsupervised task-independent learning

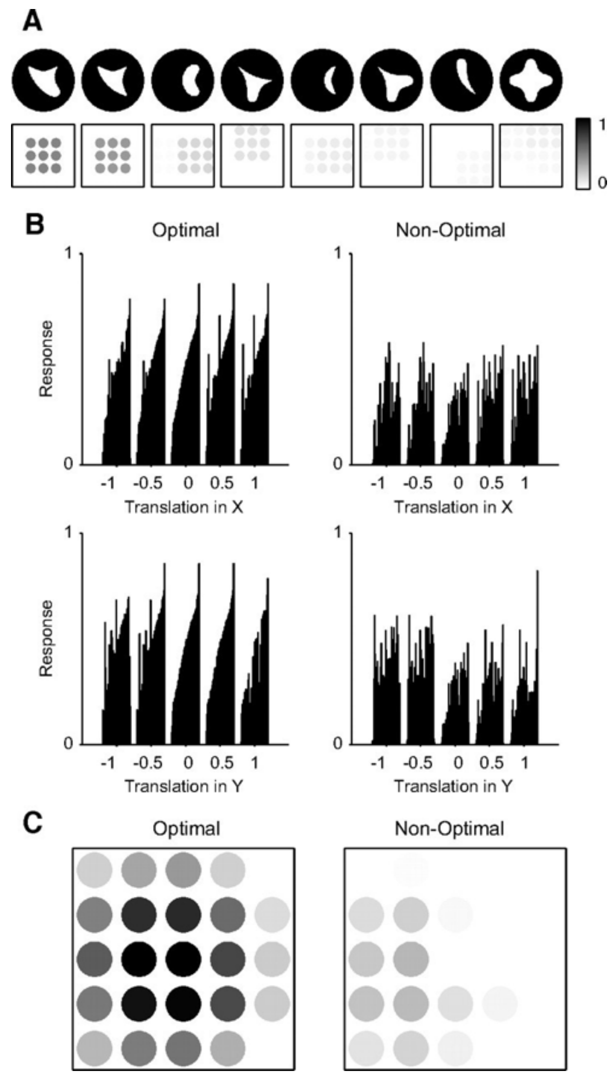
increase in complexity (number of subunits), RF size and invariance

Serre and Poggio 200/

The model captures the effects of clutter in visual responses

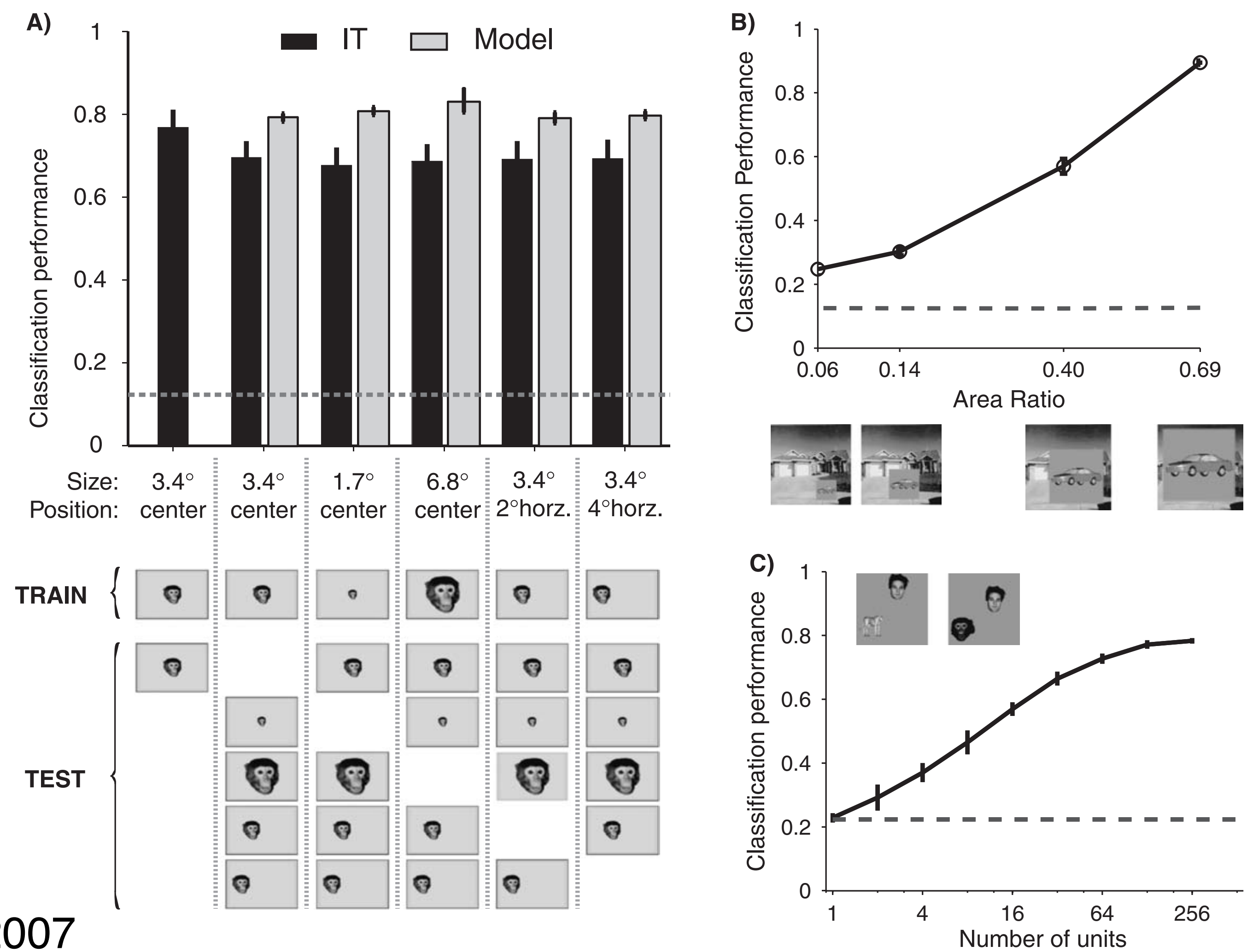


The model captures selectivity and invariance in V4 responses to curvatures



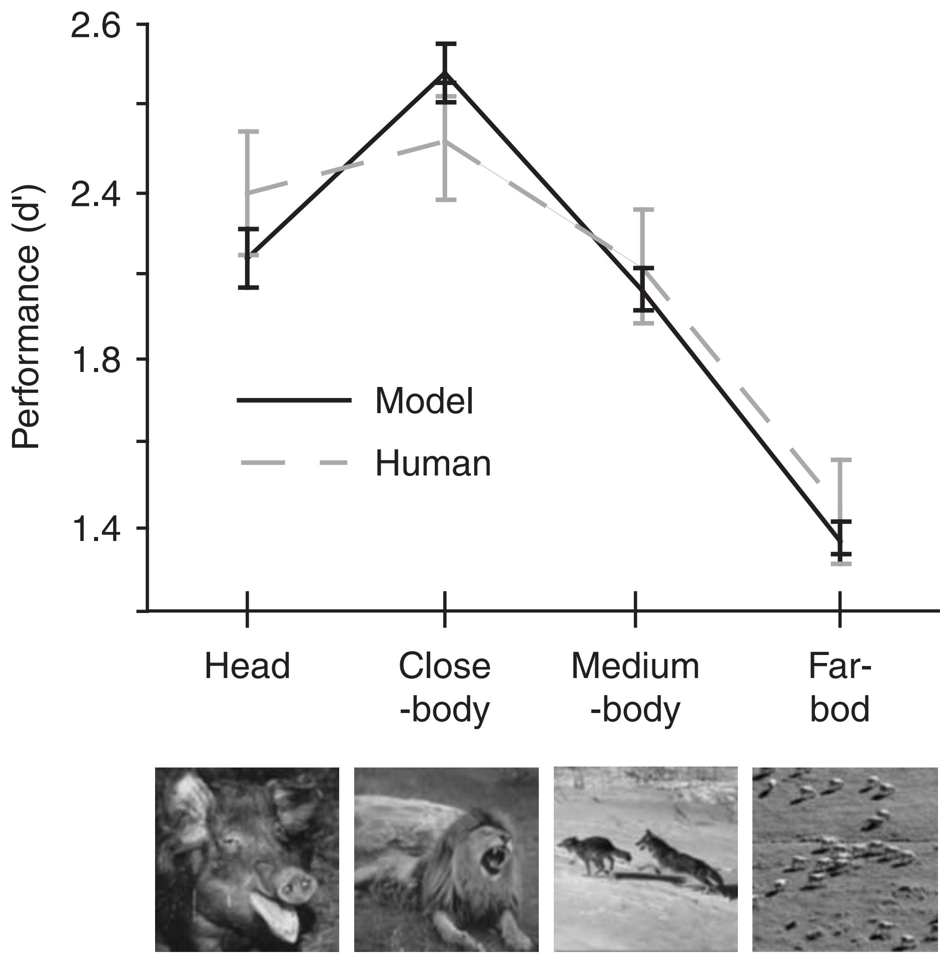
Cadieu et al 2007

The model approximates decoding of object information from IT cortex



Serre et al 2007

The model captures rapid recognition behavior



Serre et al 2007

Traditional approaches to visual recognition

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	38	43	105	255	253	253	253	253	253	174	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	43	139	224	226	252	253	252	252	252	252	252	252	158	14	0	0	0	0	0
0	0	0	0	0	0	0	0	0	178	252	252	252	252	253	252	252	252	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	109	252	252	230	132	133	132	132	189	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4	29	29	24	0	0	0	0	14	226	252	252	172	7	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	243	252	252	144	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88	189	252	252	252	14	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	212	247	252	252	252	204	9	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	32	125	193	193	193	253	252	252	252	238	102	28	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	222	252	252	252	252	253	252	252	252	177	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	223	253	253	253	253	255	253	253	253	253	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	31	123	52	44	44	44	44	143	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	5	75	9	0	0	0	0	0	0	98	242	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	61	183	252	29	0	0	0	0	18	92	239	252	252	243	65	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	147	134	134	134	134	203	253	252	252	188	83	0	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	252	252	252	252	252	252	253	230	153	8	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	49	157	252	252	252	252	252	217	207	146	45	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	7	103	235	252	172	103	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Image (or video)

→ Hand-crafted features → Learn readout (e.g. Support Vector Machine)

Example hand-crafted features

- Edges
- Textures
- Colors
- Corners
- Principal components
- Spatial frequency decomposition
- SIFT (Scale-invariant feature transform)

Deep learning

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	38	43	105	255	253	253	253	253	253	174	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	43	139	224	226	252	253	252	252	252	252	252	252	158	14	0	0	0	0	0
0	0	0	0	0	0	0	0	0	178	252	252	252	252	253	252	252	252	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	109	252	252	230	132	133	132	132	189	252	252	252	252	59	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4	29	29	24	0	0	0	0	14	226	252	252	172	7	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	243	252	252	144	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88	189	252	252	252	14	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	212	247	252	252	252	204	9	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	32	125	193	193	193	253	252	252	252	238	102	28	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	222	252	252	252	252	253	252	252	252	177	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	45	223	253	253	253	253	255	253	253	253	253	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	31	123	52	44	44	44	44	143	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	0	5	75	9	0	0	0	0	0	0	98	242	252	252	74	0	0	0	0	0	0	0	0
0	0	0	0	0	61	183	252	29	0	0	0	0	18	92	239	252	252	243	65	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	147	134	134	134	134	203	253	252	252	188	83	0	0	0	0	0	0	0	0	0
0	0	0	0	0	208	252	252	252	252	252	252	252	252	253	230	153	8	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	49	157	252	252	252	252	252	217	207	146	45	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	7	103	235	252	172	103	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

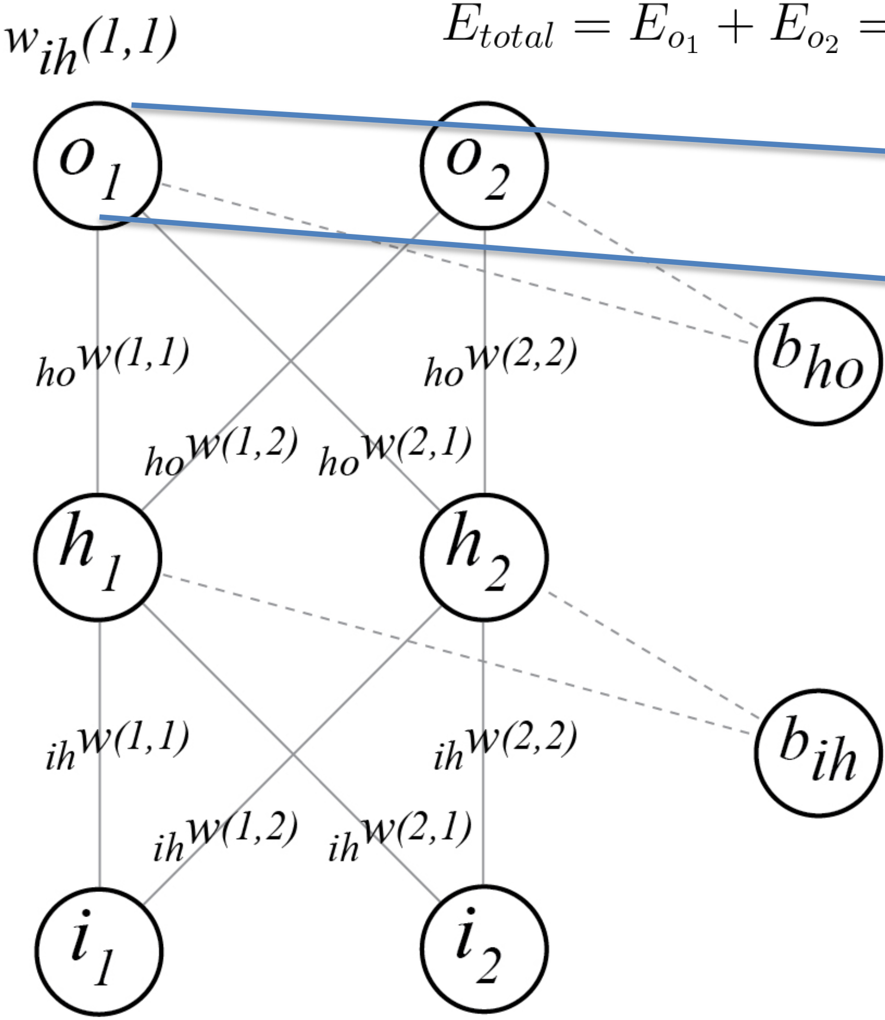
Image (or video)

→ Learn features and readout

The credit assignment problem

Back-propagation

$$E_{total} = E_{o_1} + E_{o_2} = 0.5[(target_{o_1} - o_1)^2 + (target_{o_2} - o_2)^2]$$



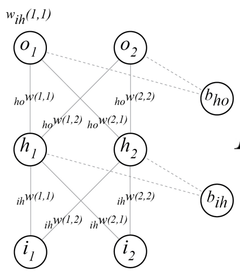
$$o_1 = \frac{1}{1 + e^{-net_{o1}}}$$

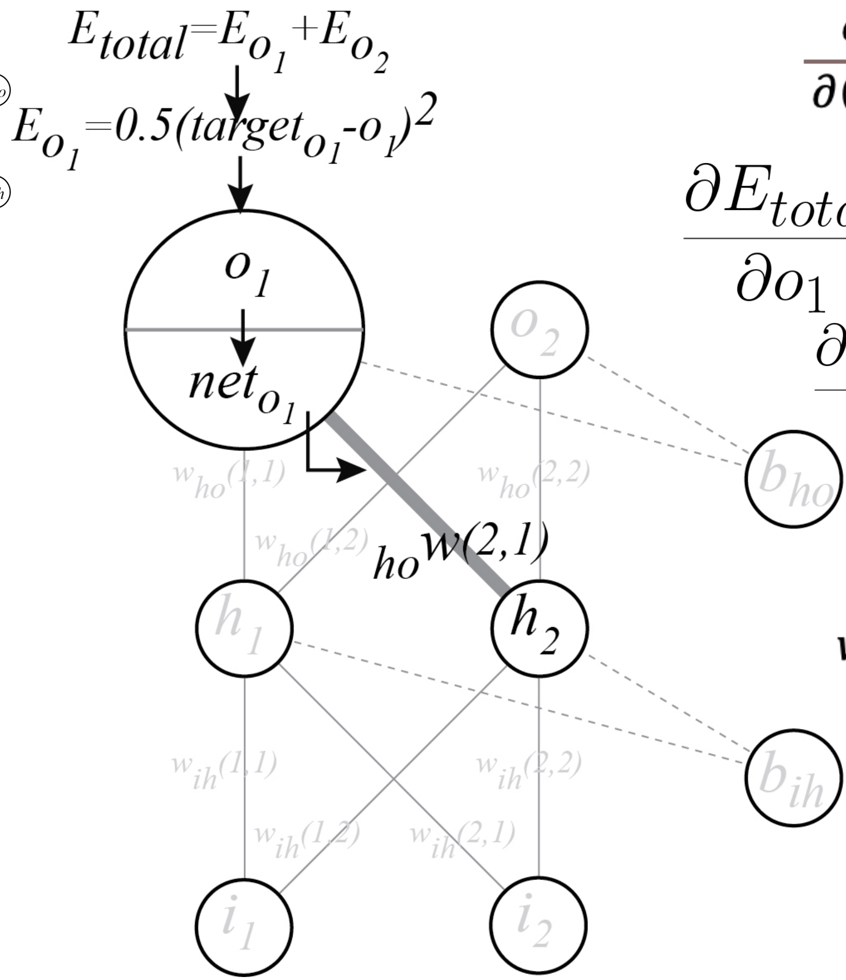
$$net_{o_1} = h_1 * w_{ho}(1,1) + h_2 * w_{ho}(2,1) + b_{ho}$$

$$h_1 = \frac{1}{1 + e^{-net_{h1}}}$$

$$net_{h_1} = i_1 * w_{ih}(1,1) + i_2 * w_{ih}(2,1) + b_{ih}$$

Back-propagation





$$\frac{\partial E_{total}}{\partial (w_{ho}2,1)} = \frac{\partial E_{total}}{\partial o_1}\frac{\partial o_1}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial (w_{ho}2,1)}$$

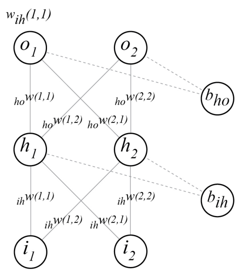
$$\frac{\partial E_{total}}{\partial o_1} = 2 * 0.5 * (o_1 - target_{o_1})$$

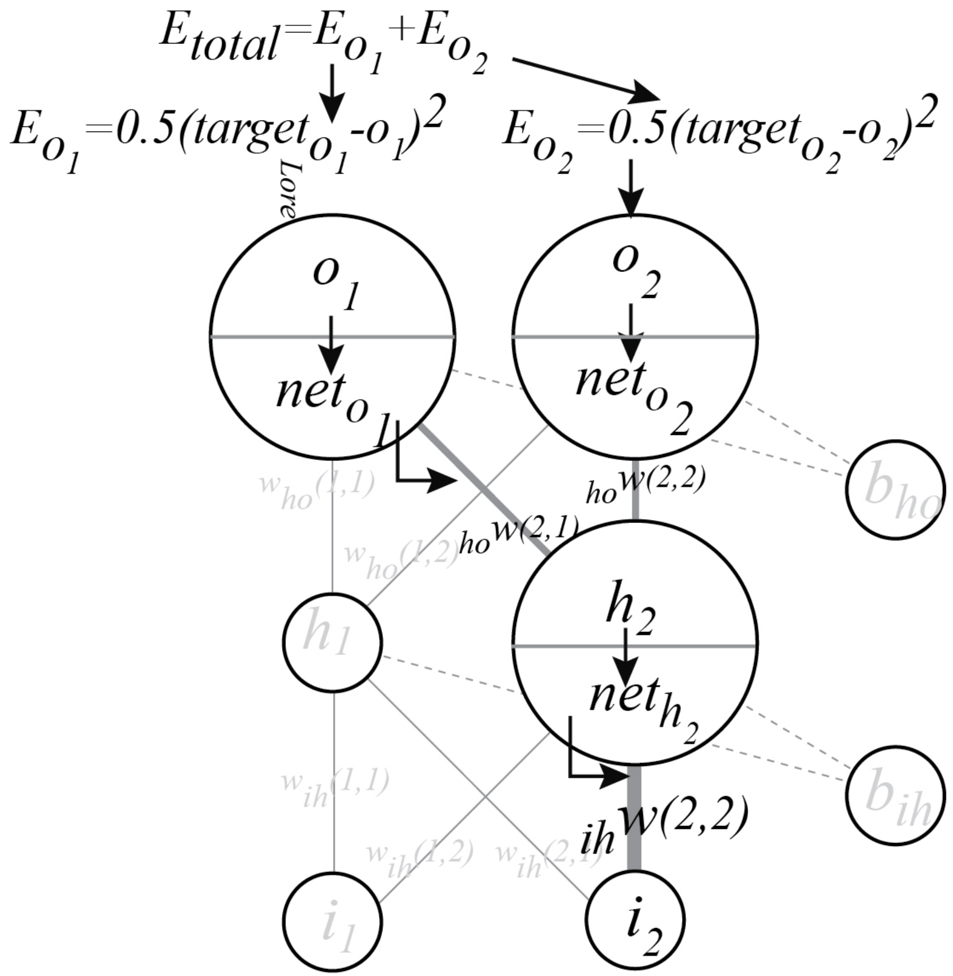
$$\frac{\partial E_{total}}{\partial o_1} = 2 * 0.5 * (o_1 - target_{o_1})$$

$$\frac{\partial net_{o1}}{\partial (w_{ho}2,1)} = h_2$$

$$w_{ho}(2,1) \rightarrow w_{ho}(2,1) - \varepsilon \frac{\partial E_{total}}{\partial w_{ho}(2,1)}$$

Back-propagation





Is back-propagation biologically plausible?

Symmetric feed-forward and feed-back weights

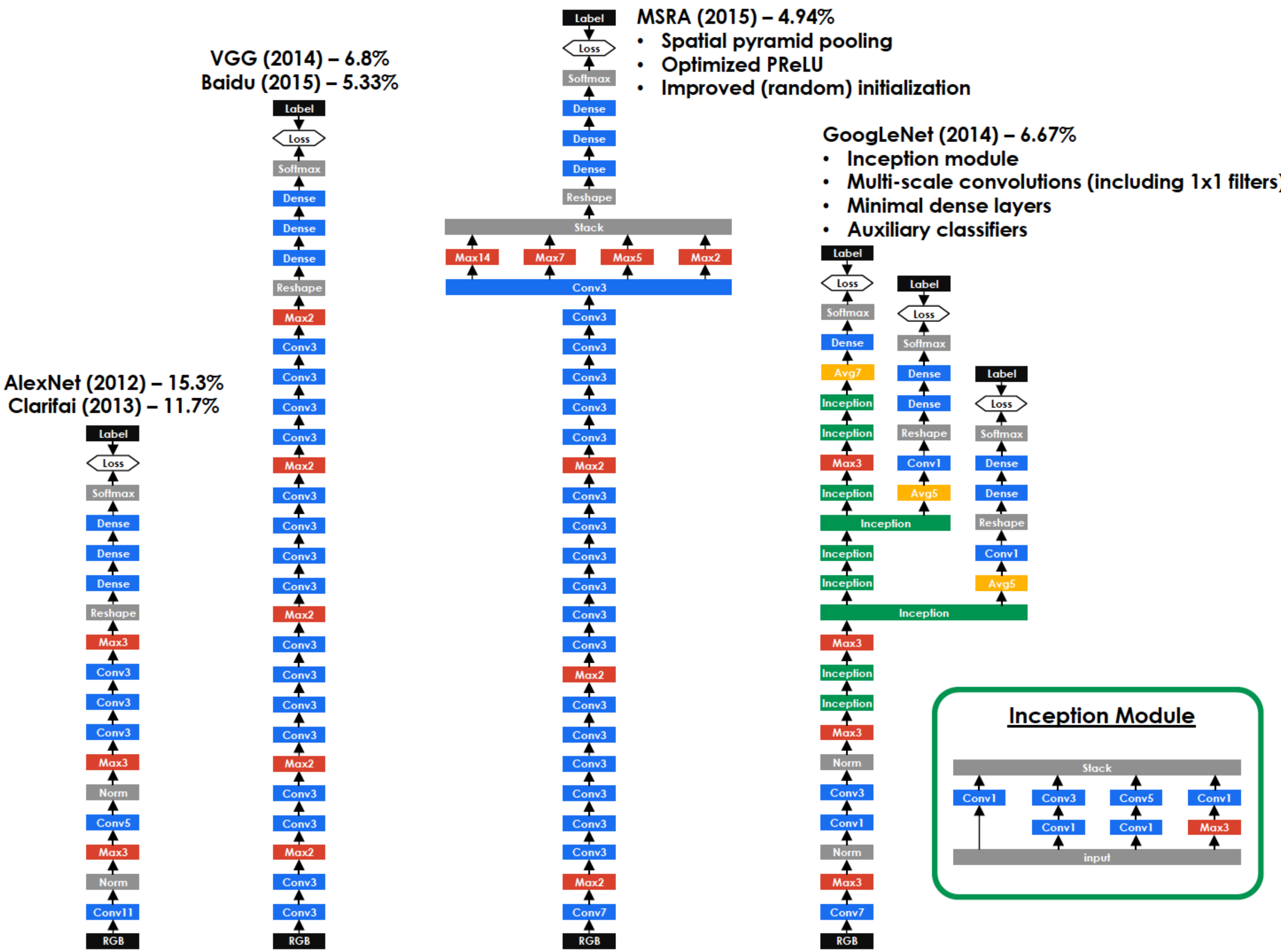
Signed error signals

Large gradients

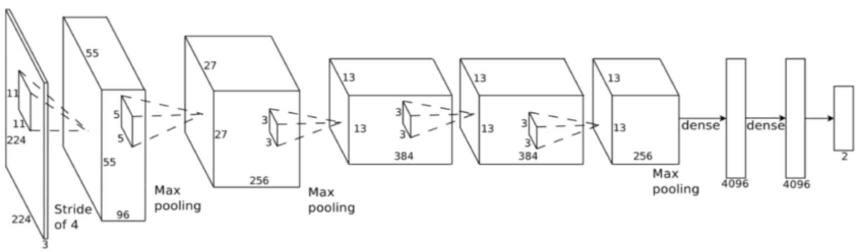
Feedback alters neuronal activity (and weights only indirectly)

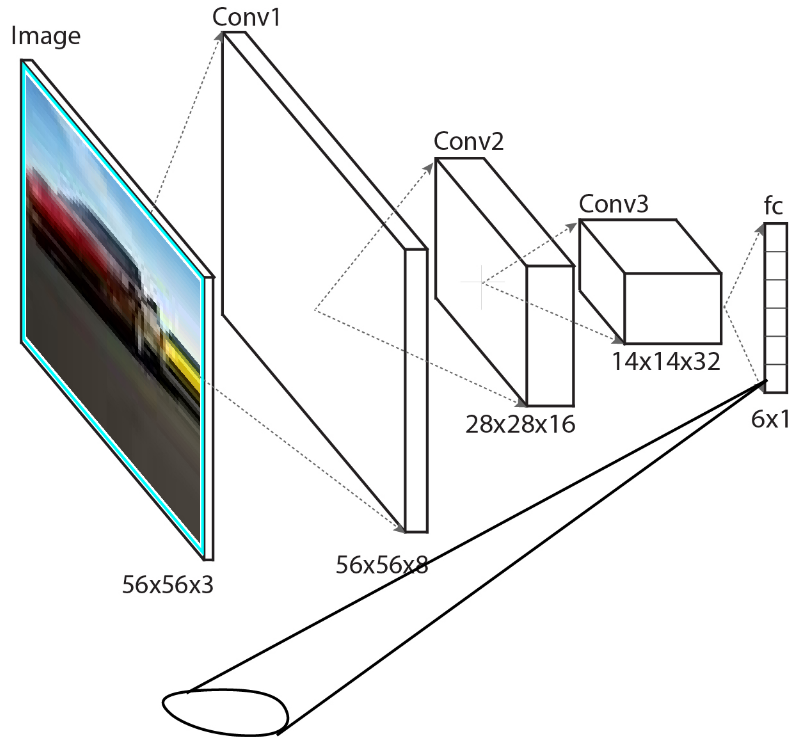
Supervised learning requires many training examples

Deeper and deeper

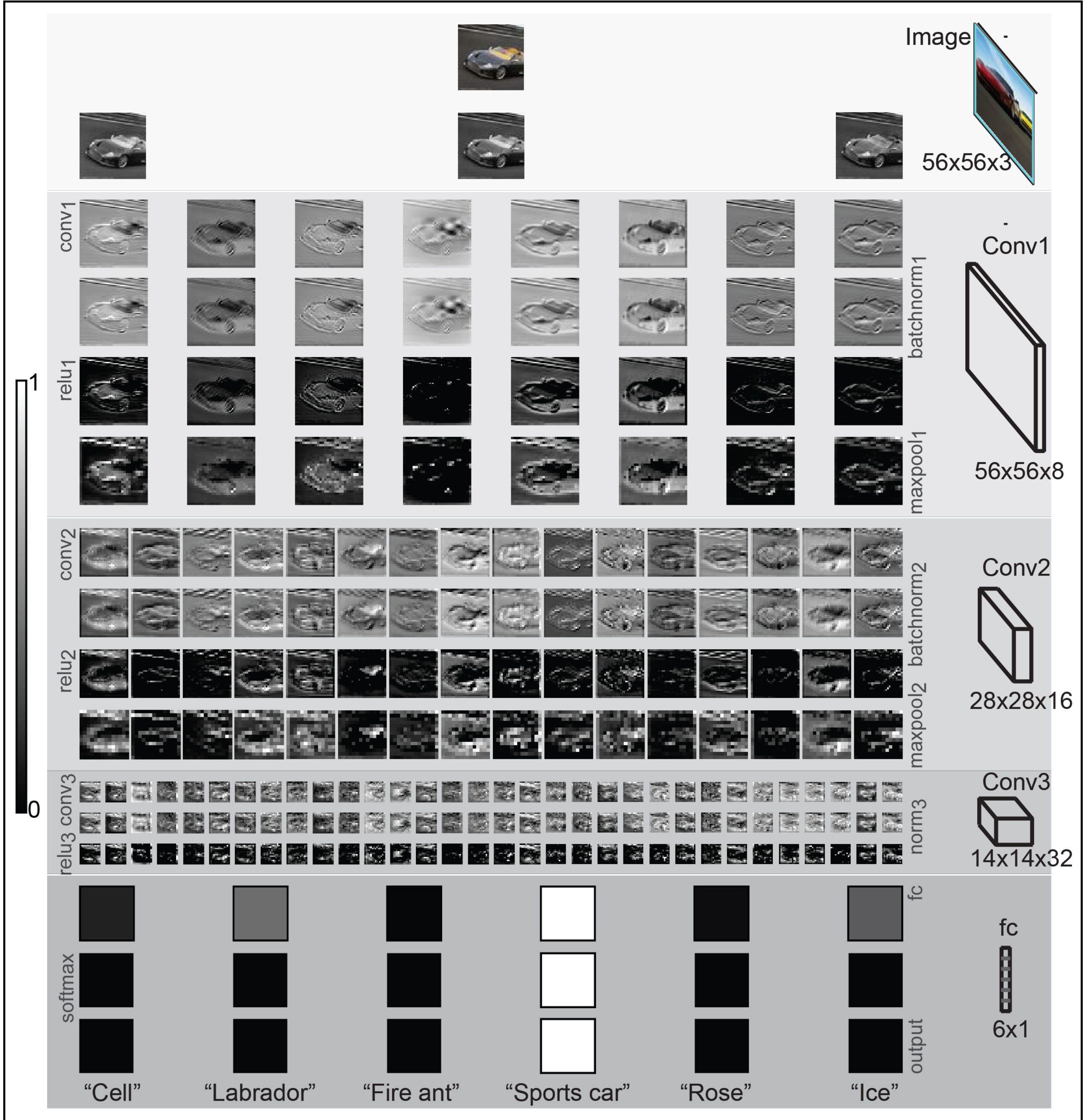


Putting it all together



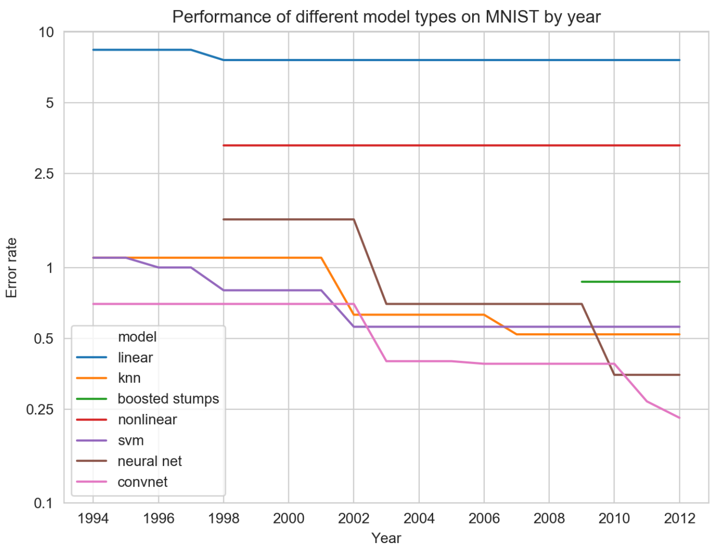


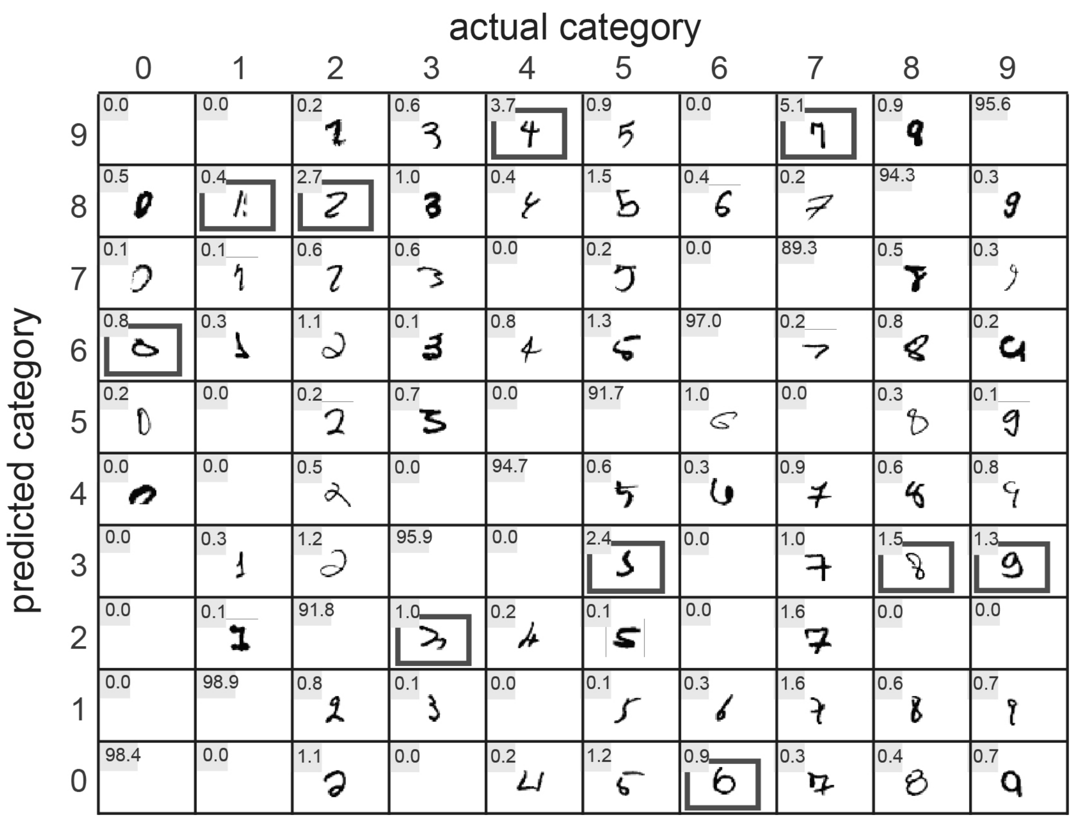
A CNN in action



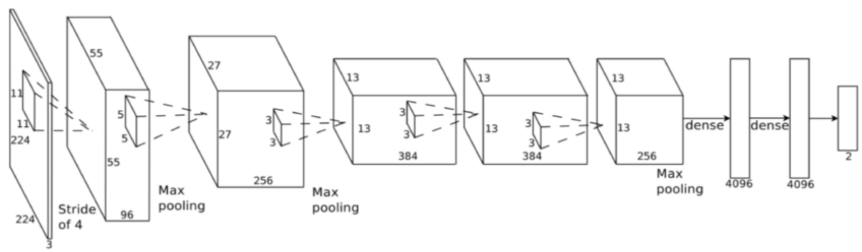
Kreiman, 2019

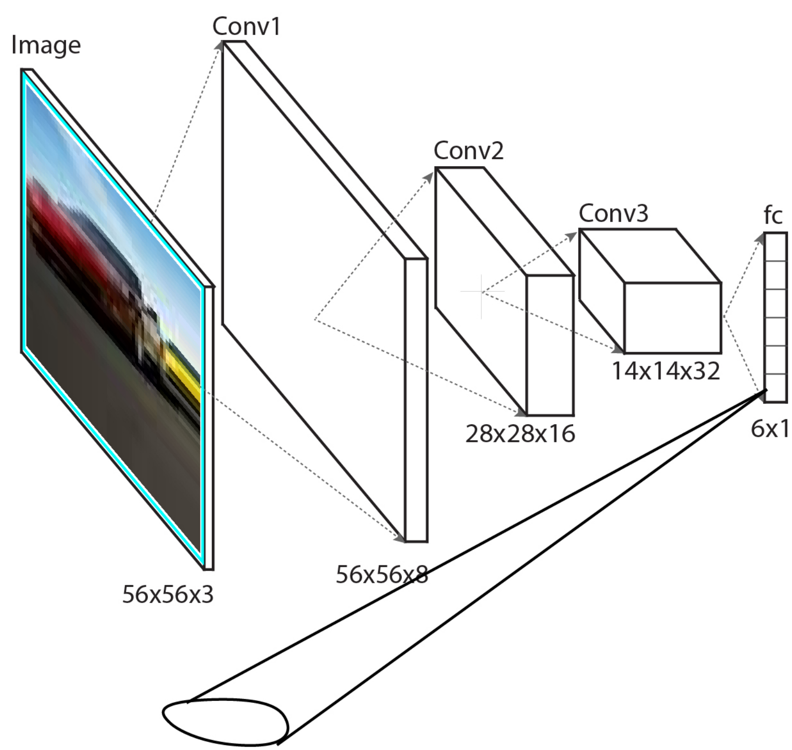
To err is human and algorithmic



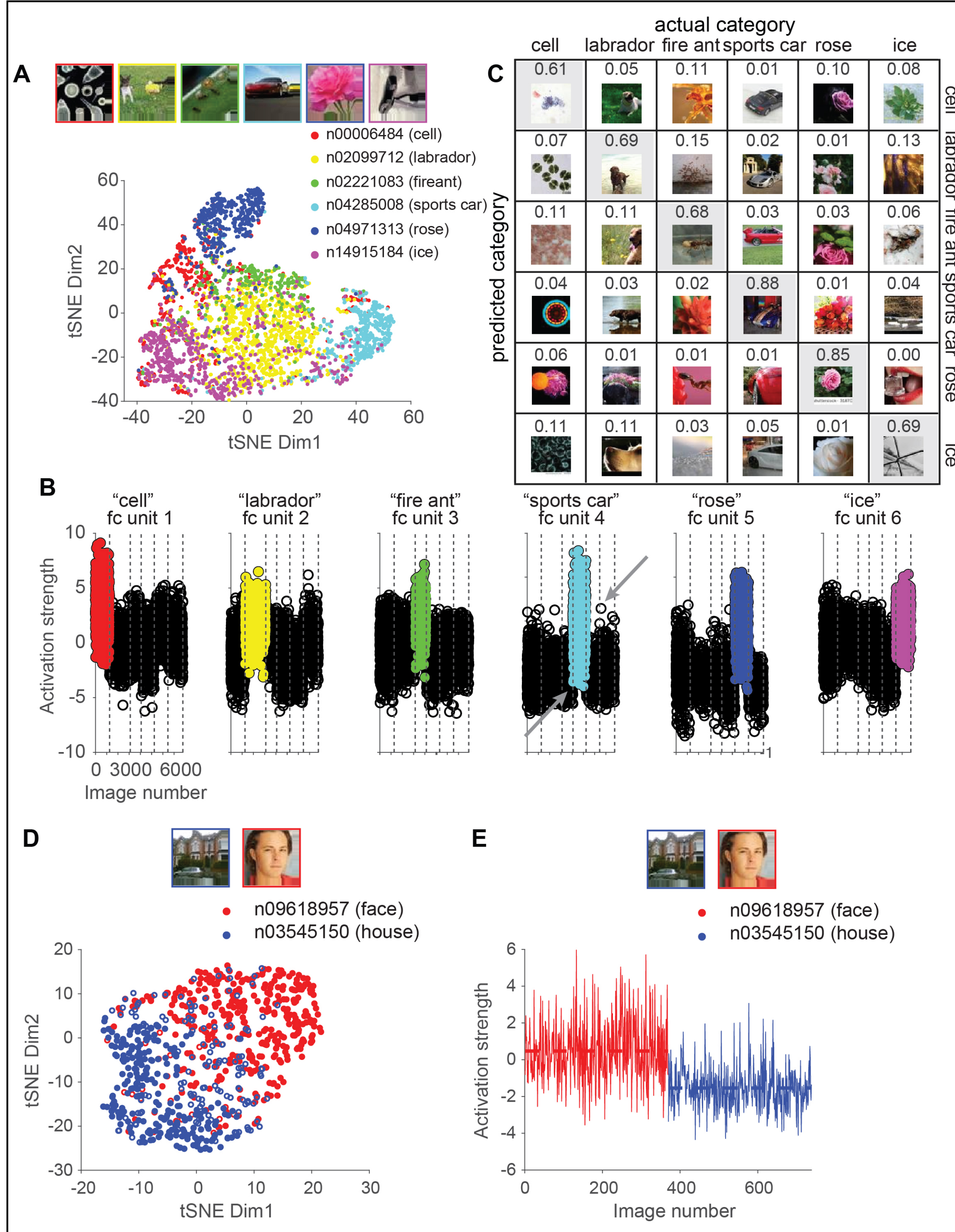


Putting it all together





CNNs in action: example



Kreiman, 2019

Computational models versus biology

	Biology	Computer vision
Hierarchy	✔	✔
Receptive field increase through hierarchy	✔	✔
Convolution-like operations	✔	✔
Backpropagation	?	✔
Supervised learning	~	✔
Unsupervised learning	✔	~
Interactions between areas	✔	~

Summary

- Visual recognition ~ extraction of task-dependent adequate features plus read-out
- Computation emerges from combination of simple elementary functions: convolution, normalization, rectification, pooling
- Hierarchical models capture essential neural and behavioral properties of visual processing
- Weights can be learned via back-propagation
- Current models provide only a coarse approximation to the complexities of the visual system

Visual Object Recognition

Computational Models and Neurophysiological Mechanisms

Neuro 130/230. Harvard College/GSAS 78454