I want machines to see

Jonas Kubilius

Brain & Cognition / KU Leuven (Belgium)

DataLux / 2015-04-08

klab.lt

The goal

How to get from an image to the knowledge about its contents as perceived by a human observer.

Compete against GoogLeNet

ILSVRC 2014 demo by Andrej Karpathy

Compete against GoogLeNet

ILSVRC 2014 demo by Andrej Karpathy

How did we get here?

Why is vision difficult?

1. The invariance problem
2. Discovering structure

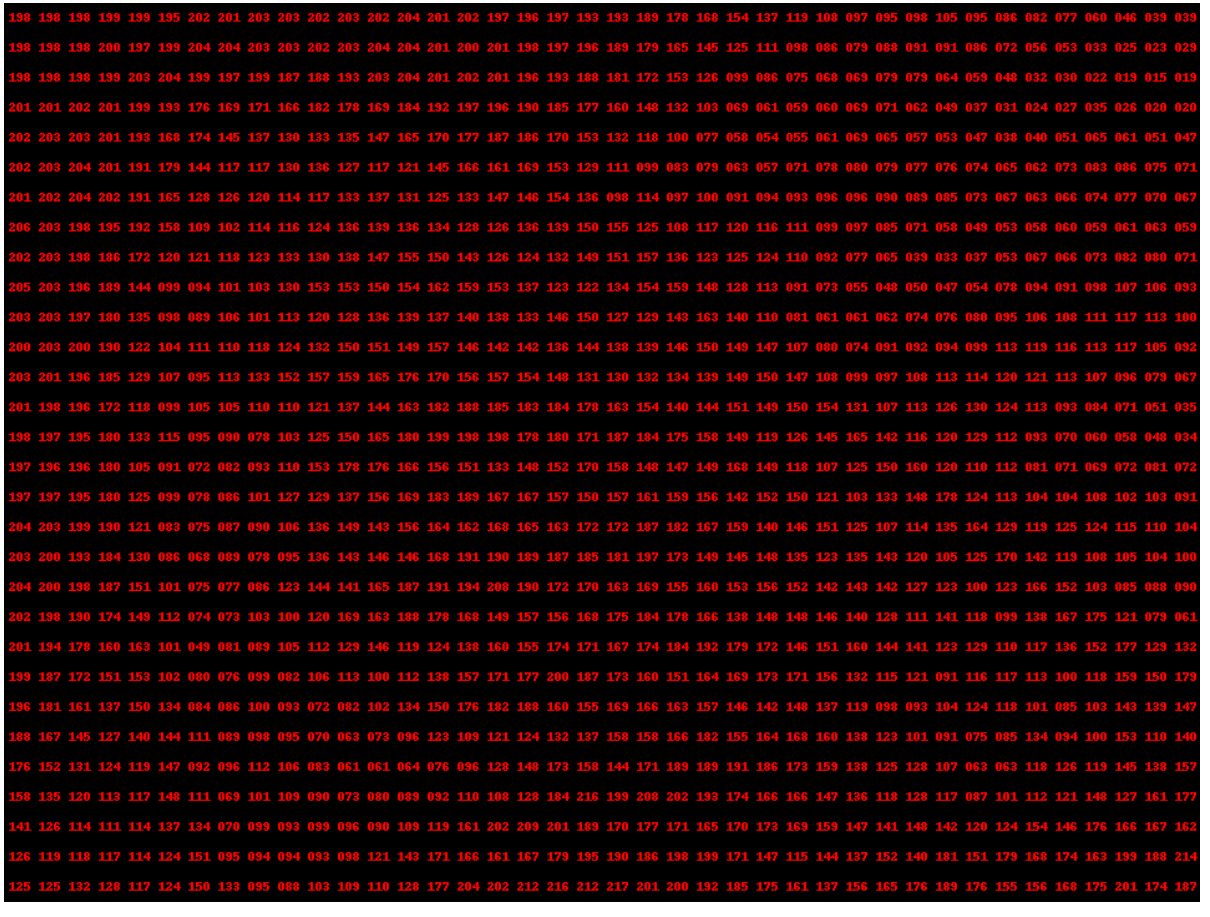
The invariance problem

cc by 2.0 – Wendy Cope / Flickr

Discovering structure

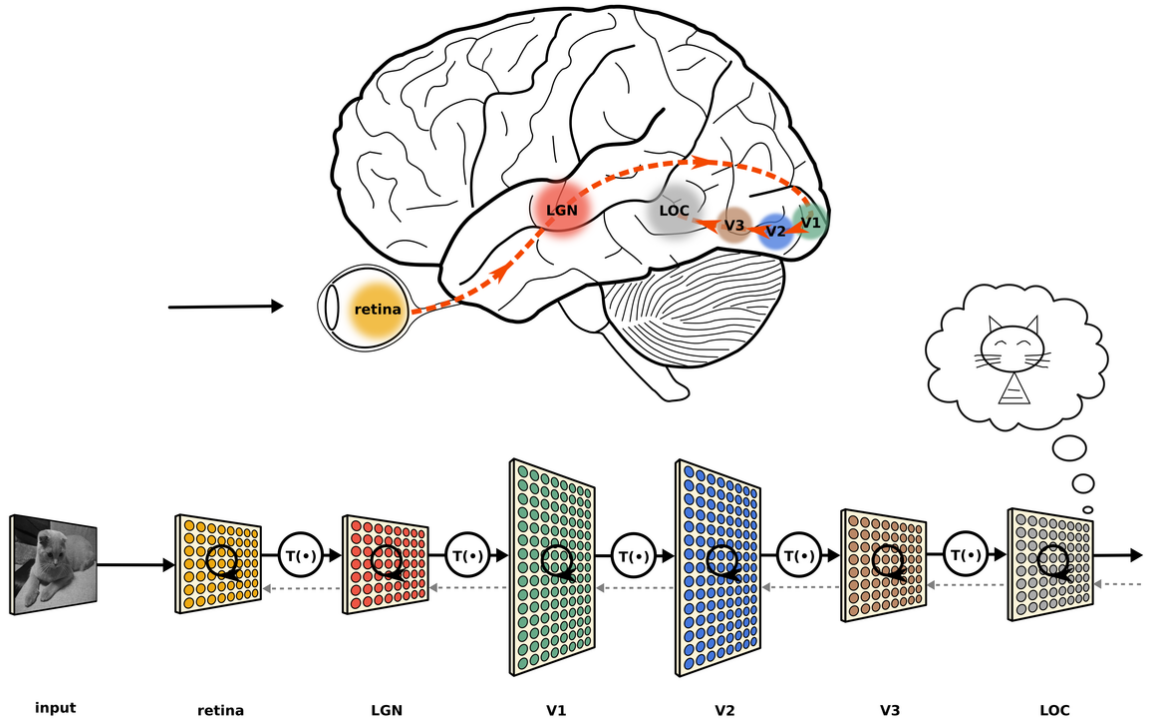
Discovering structure

Discovering structure

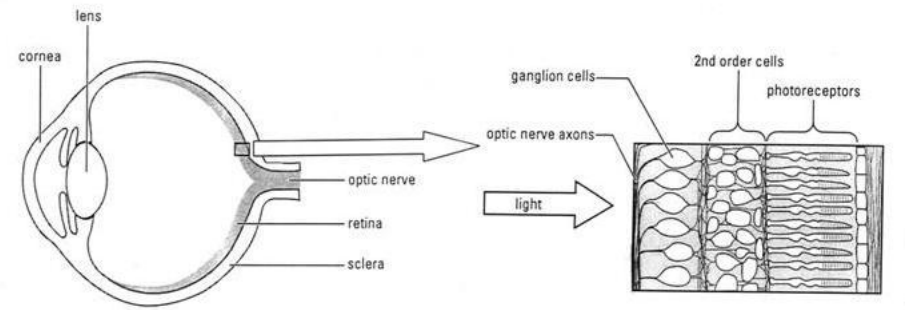


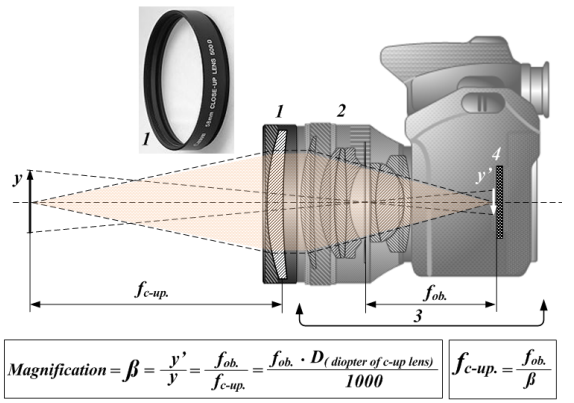
Human visual system

Human brain

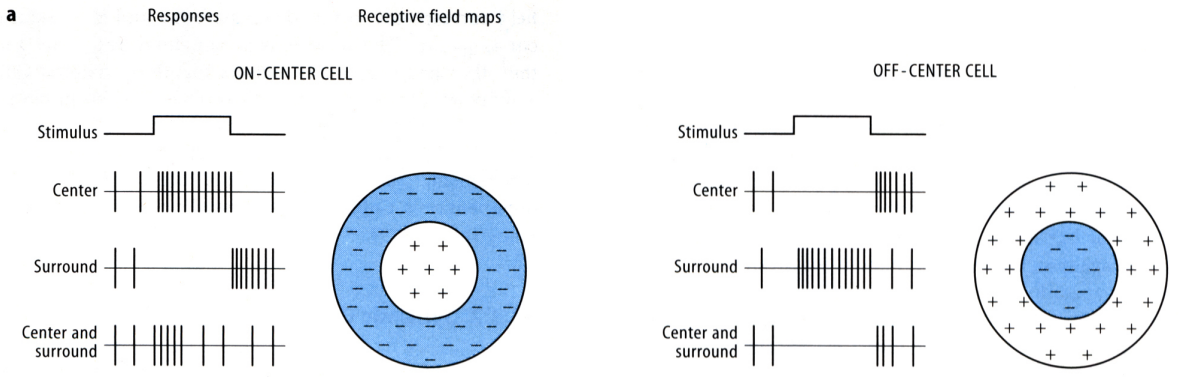


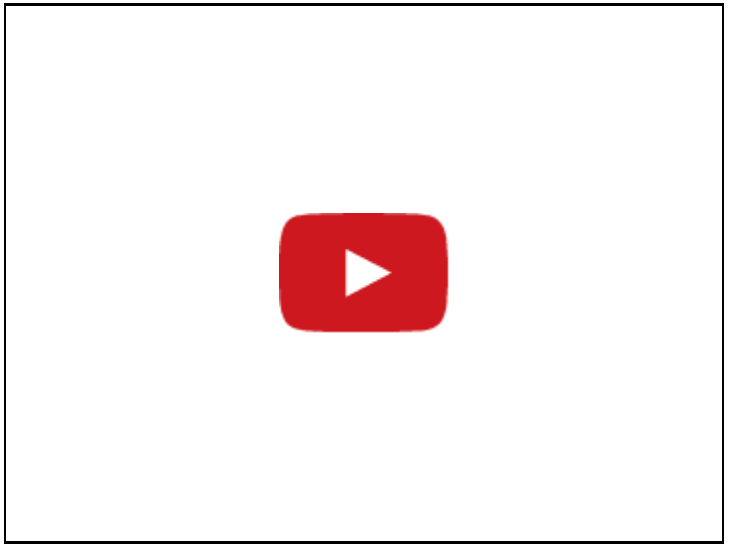
Retina





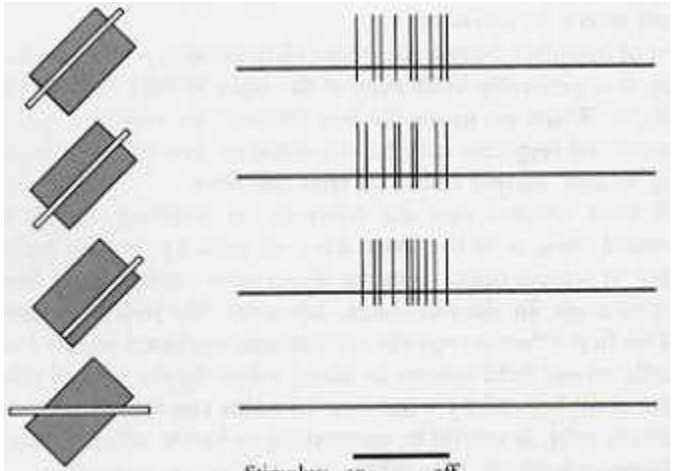
Retinal ganglion cells





V1: Simple cell

V1: Complex cell

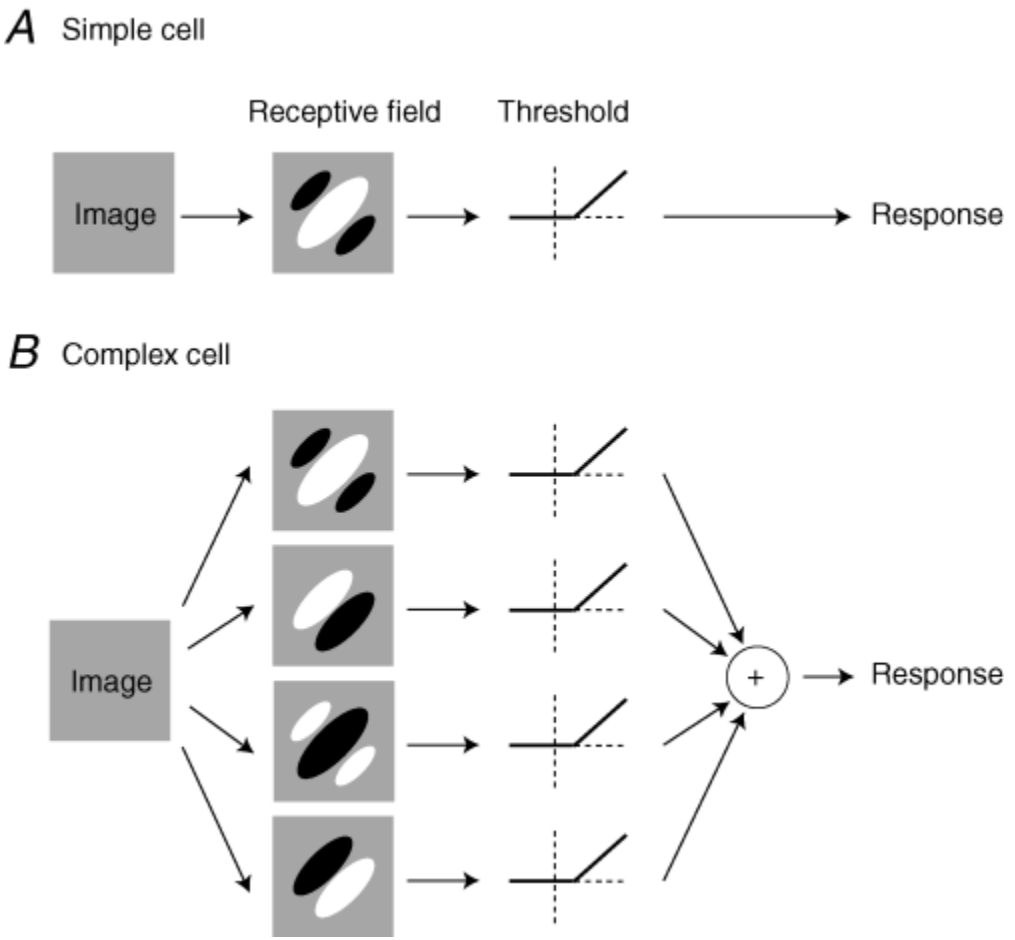


V1 function

Simple cells: **feature selectivity**

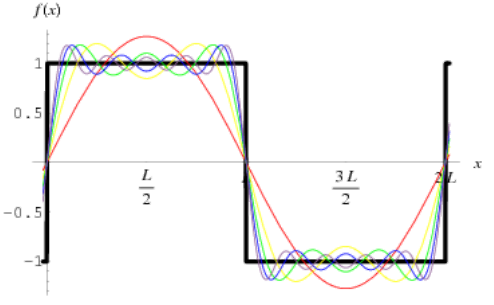
Complex cells: **feature invariance**

Complex cell connectivity

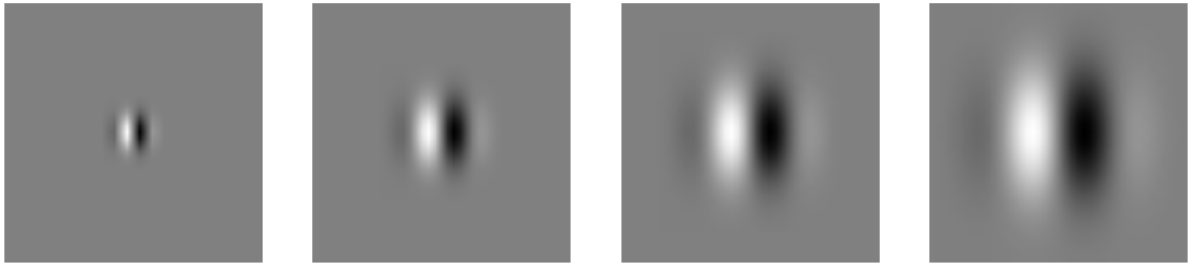


fair use – Carandini (The Journal of Physiology, 2006)

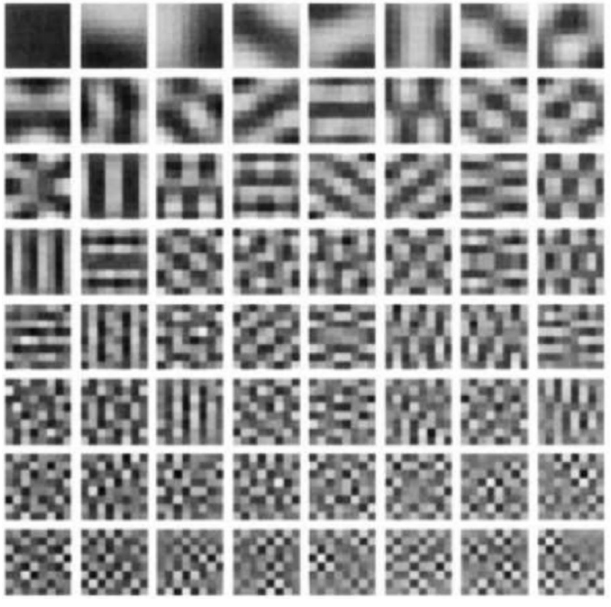
V1: Local frequency filtering



Weisstein / MathWorld



Principal components analysis



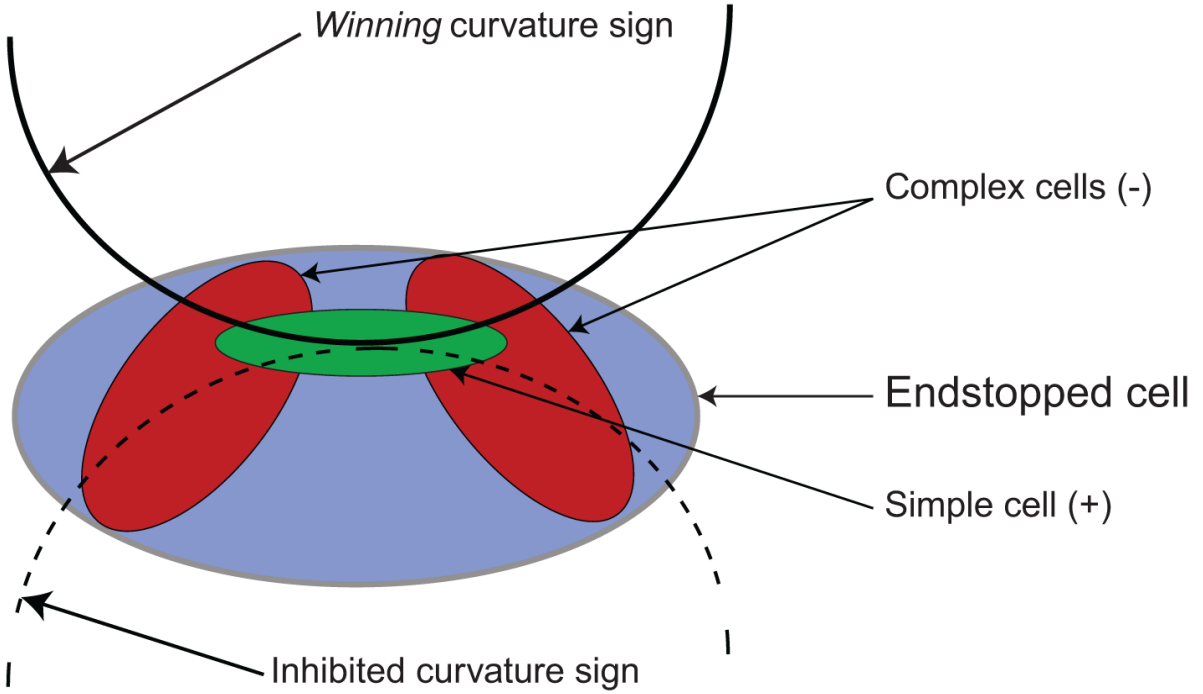
fair use – Olshausen & Field (Nature, 1996)

Sparse coding

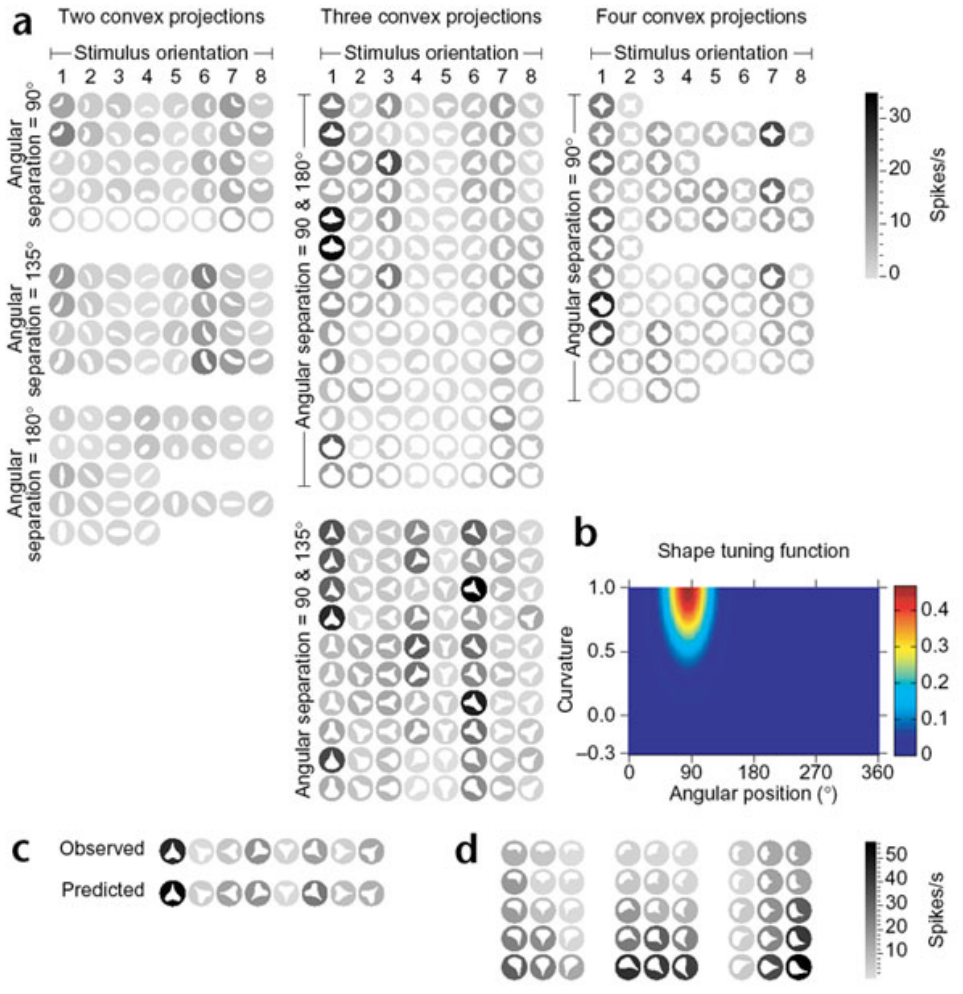
Principal Components Analysis (PCA) + sparcity, or Independent Components Analysis (ICA)

V2: End-stopping

V2: Curvature detection

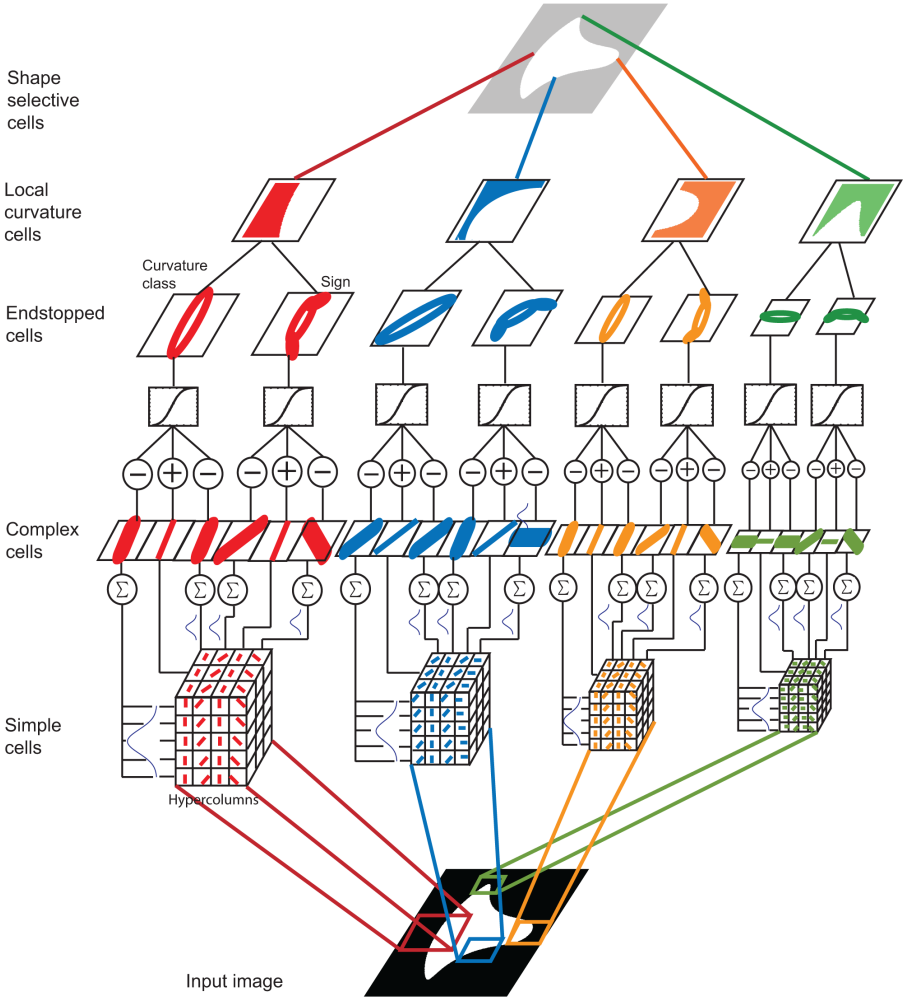


V4: Curvature detection

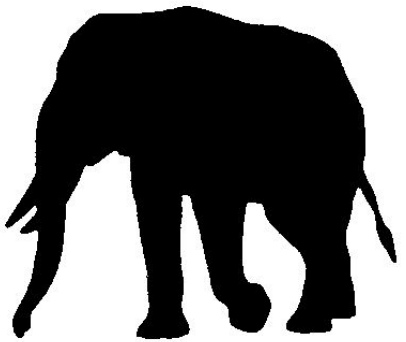


fair use – Pasupathy & Connor (Nature Neuroscience, 2002)

Putting it all together



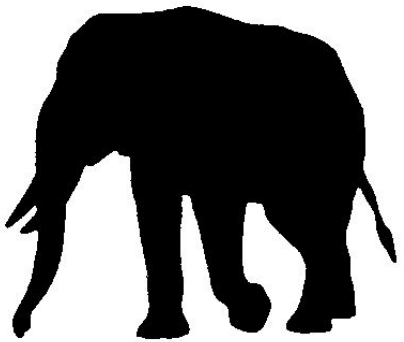
Recognition from shape only?



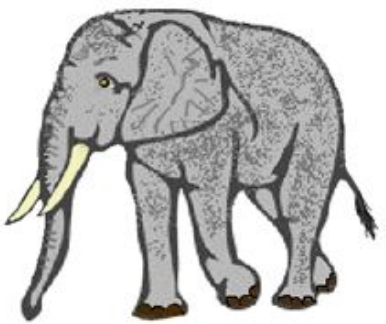
100%

34%

Recognition from shape only?



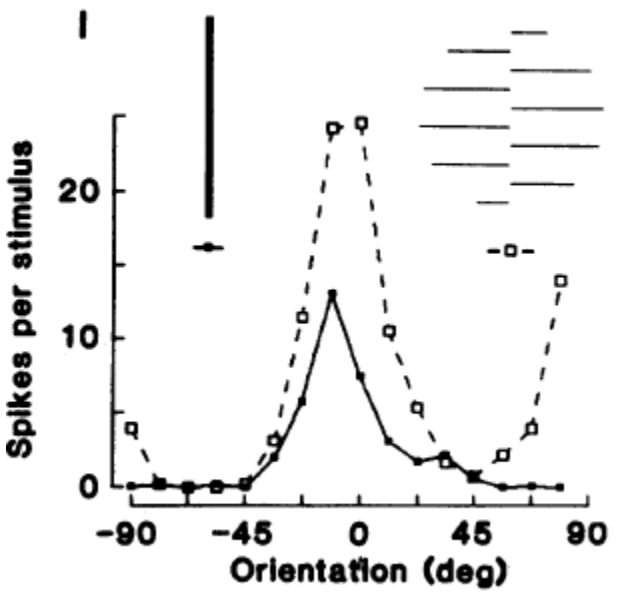
100%



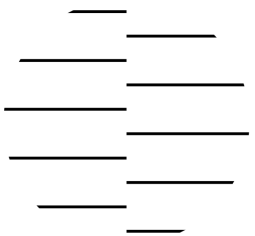
fair use – Wagemans et al. (Perception, 2008)

You won't get far with edges

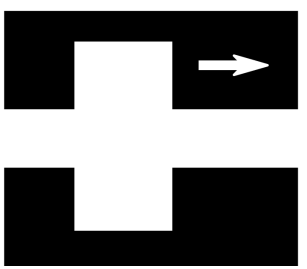
What is V2 doing?



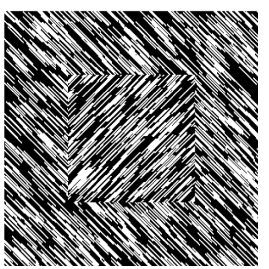
A



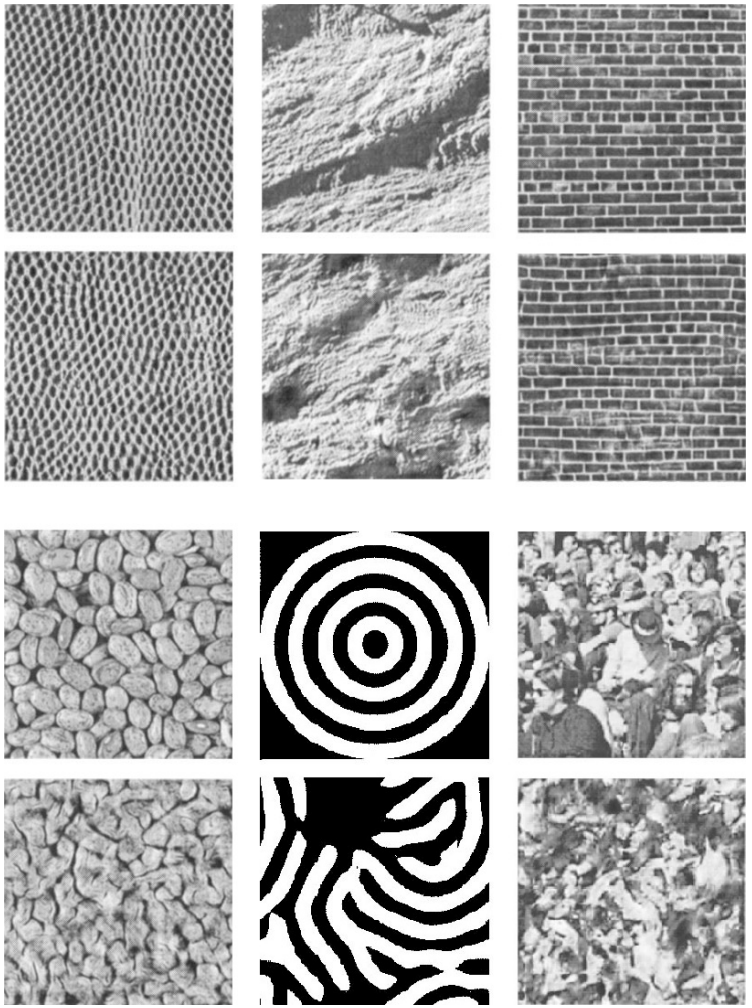
B



C

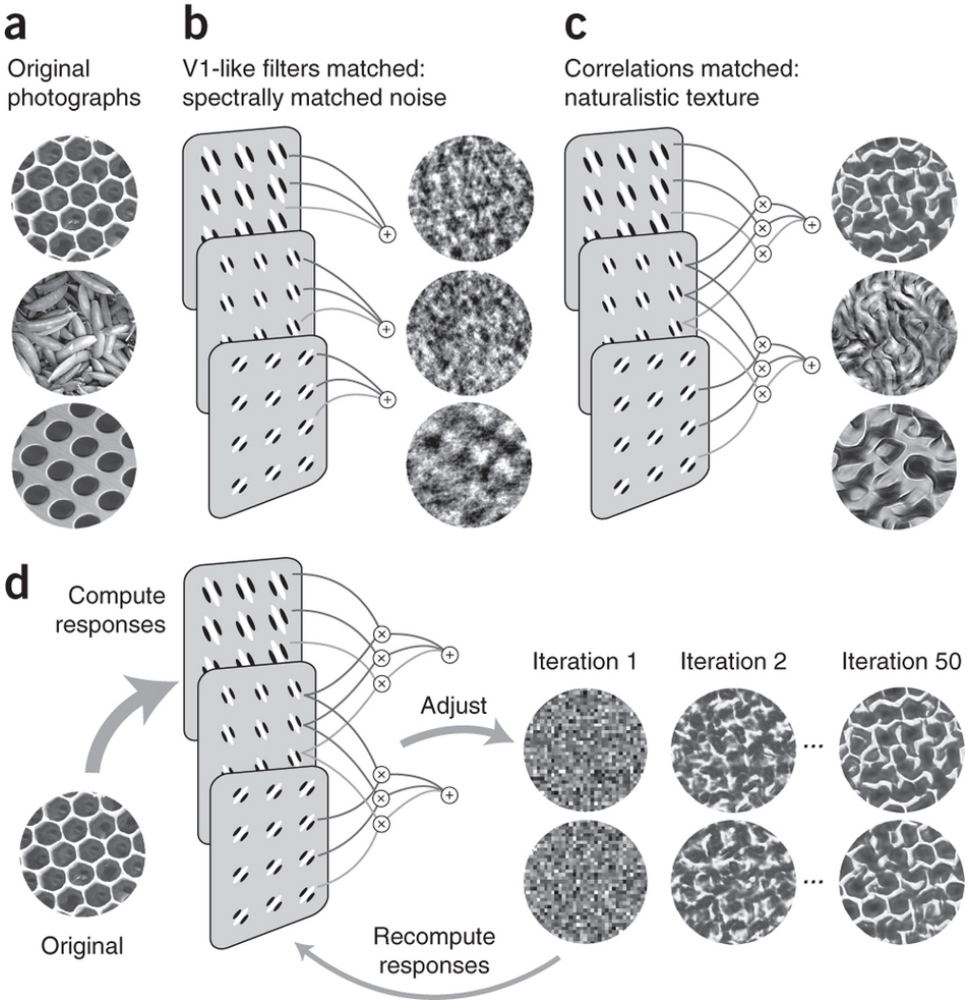


Texture processing

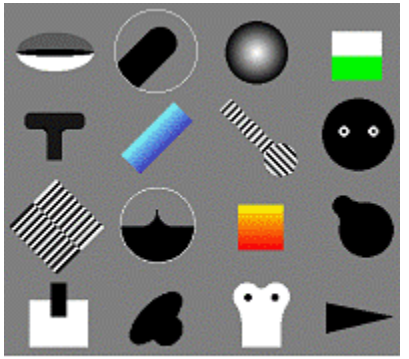


fair use – Portilla & Simoncelli (IJCV, 2000)

V2 processes textures



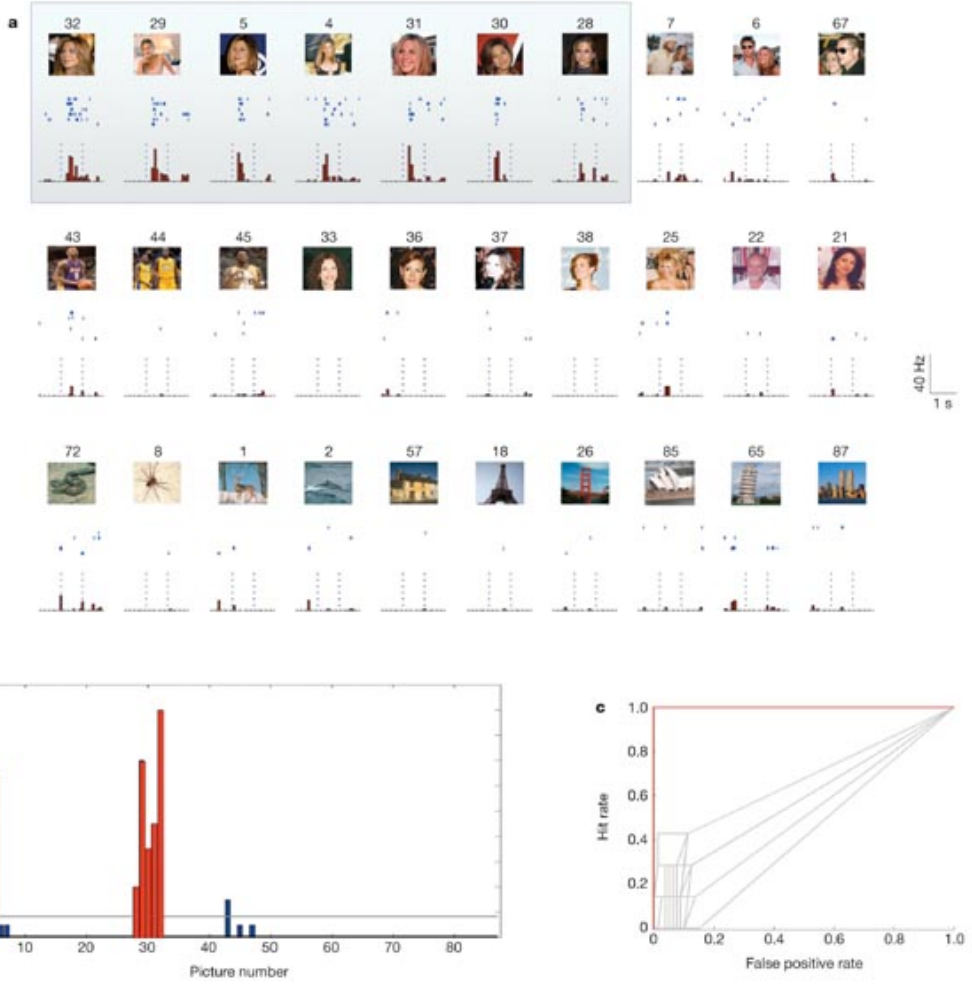
Later stages



Examples of features to which TE cells selectively respond

fair use – Tanaka (Annual Review of Neuroscience, 1996) via RIKEN

Jeniffer Aniston cell

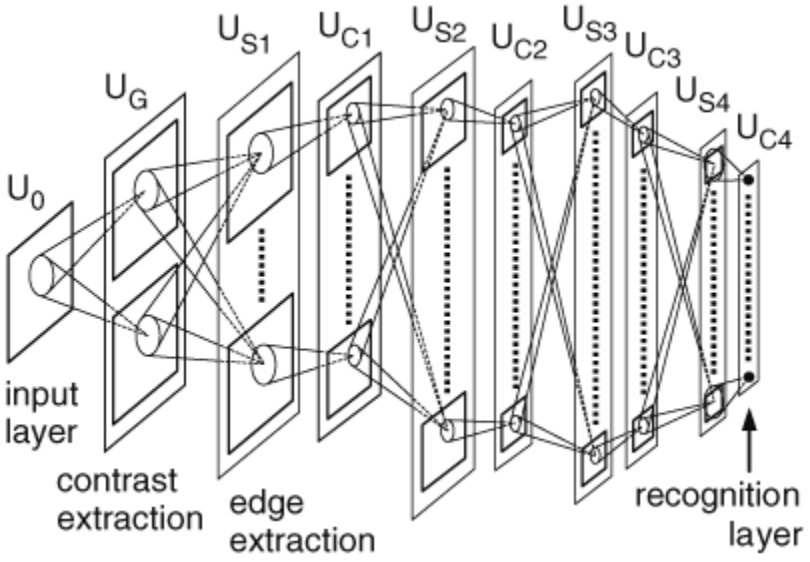


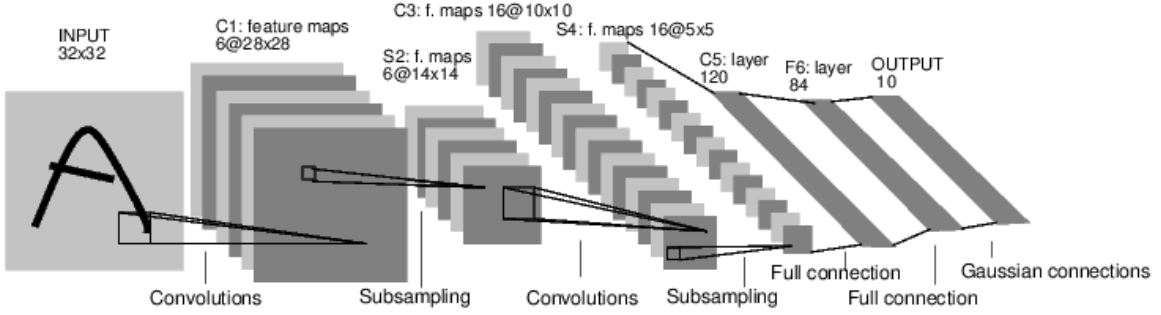
fair use – Quiroga et al. (Nature, 2005)

Deep Neural Networks

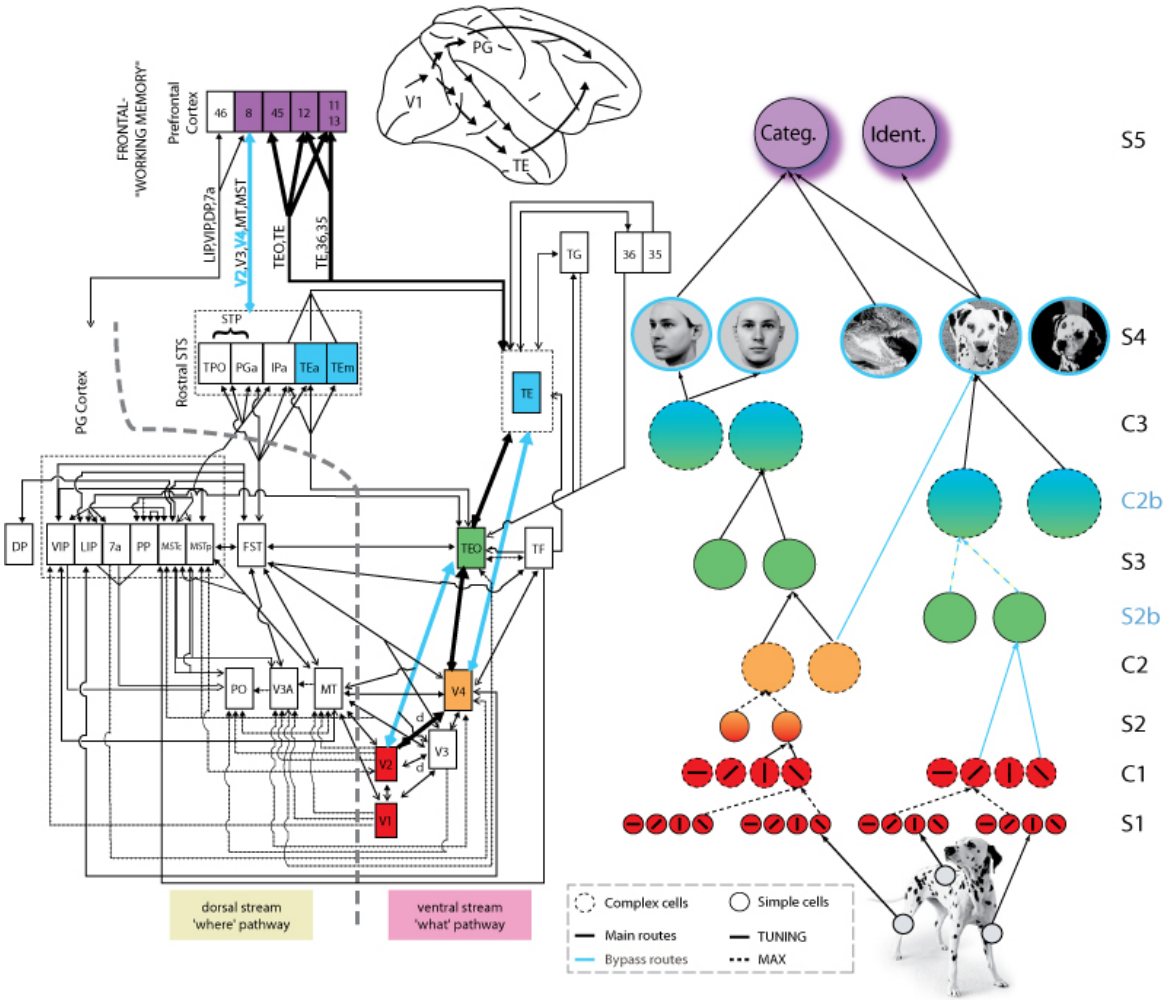
(actually, only convolutional)

Early architectures

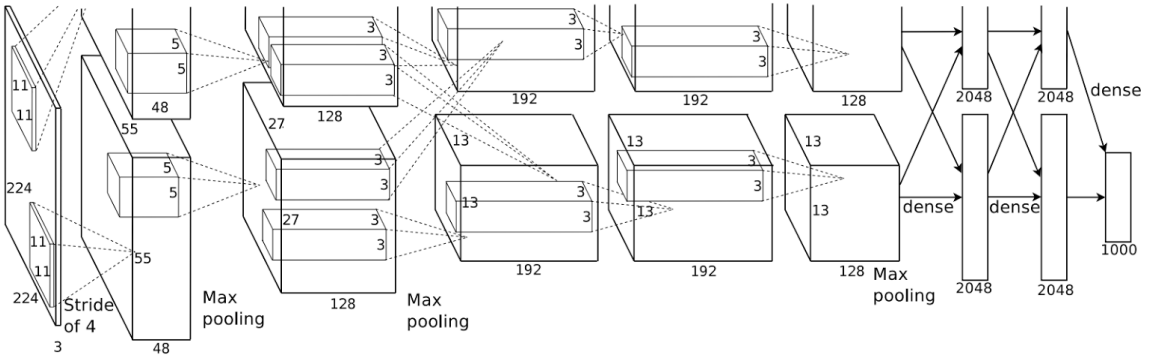




HMAX

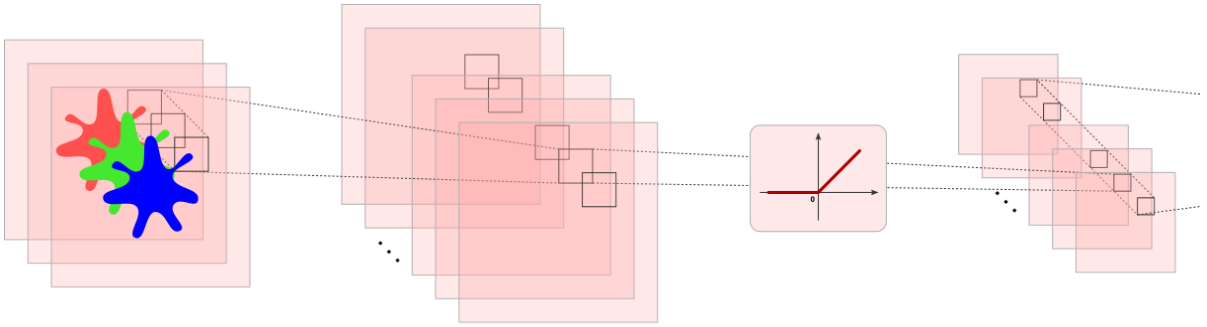


AlexNet



fair use – Krizhevsky et al. (NIPS, 2012)

Basic building blocks



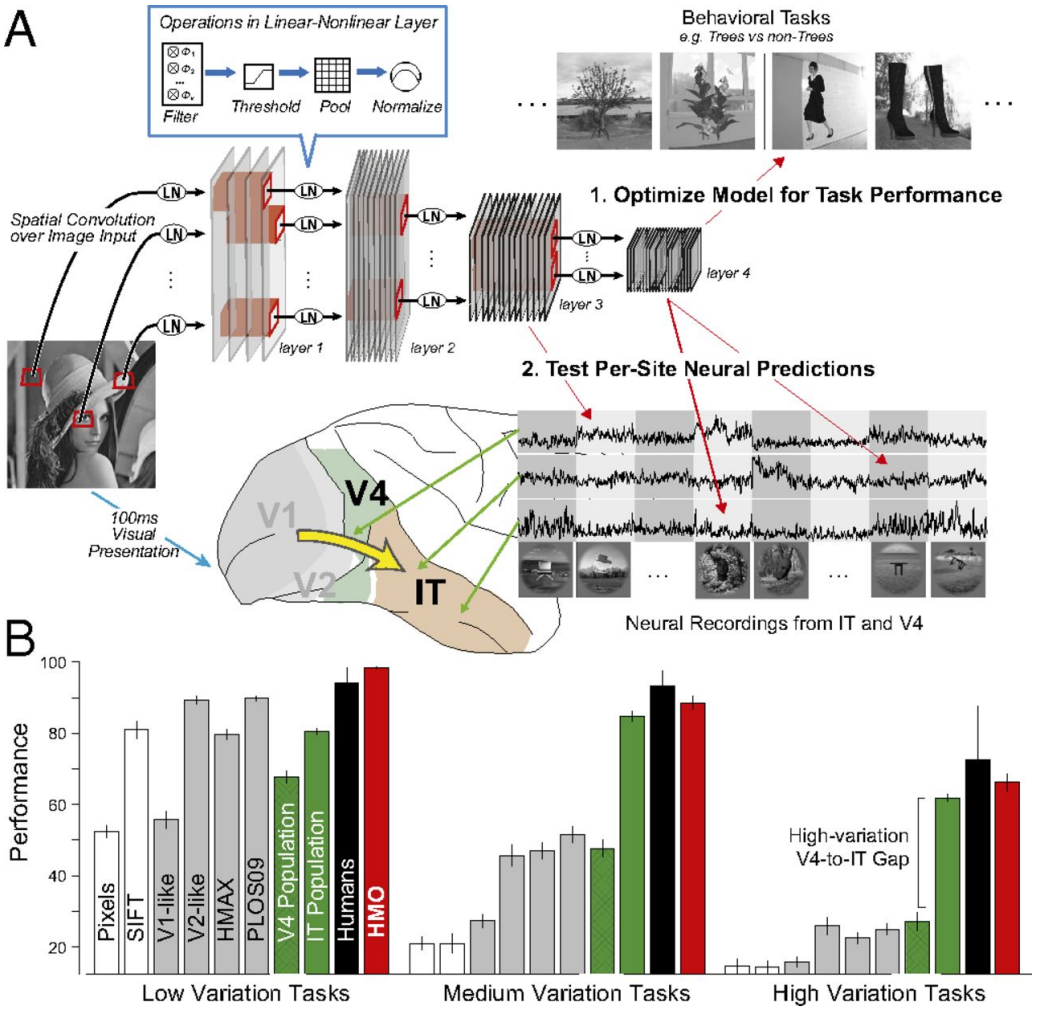
RGB input
224×224×3

convolution
224×224×64

ReLU
$f(x) = \max(0,x)$

pooling
112×112×64

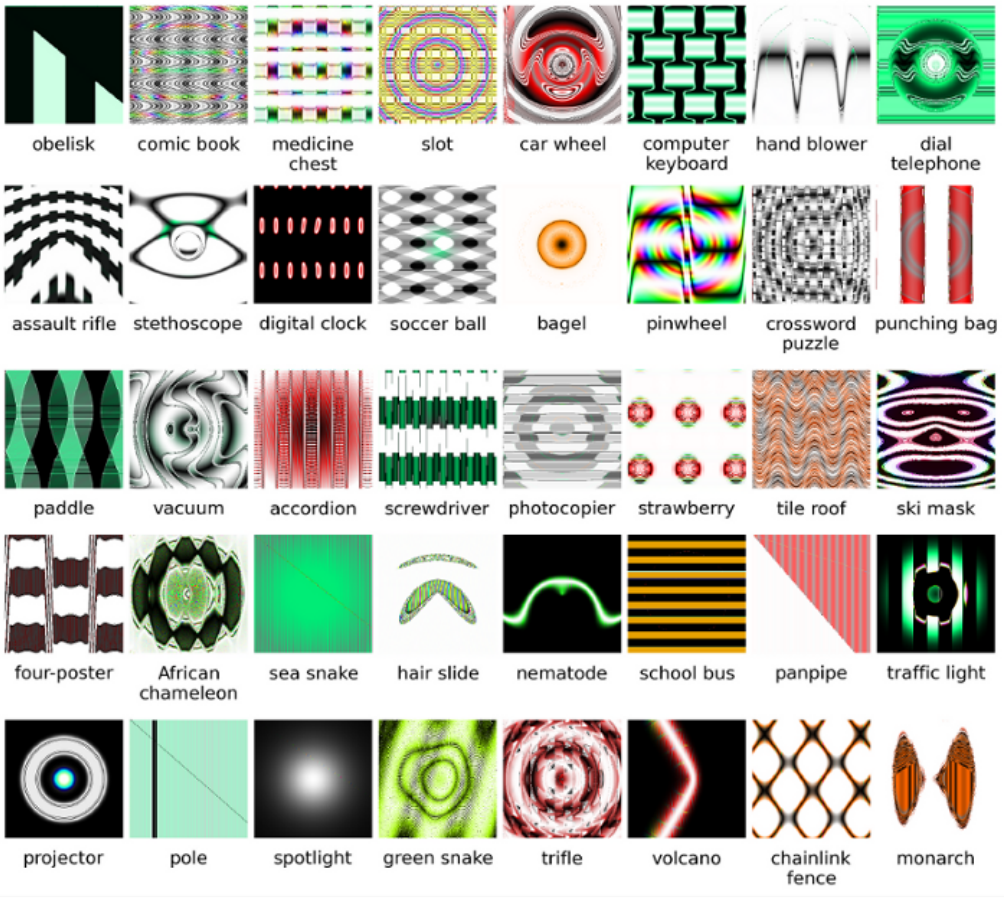
HMO



fair use – Yamins et al. (PNAS, 2014)

Fooling ConvNets

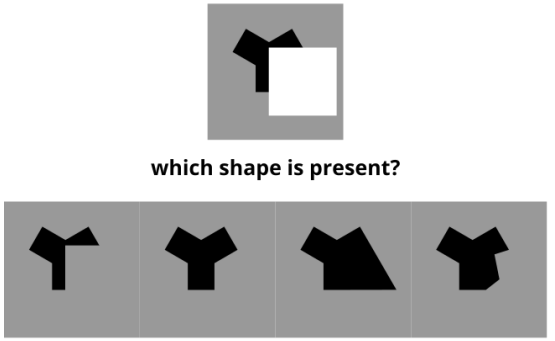
With 99.12% confidence, this is what I see...



fair use – Nguyen et al. (CVPR, 2015)

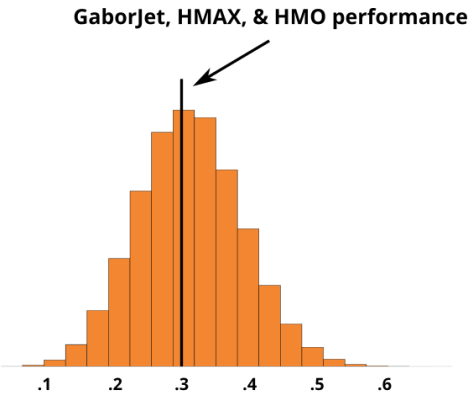
Is this what we fought for?

Occlusion



which shape is present?

a dataset of partially occluded and unoccluded shapes



random model accuracy

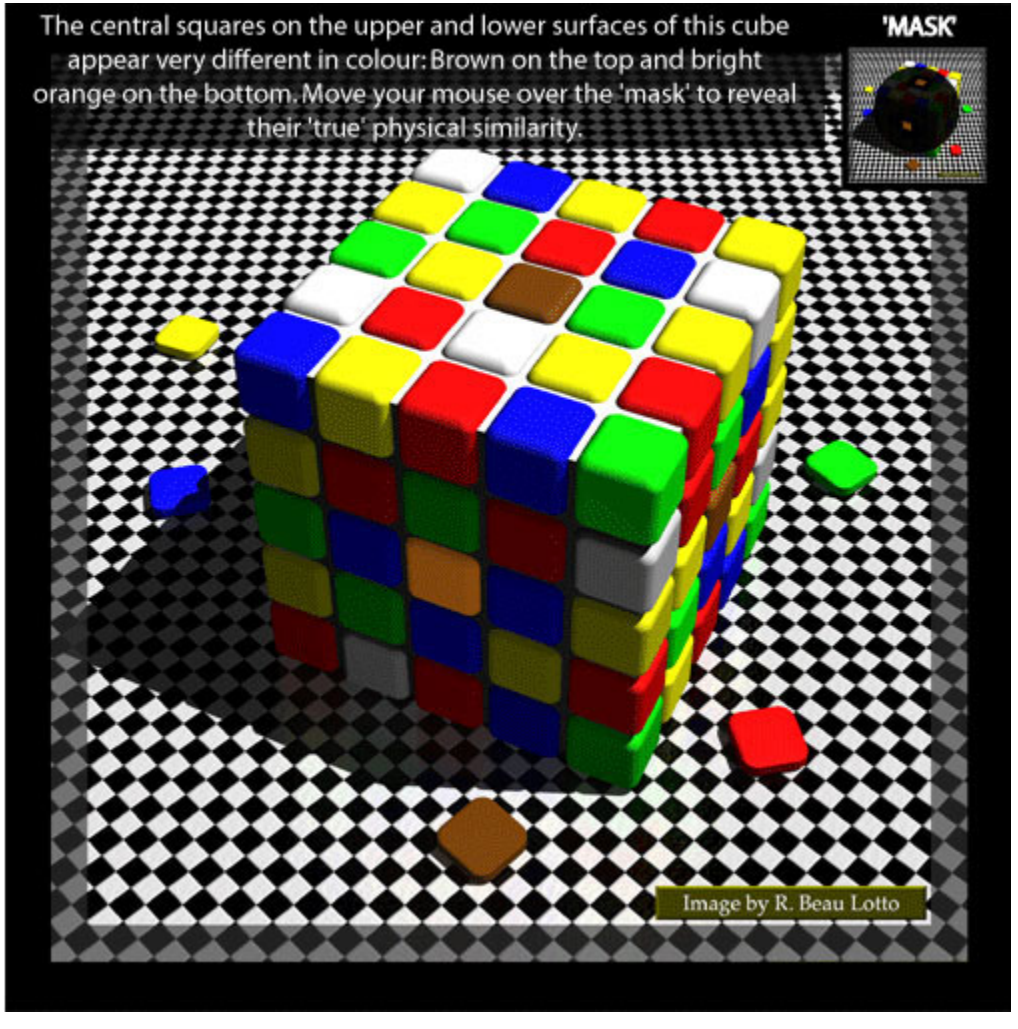
I want machines to ~~see~~

perceive

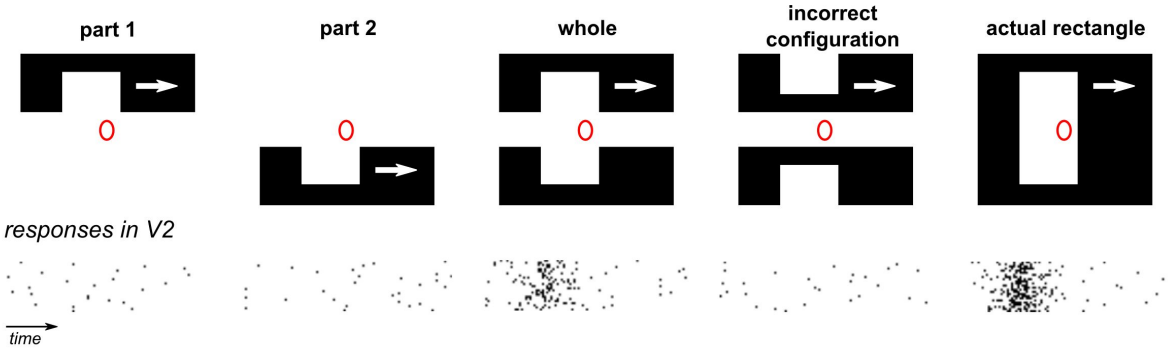
The dress

fair use – Wikipedia.png)

Color tile


The central squares on the upper and lower surfaces of this cube appear very different in colour: Brown on the top and bright orange on the bottom. Move your mouse over the 'mask' to reveal their 'true' physical similarity.
'MASK'
Image by R. Beau Lotto

Illusory contours

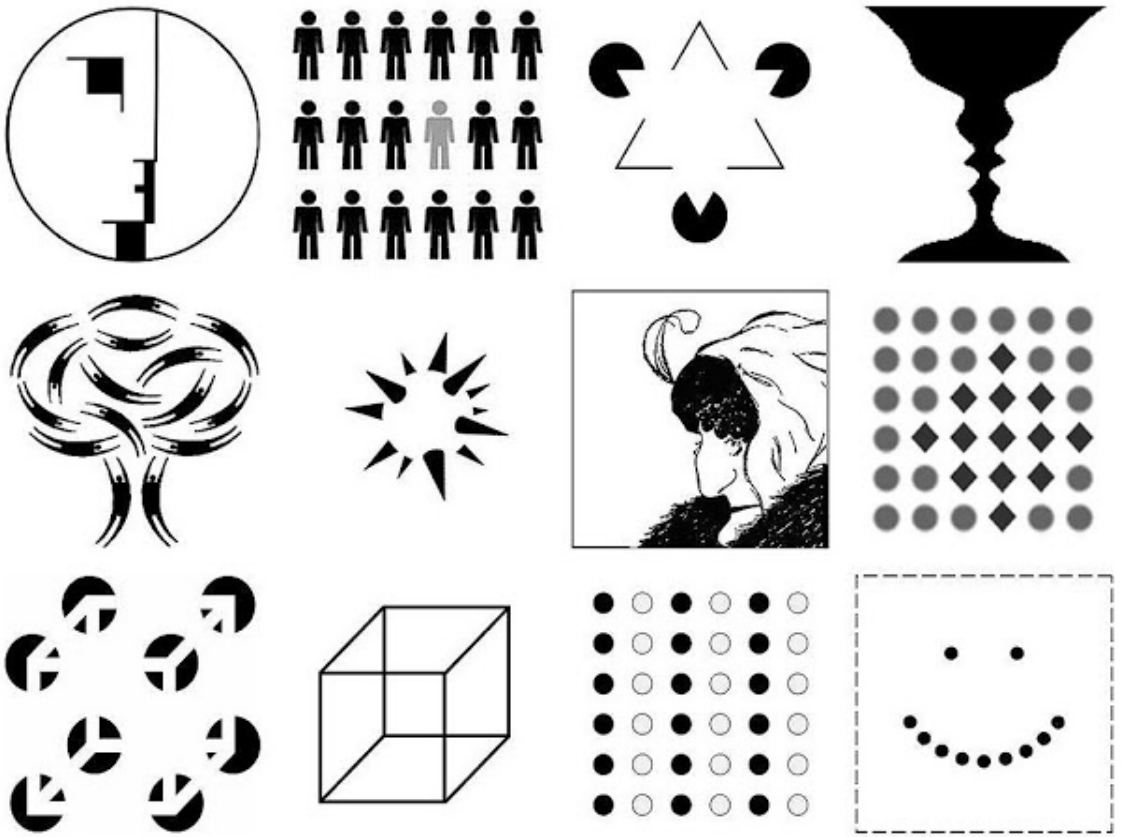


von der Heydt et al. (Science, 1984) / via Kubilius et al. (Cortex, in press)

Integration

- Need to put things together
- Feedforward architectures are not sufficient
- Must work on multiple tasks

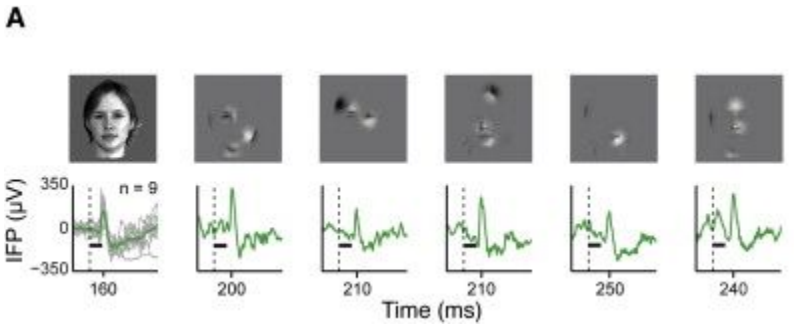
Gestalts



Border-ownership

fair use – Zhou et al. (The Journal of Neuroscience, 2000)

It takes time to perceive



fair use – Liu et al. (Neuron, 2014)

What does it look like?

fair use – LabelMe

What does it look like?

fair use – LabelMe

More tasks

The visual system does a lot of things:

- Categorization
- Navigation
- Acting on environment
- ...

A

B

C

D

Visual Turing test

1. Q: Is there a person in the blue region?	A: yes
2. Q: Is there a unique person in the blue region? (Label this person 1)	A: yes
3. Q: Is person 1 carrying something?	A: yes
4. Q: Is person 1 female?	A: yes
5. Q: Is person 1 walking on a sidewalk?	A: yes
6. Q: Is person 1 interacting with any other object?	A: no
⋮	
9. Q: Is there a unique vehicle in the yellow region? (Label this vehicle 1)	A: yes
10. Q: Is vehicle 1 light-colored?	A: yes
11. Q: Is vehicle 1 moving?	A: no
12. Q: Is vehicle 1 parked and a car?	A: yes
⋮	
14. Q: Does vehicle 1 have exactly one visible tire?	A: no
15. Q: Is vehicle 1 interacting with any other object?	A: no
17. Q: Is there a unique person in the red region?	A: no
18. Q: Is there a unique person that is female in the red region?	A: no
19. Q: Is there a person that is standing still in the red region?	A: yes
20. Q: Is there a unique person standing still in the red region? (Label this person 2)	A: yes
⋮	
23. Q: Is person 2 interacting with any other object?	A: yes
24. Q: Is person 1 taller than person 2?	A: amb.
25. Q: Is person 1 closer (to the camera) than person 2?	A: no
26. Q: Is there a person in the red region?	A: yes
27. Q: Is there a unique person in the red region? (Label this person 3)	A: yes
⋮	
36. Q: Is there an interaction between person 2 and person 3?	A: yes
37. Q: Are person 2 and person 3 talking?	A: yes

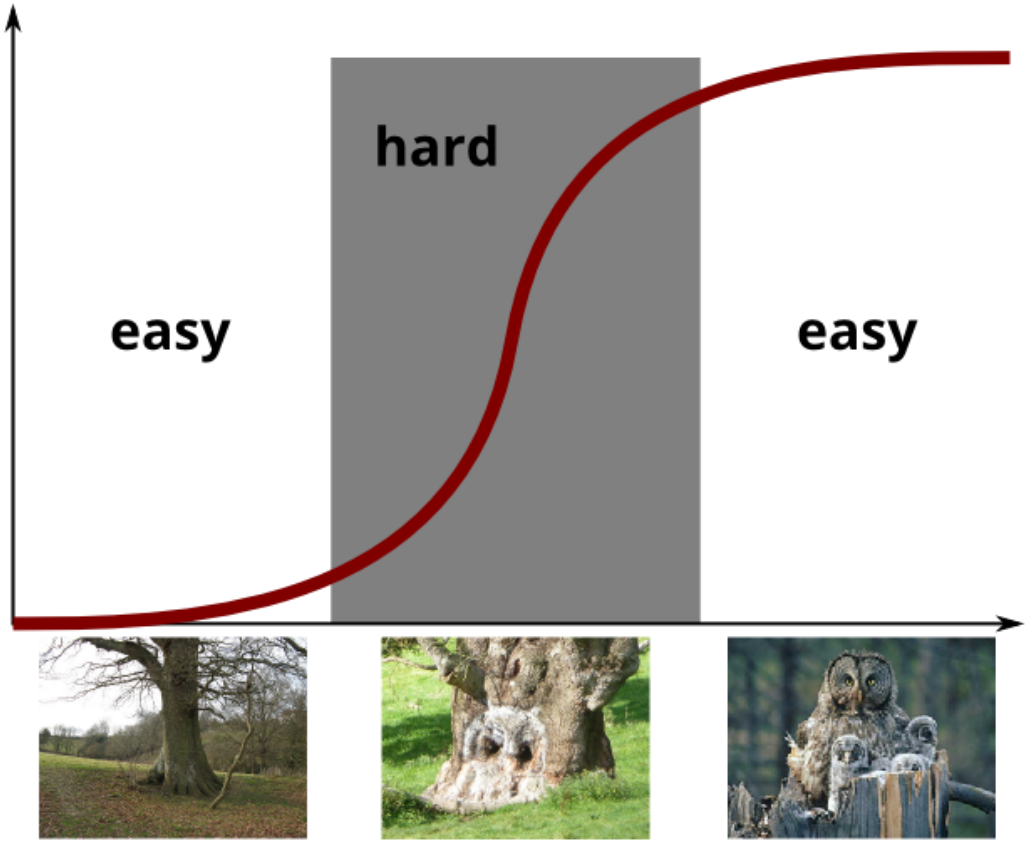
fair use – Geman et al. (PNAS, 2015)

A machine who mistook a knife for a cat

or why understanding human perception matters

Decision making

Decisions are difficult when only partial or ambiguous information is available.



The trolley problem

the switch
Foot, 1967

the fat man
Thomson, 1976

the fat villain

the loop
Costa, 1987

the man in the yard
Unger, 1992

Tools must act clever

1. Tools are too limited (e.g., default location in Google Maps)
2. Programmers often are bad at accounting for human nature (e.g., awful GUIs)

Thank you!

slides available at klab.lt