Teaching machines to see

A quest to visual intelligence

Lamberto Ballan

University of Padova, Italy

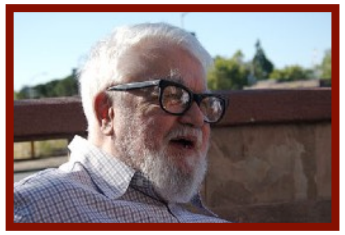
Open day: Math &
Computer Science

DIPARTIMENTO
MATEMATICA
Department of Mathematics "Tullio Levi-Civita"

Artificial Intelligence

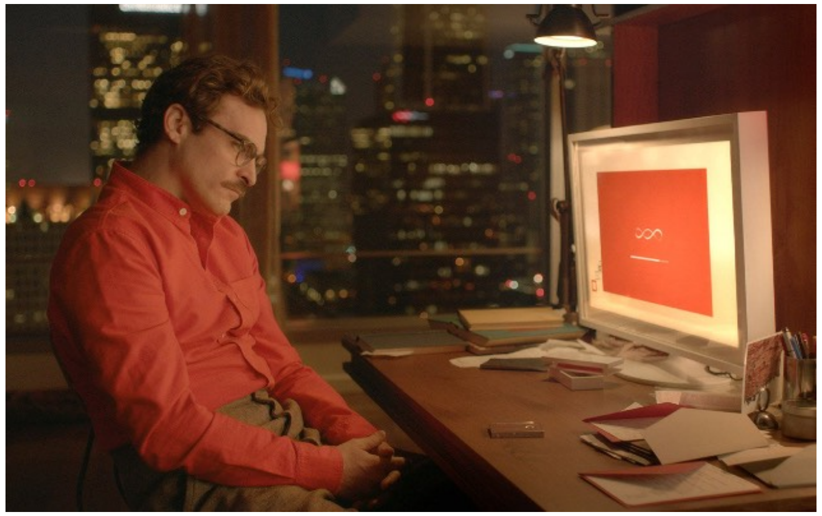
Alan Turing

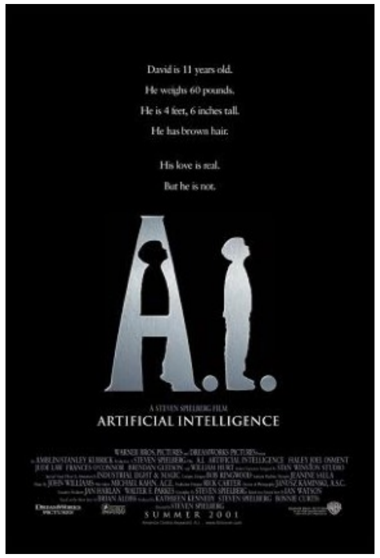
1950



1955

John McCarthy





IBM

Google

nature
At last – a computer program that can beat a champion Go player
ALL SYSTEMS GO

2018

Eyes and Vision

HAWK CLOSE-UP BY IBLACKSTOCK WWW.CRESTOCK.COM

What is (computer) vision?

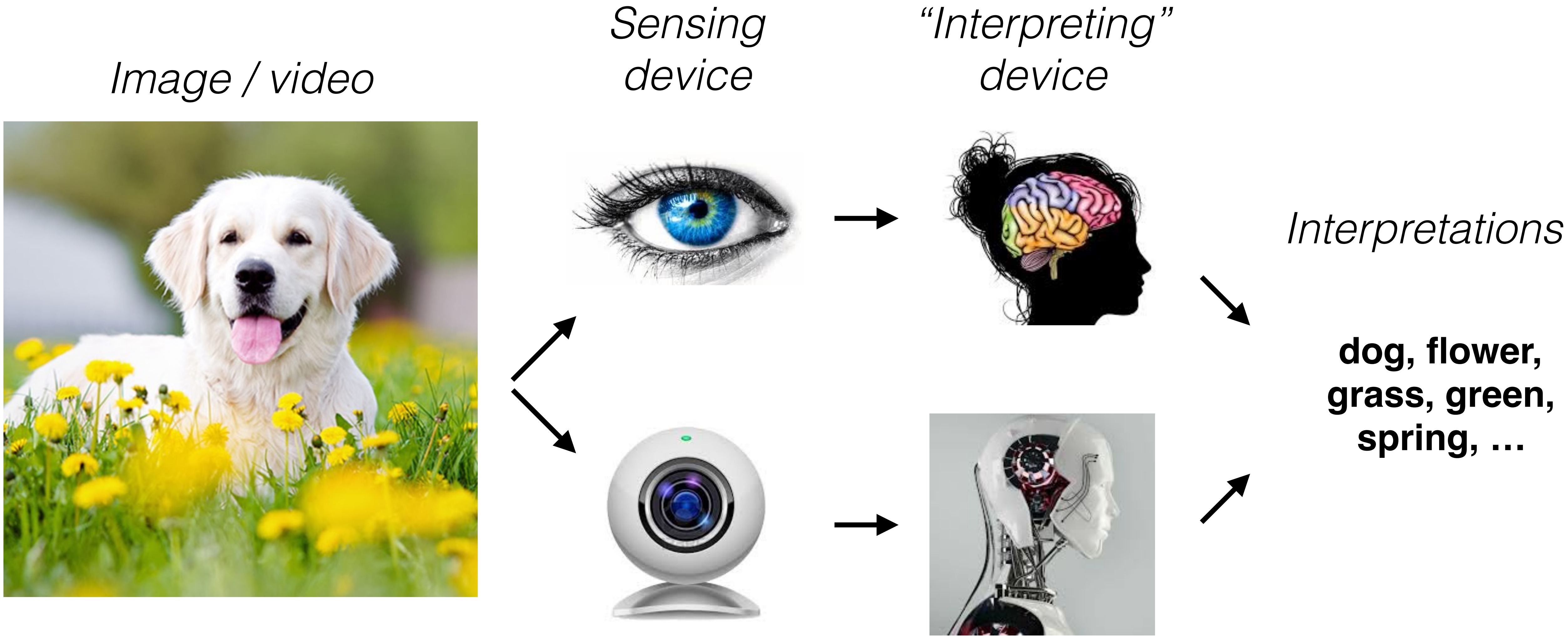


Image / video

Sensing device

“Interpreting” device

Interpretations

dog, flower, grass, green, spring, …

The goal of computer vision

- Bridge the gap between pixels and “meaning”

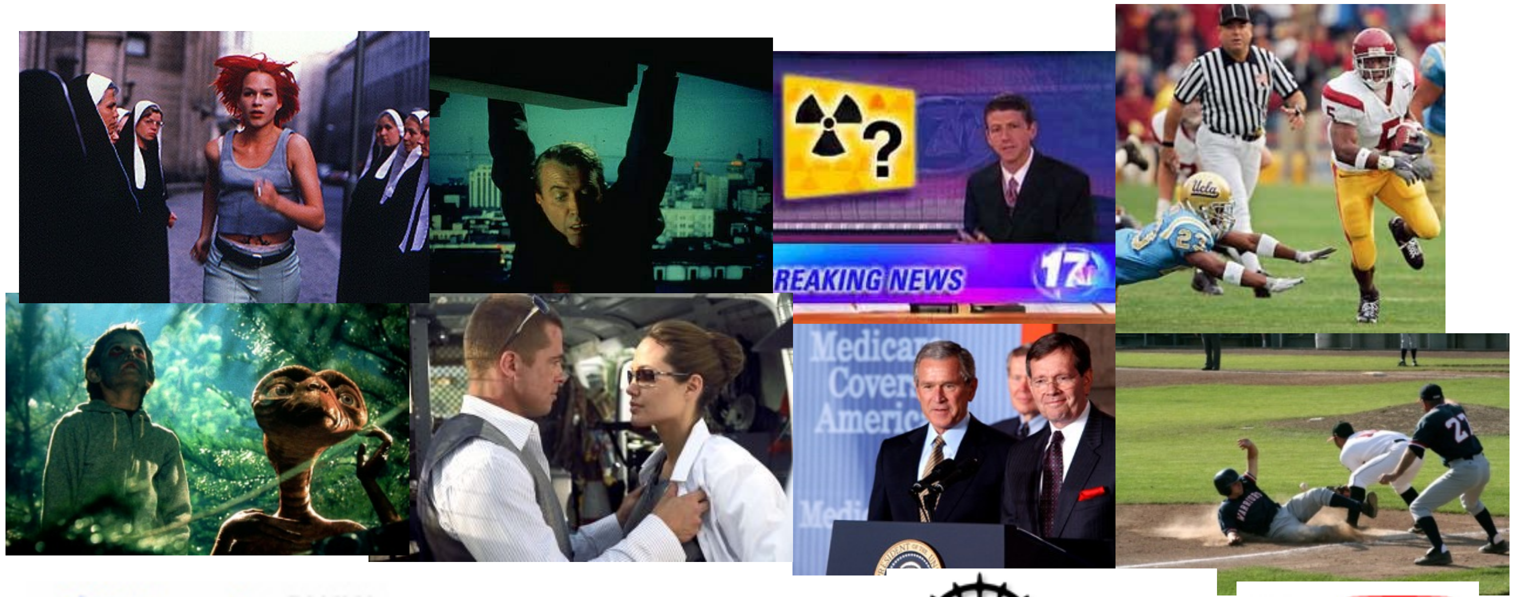
What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer see

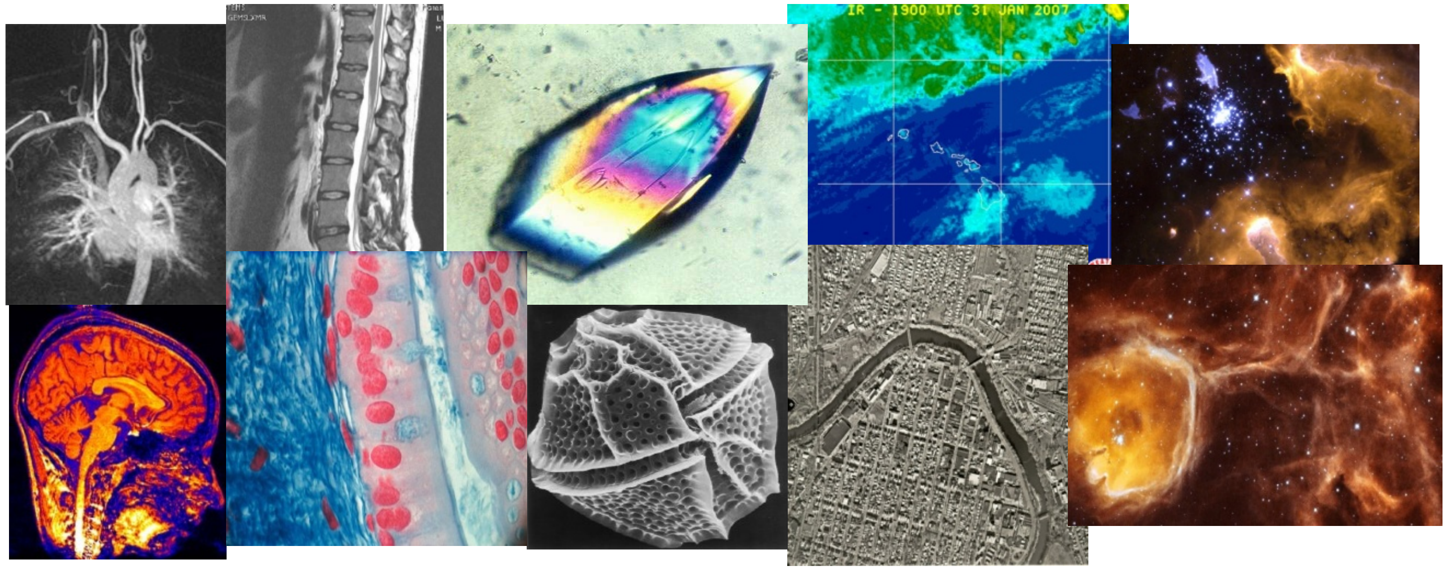
Why study computer vision?

- Big visual data: almost 90% of web data is visual



Google Image Search · Google Photos · flickr · webshots · picsearch · YouTube Broadcast Yourself

Surveillance and security

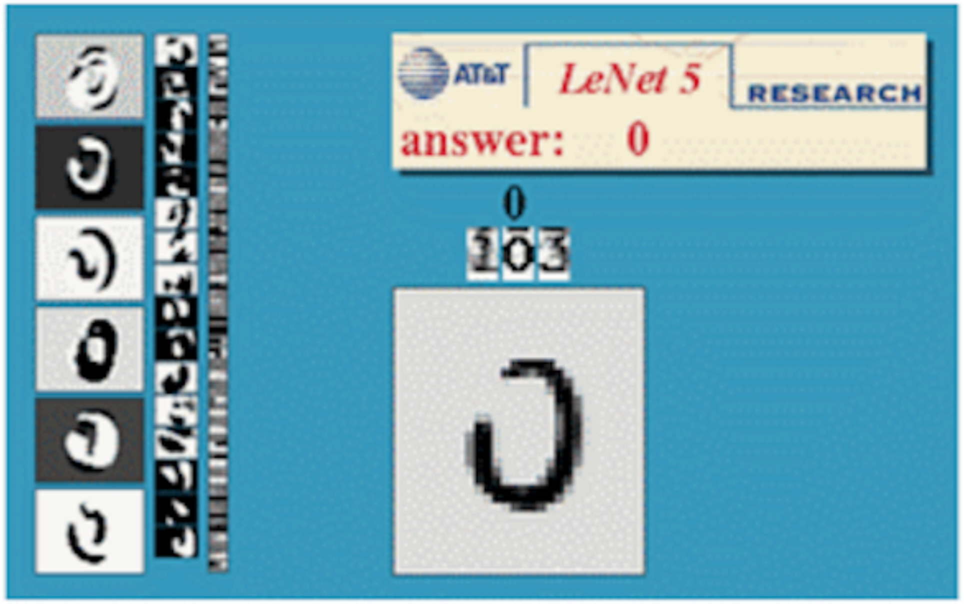


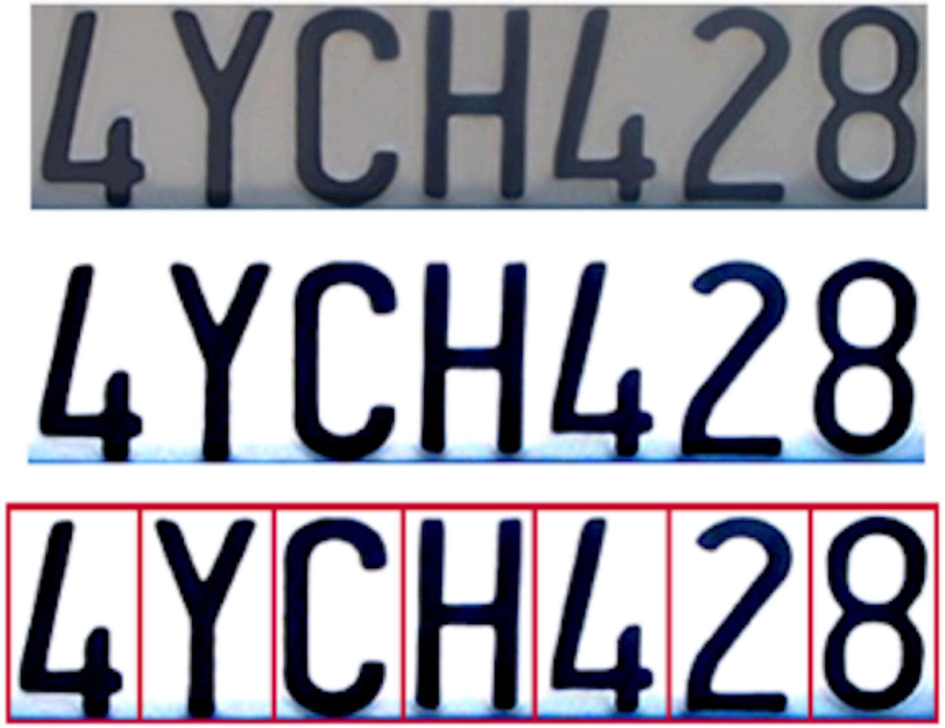
Medical and scientific images

Image credit: L. Fei-Fei

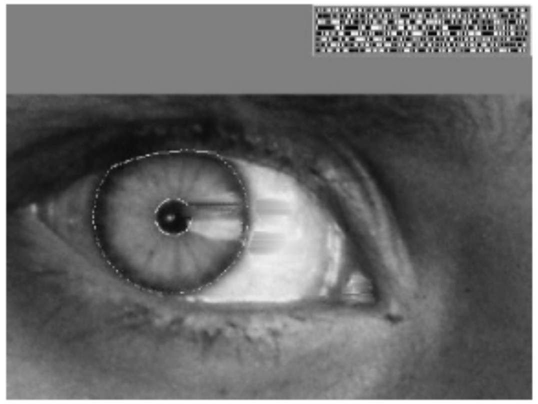
Applications

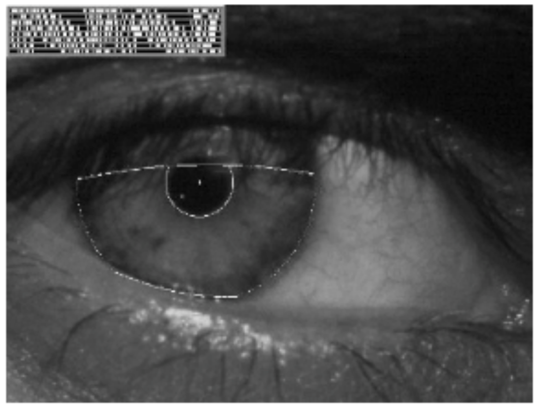
Digit recognition

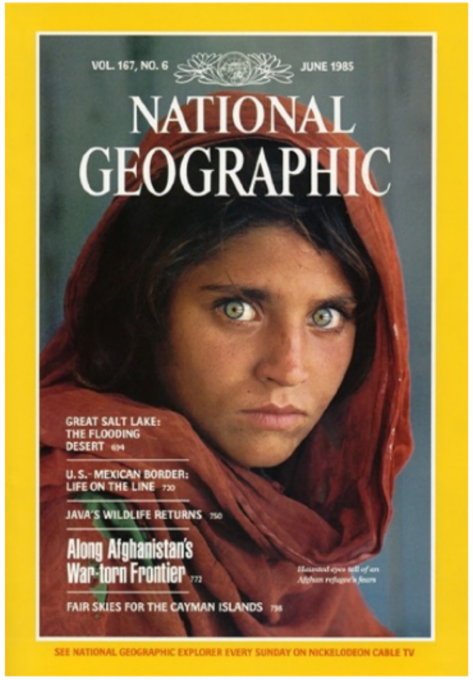




Biometric

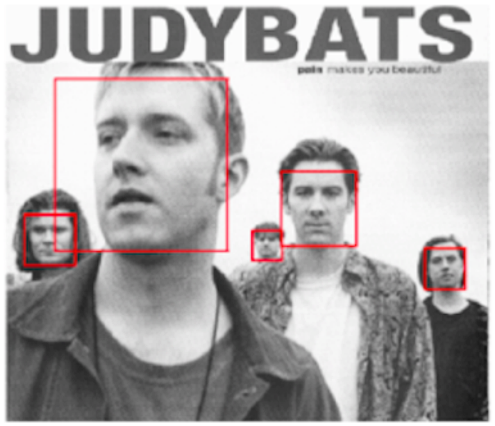


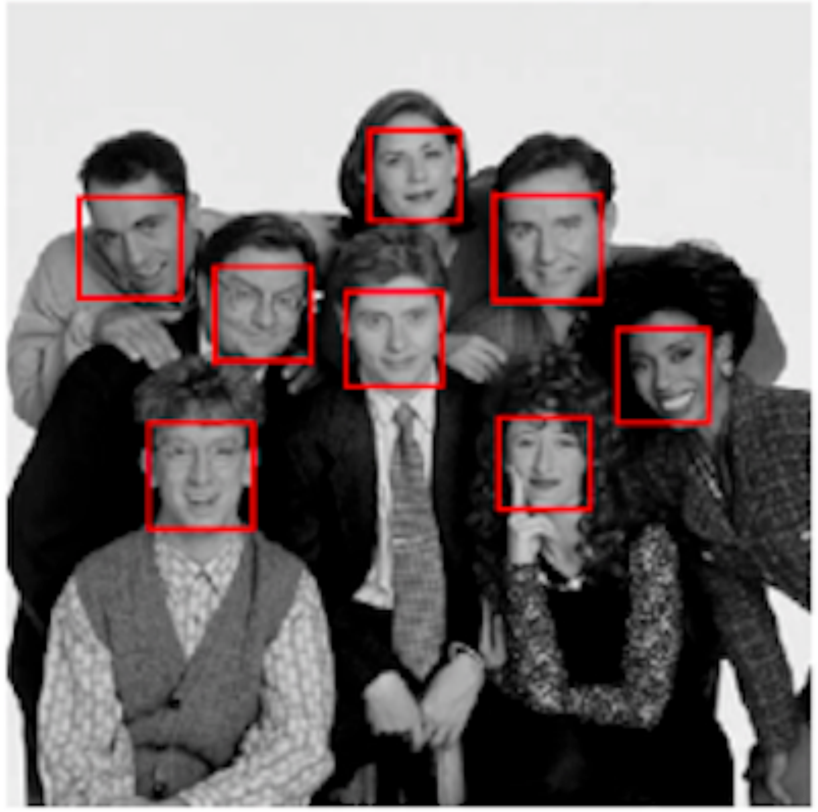



VOL. 167, NO. 6
JUNE 1985
NATIONAL GEOGRAPHIC
GREAT SALT LAKE: THE FLOODING DESERT
U.S.-MEXICAN BORDER: LIFE ON THE LINE
JAVA'S WILDLIFE RETURNS
Along Afghanistan's War-torn Frontier
FAIR SKIES FOR THE CAYMAN ISLANDS
SEE NATIONAL GEOGRAPHIC EXPLORER EVERY SUNDAY ON NICKELODEON CABLE TV

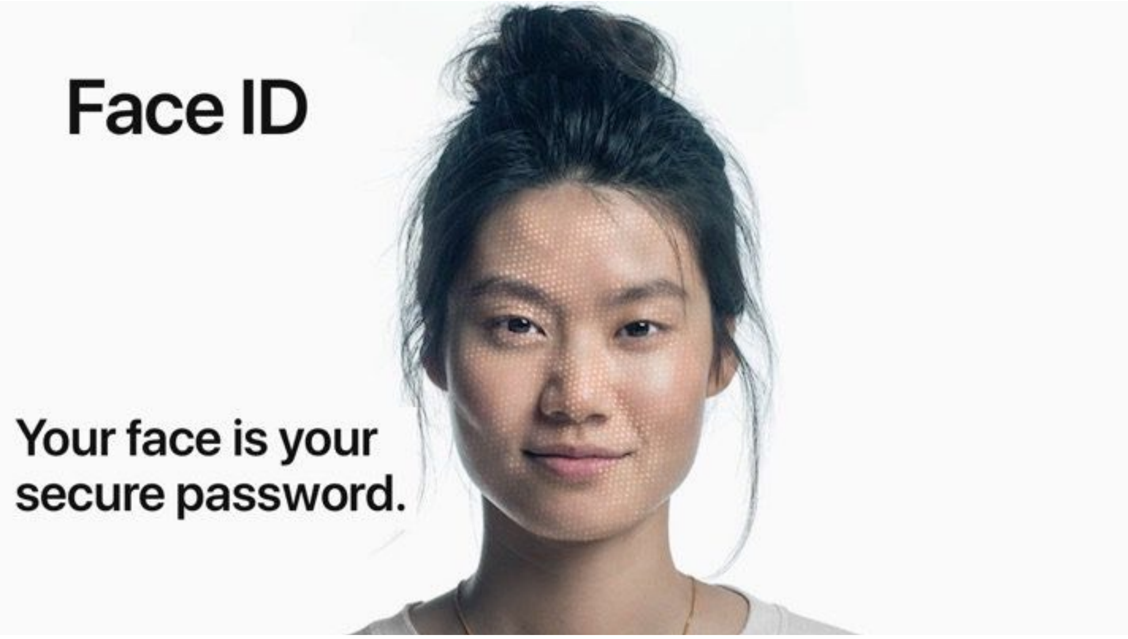
Applications

Face detection





Face recognition



Applications

Visual search

Google Goggles

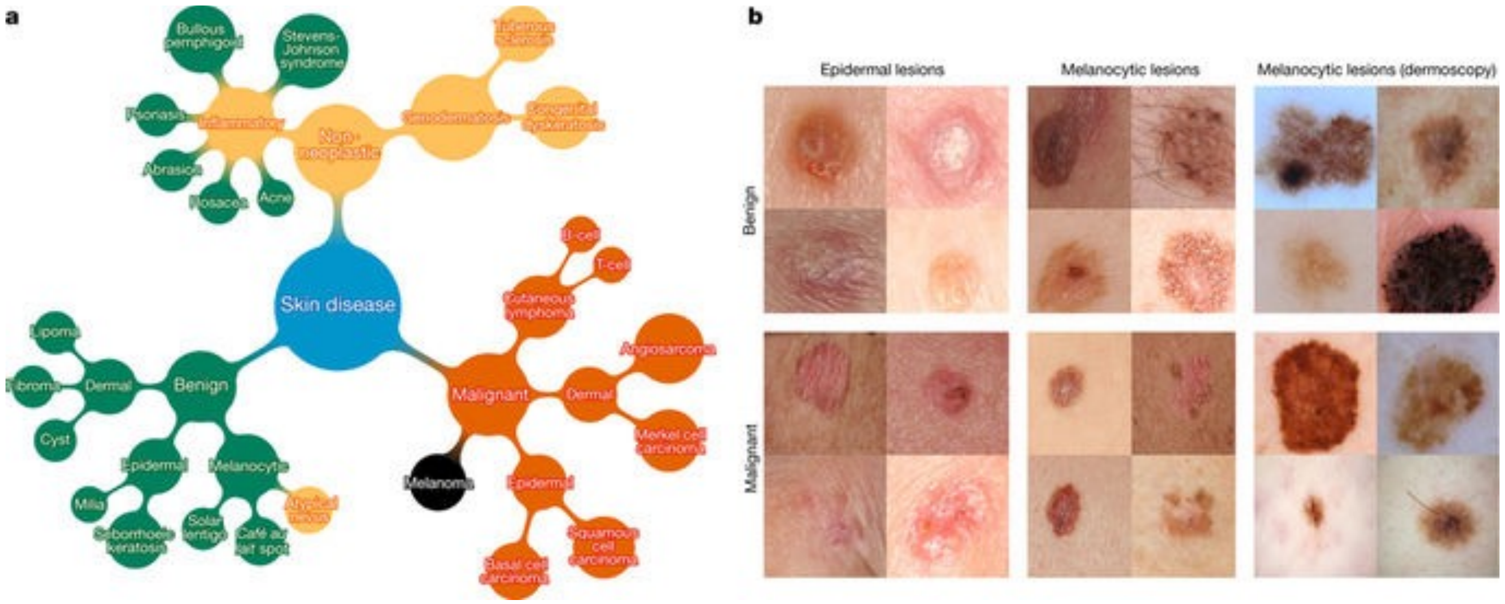
Scanning...

Masterworks of Art - Frida Kahlo and Di...

Art

LOST IN SHANGRI-LA
Lost in Shangri-La: A True Story of Survival, Adven...
Price: $16.03
Shipping: $3.99
Ships from & sold by Amazon.com
Buy Now
MITCHELL ZUCKOFF

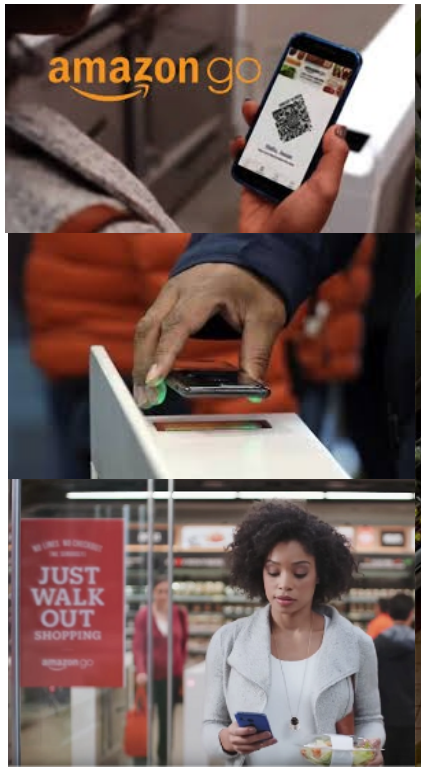
Healthcare

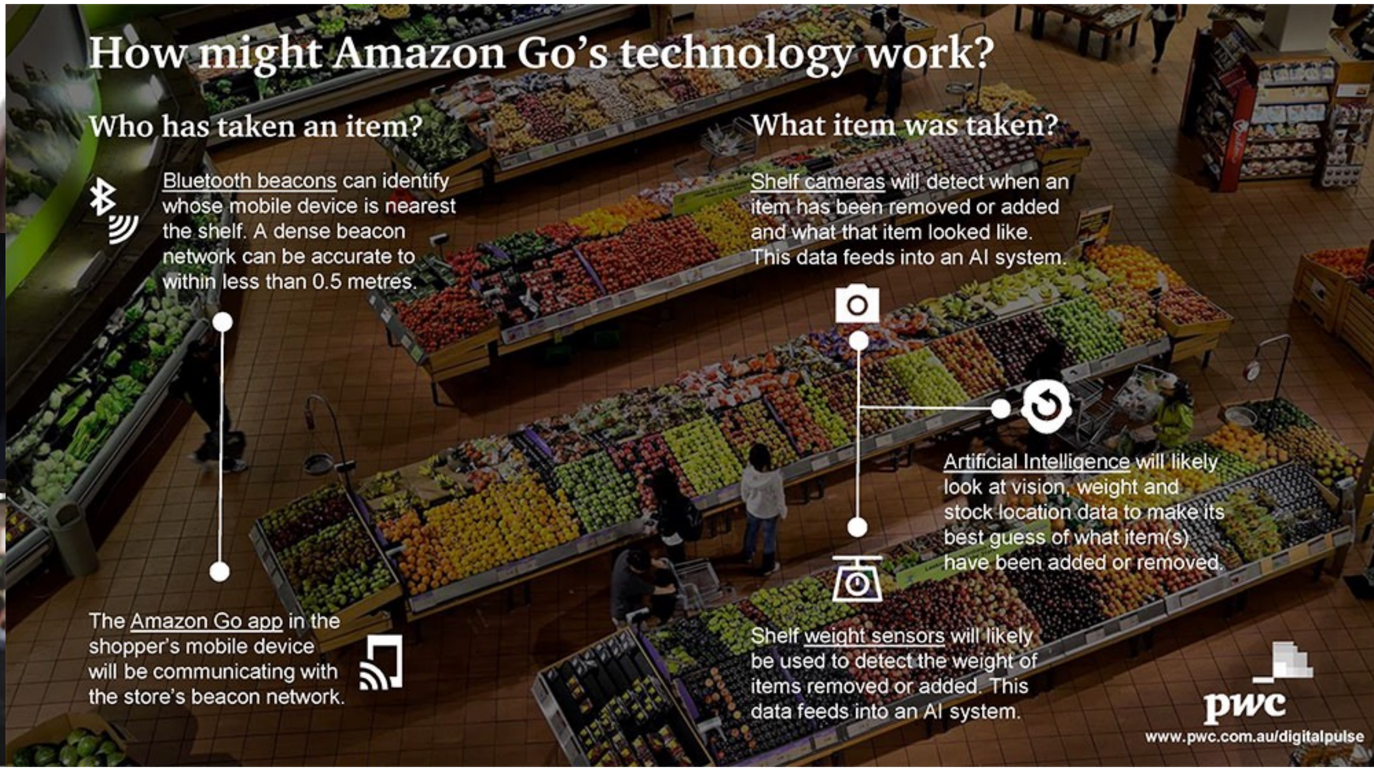


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
LESIONS LEARNT
Artificial intelligence powers detection of skin cancer from images

Applications (not far from now…)

Retail/Supermarket





How might Amazon Go's technology work?

Who has taken an item?

Bluetooth beacons can identify whose mobile device is nearest the shelf. A dense beacon network can be accurate to within less than 0.5 metres.

The Amazon Go app in the shopper's mobile device will be communicating with the store's beacon network.

What item was taken?

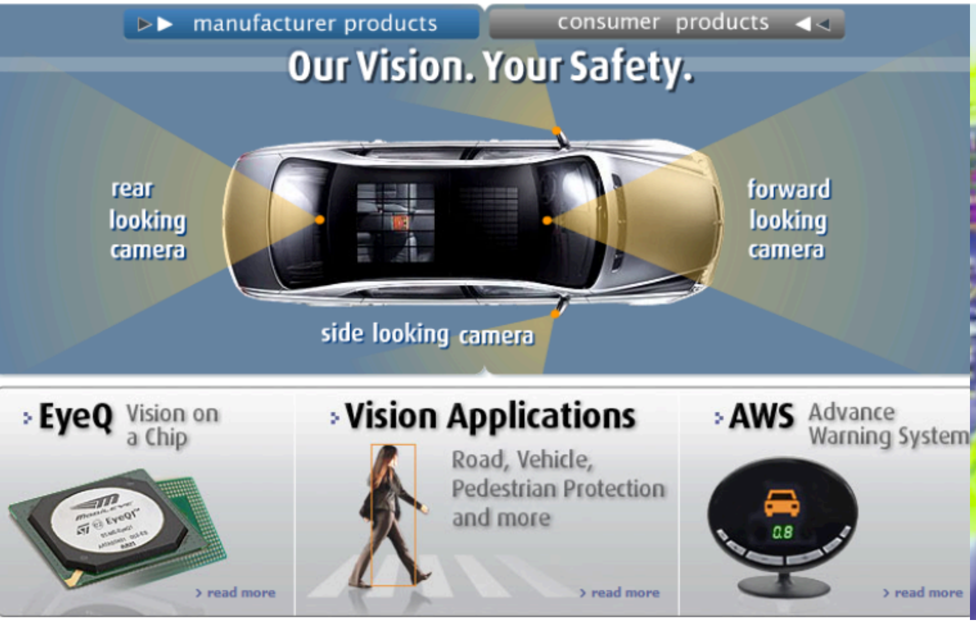
Shelf cameras will detect when an item has been removed or added and what that item looked like. This data feeds into an AI system.

Artificial Intelligence will likely look at vision, weight and stock location data to make its best guess of what item(s) have been added or removed.

Shelf weight sensors will likely be used to detect the weight of items removed or added. This data feeds into an AI system.

pwc
www.pwc.com.au/digitalpulse

Self-driving cars



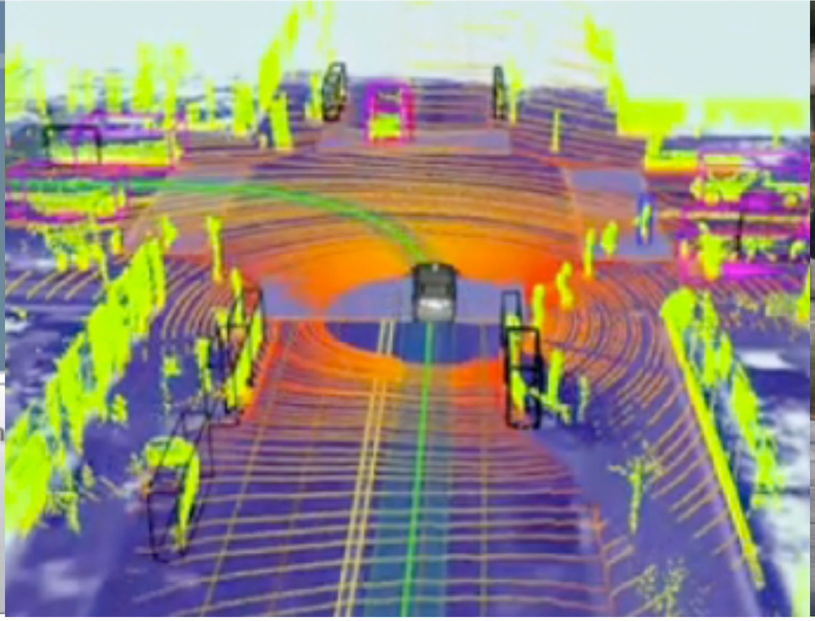
manufacturer products | consumer products

Our Vision. Your Safety.

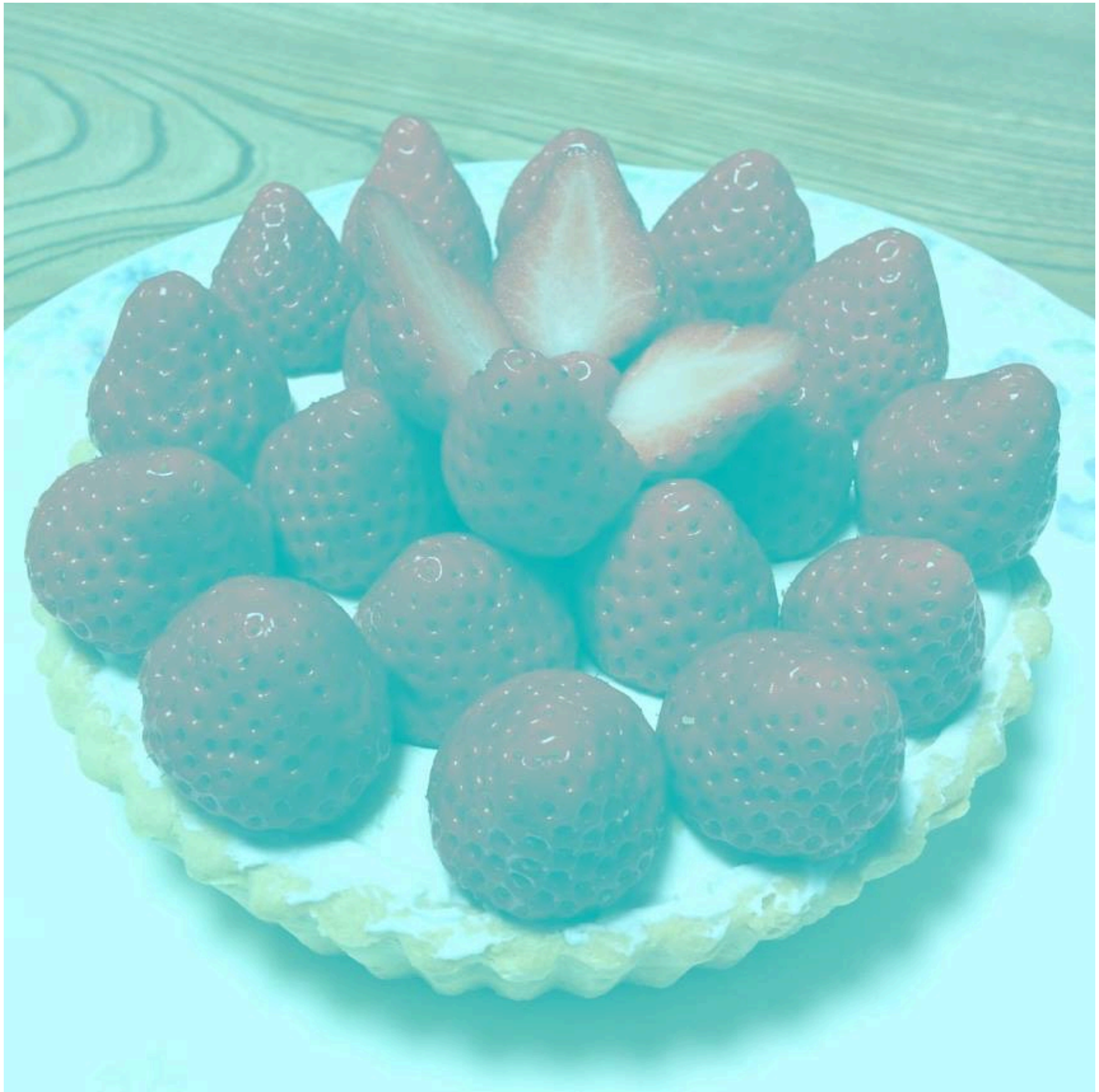
EyeQ Vision on a Chip
› read more

Vision Applications
Road, Vehicle, Pedestrian Protection and more
› read more

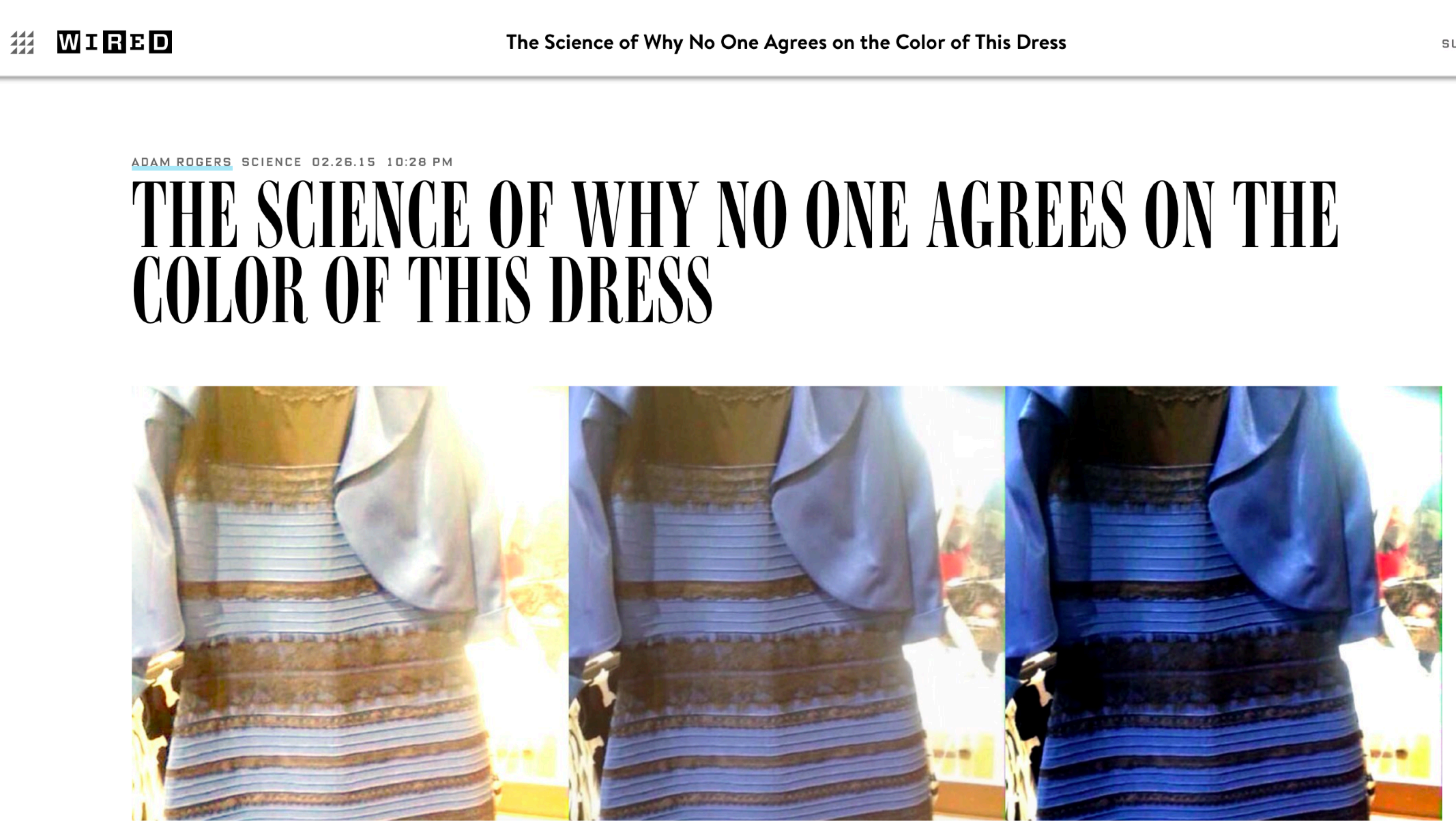
AWS Advance Warning System
› read more



Vision is challenging: *let's play a game (1)*



Vision is challenging: *let’s play a game (1)*



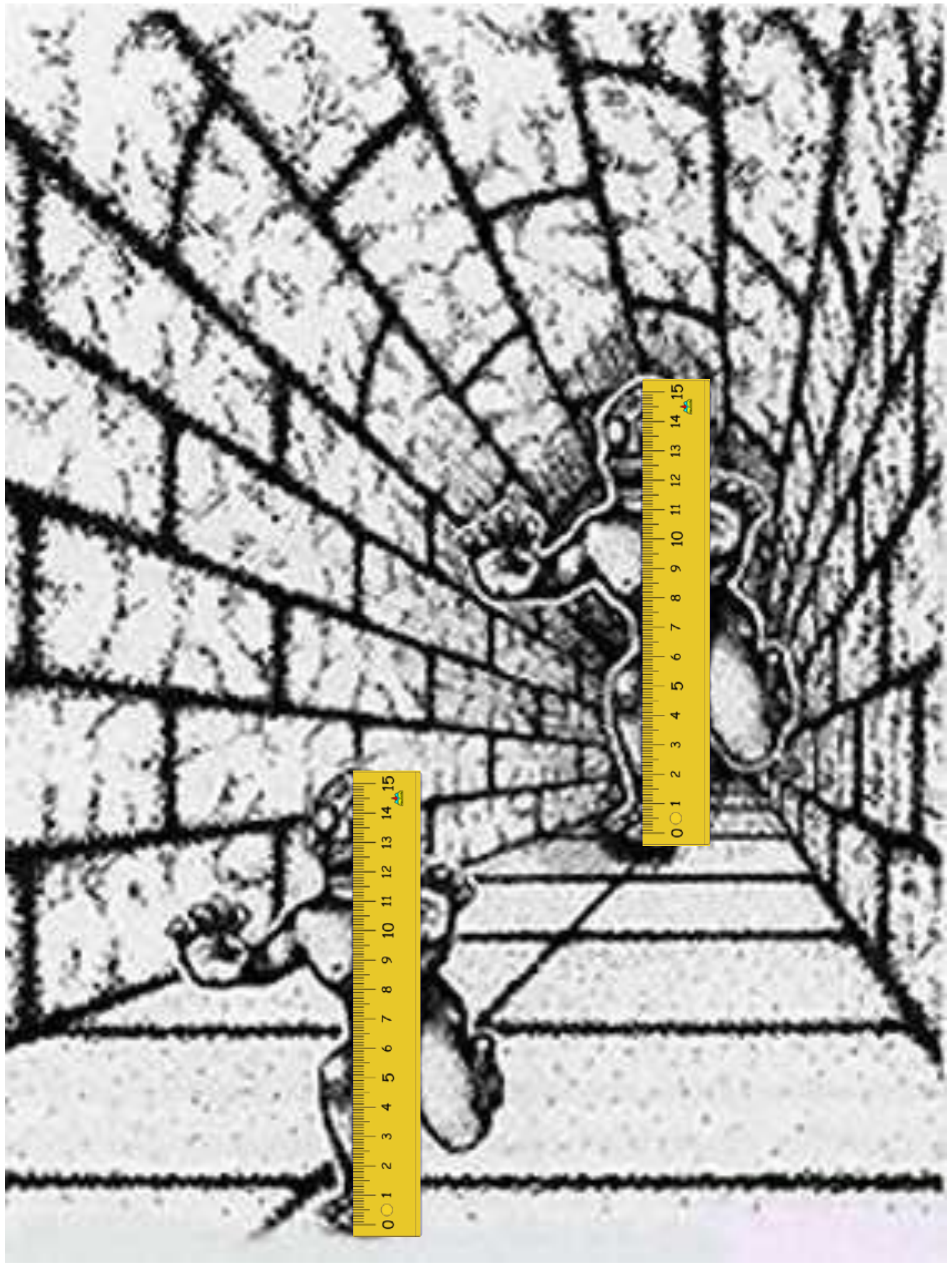
WIRED The Science of Why No One Agrees on the Color of This Dress SU

ADAM ROGERS SCIENCE 02.26.15 10:28 PM

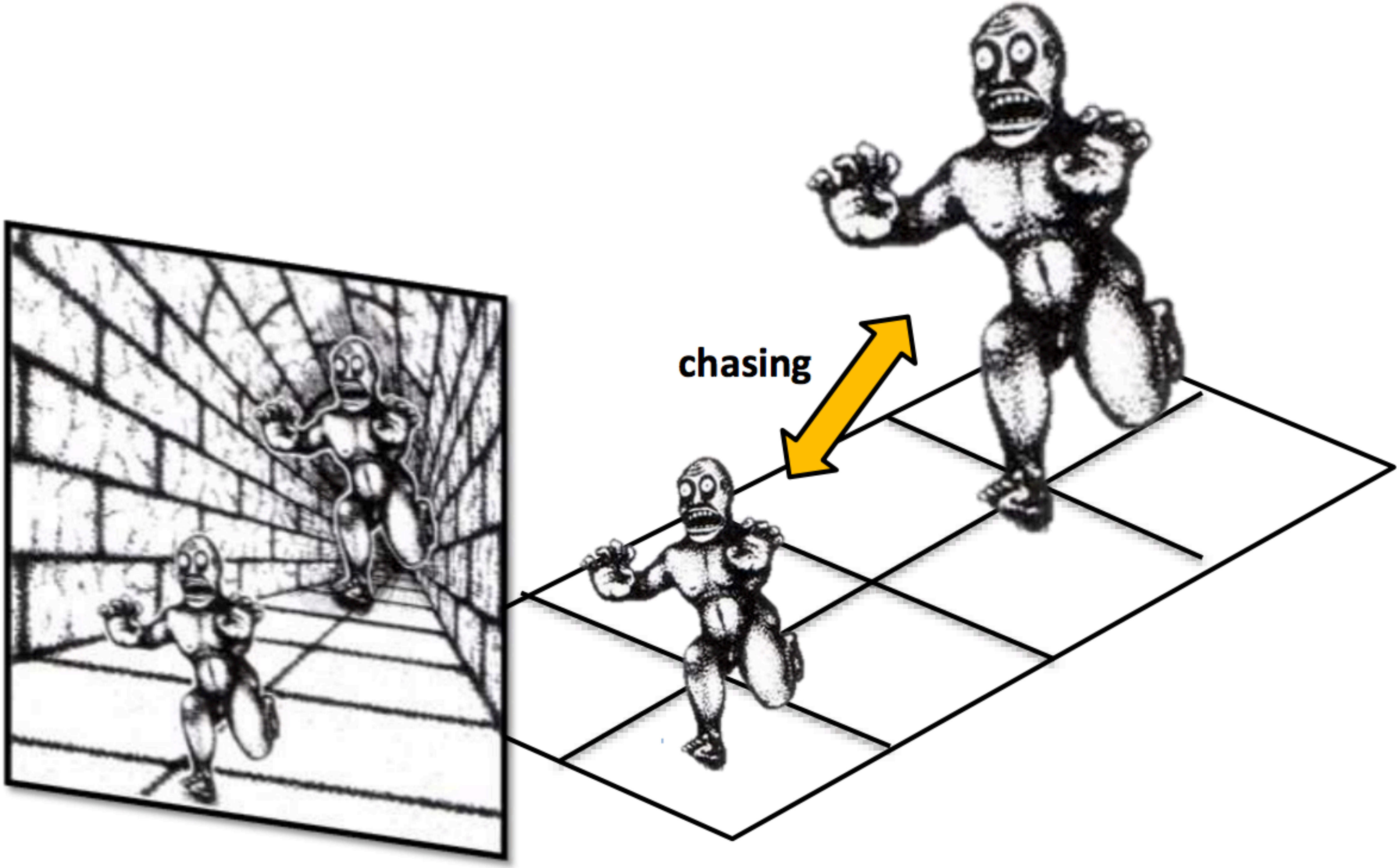
THE SCIENCE OF WHY NO ONE AGREES ON THE COLOR OF THIS DRESS

Feb 2015: people across social media have been arguing about whether this picture depicts a dress as blue with black lace fringe or white with gold lace fringe

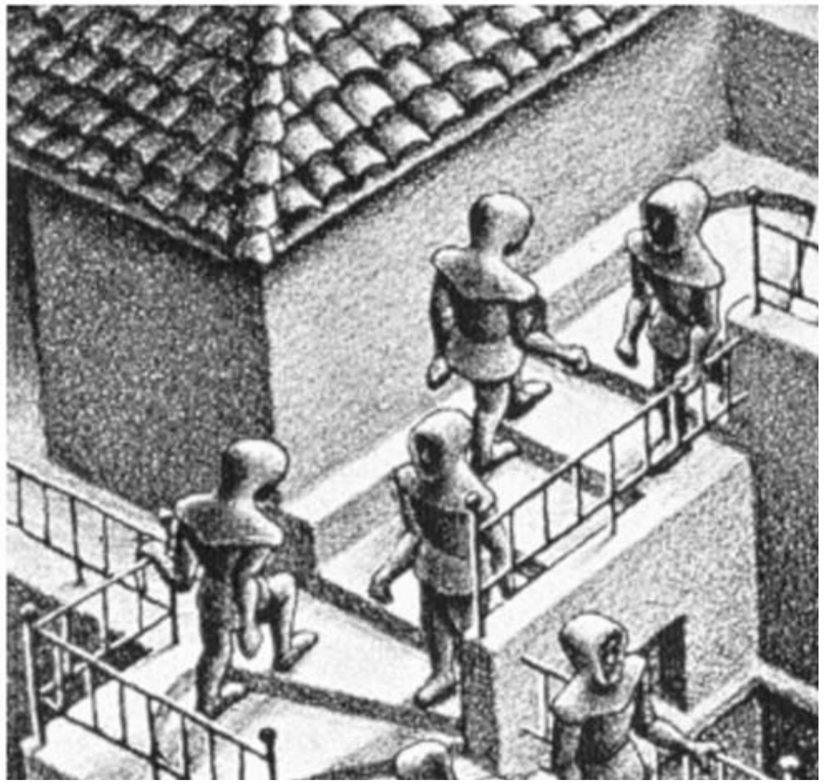
Vision is challenging: *let's play a game (2)*



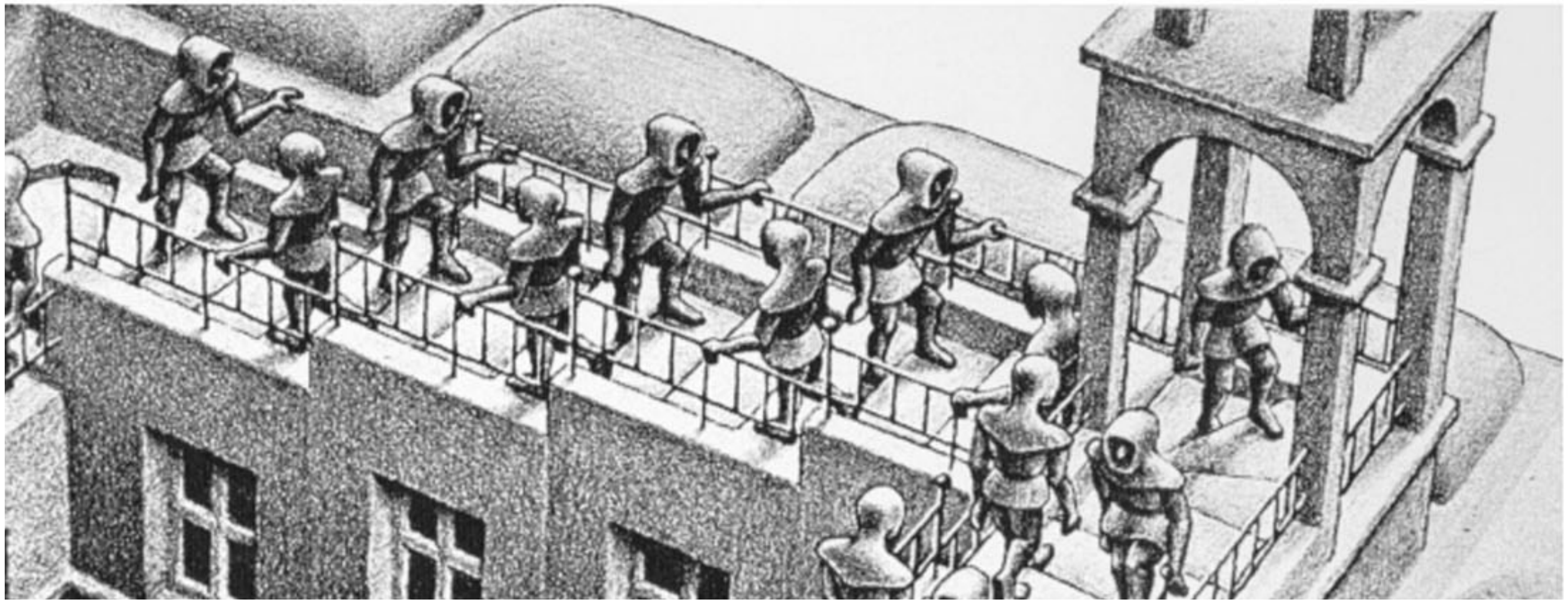
Vision is challenging: *let's play a game (2)*



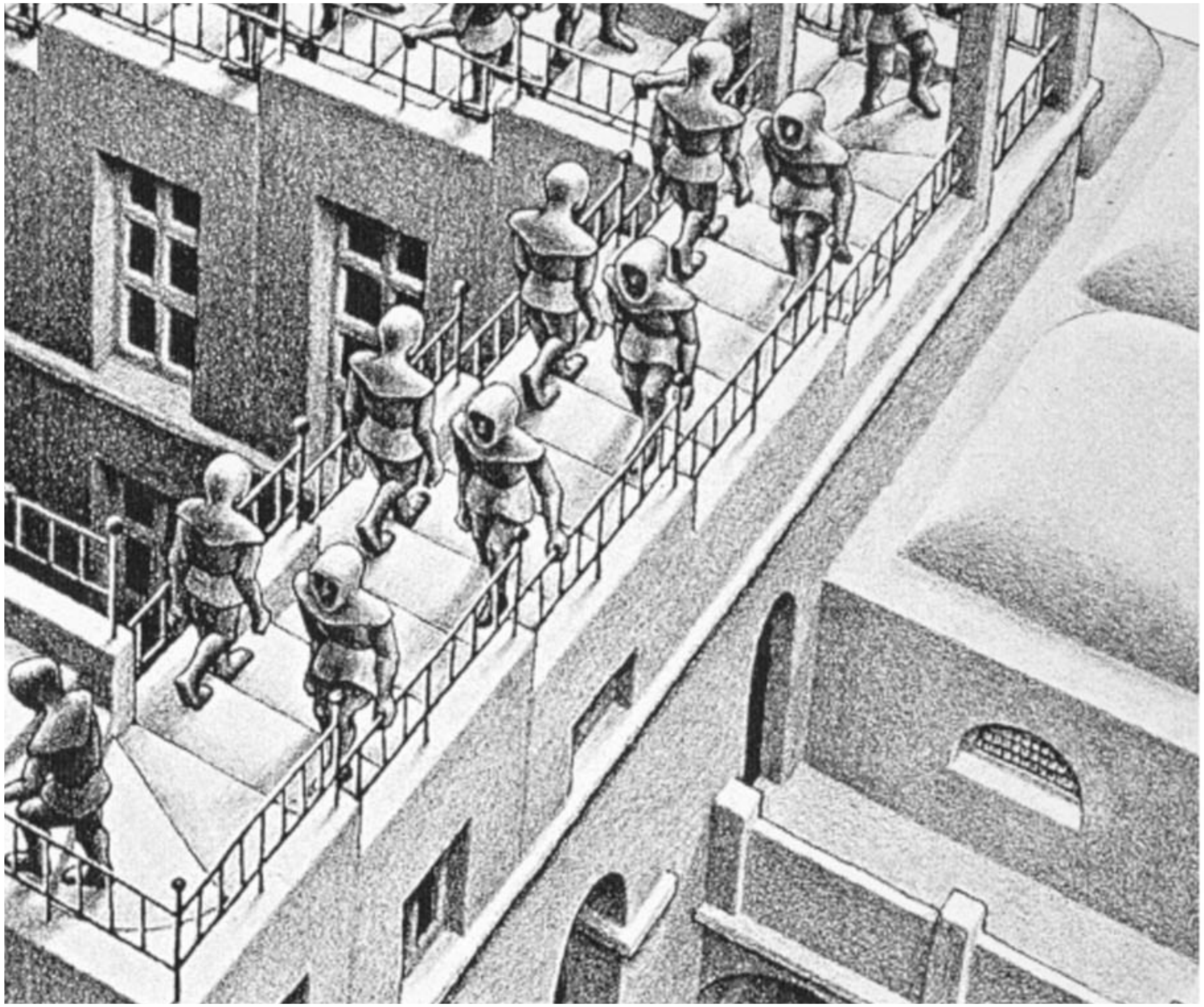
Vision is challenging: *let's play a game (3)*



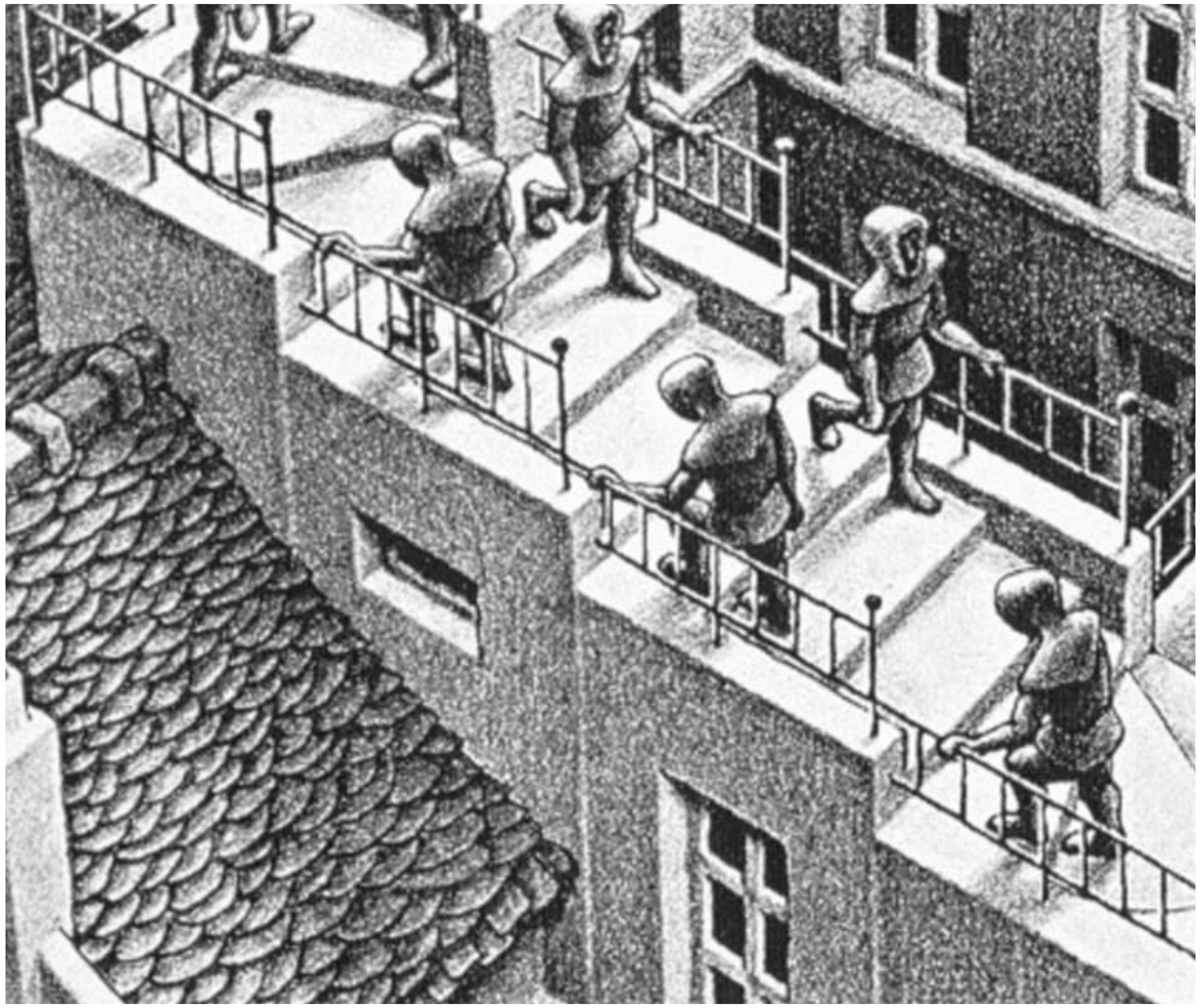
Vision is challenging: *let's play a game (3)*



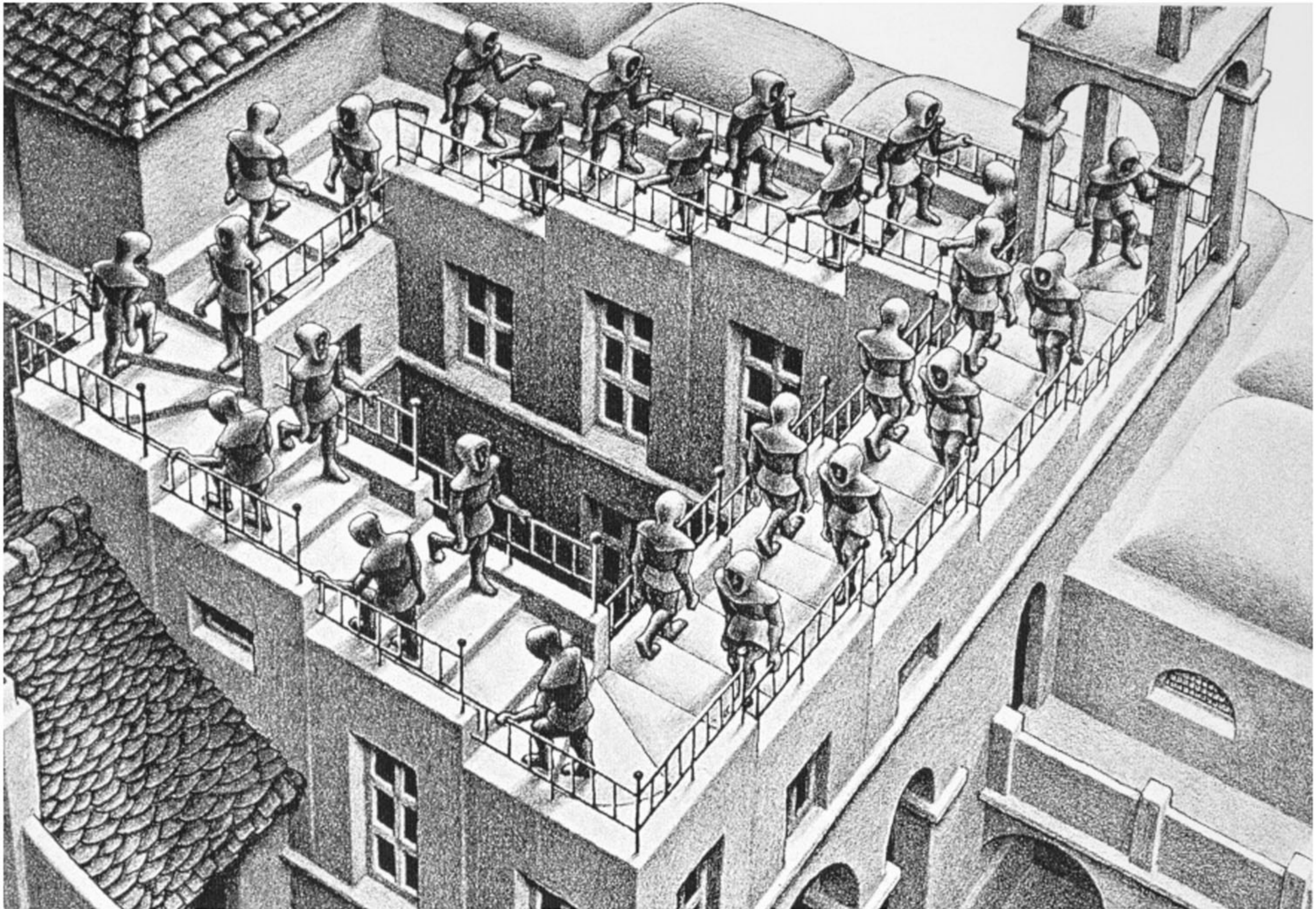
Vision is challenging: *let's play a game (3)*



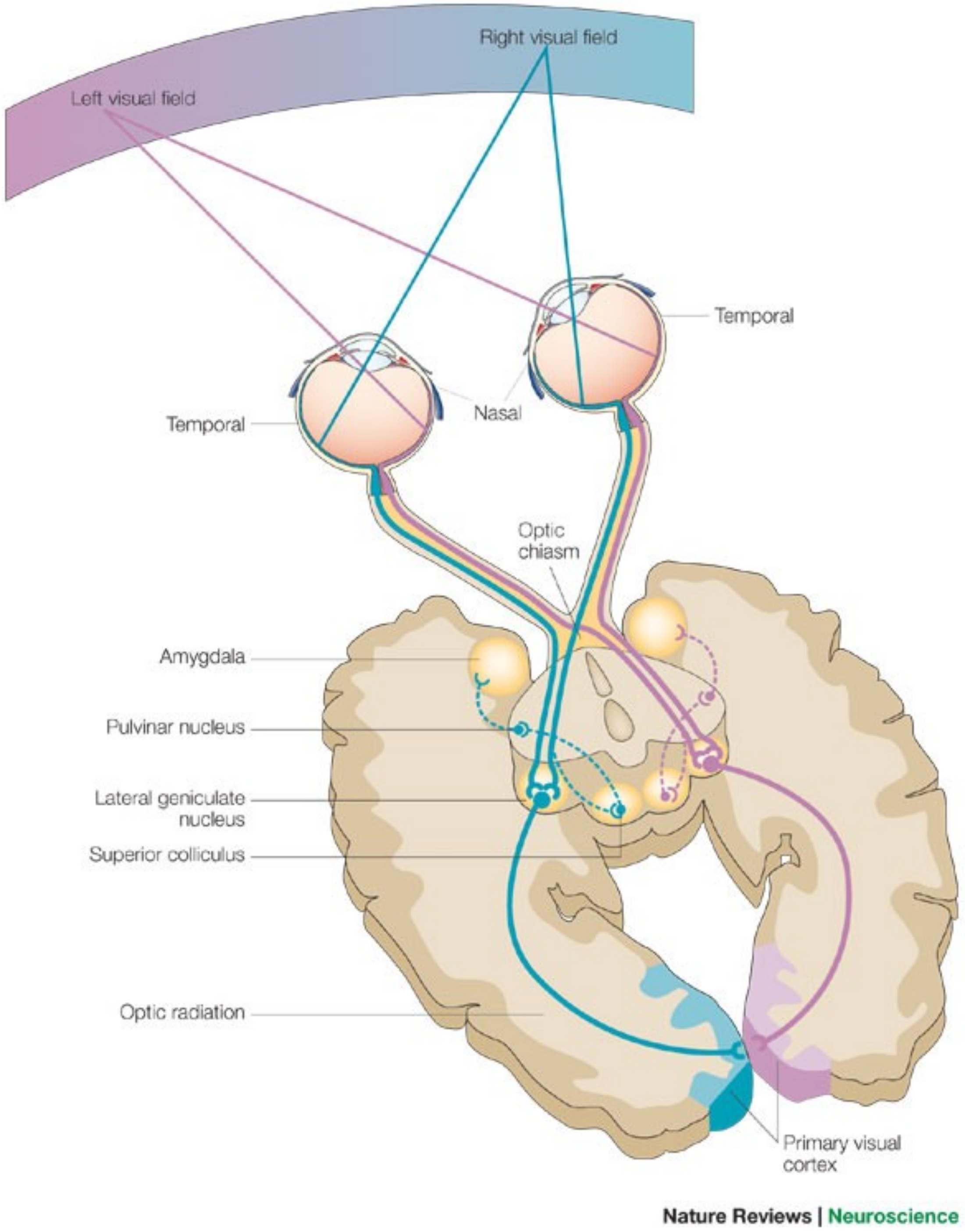
Vision is challenging: *let's play a game (3)*



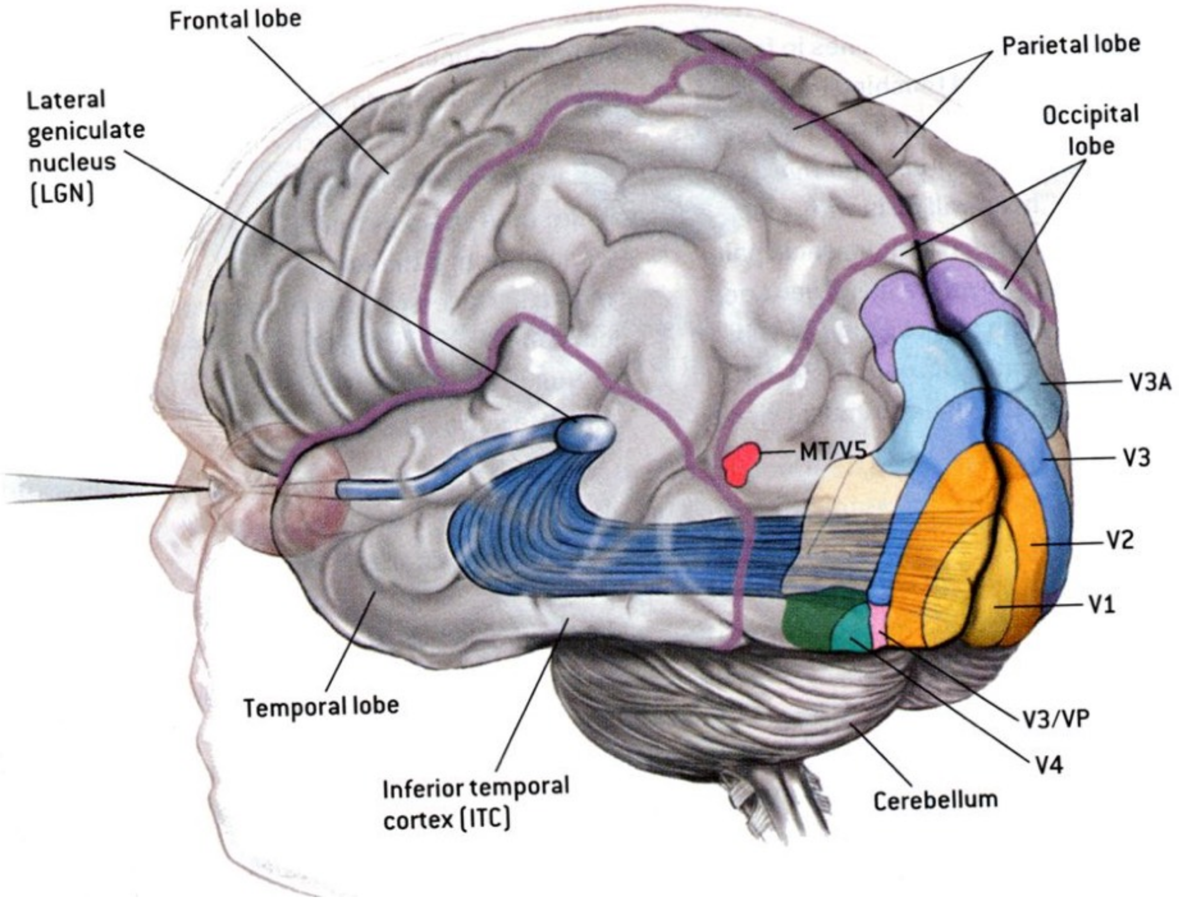
Vision is challenging: *let's play a game (3)*



Vision begins with the eyes…

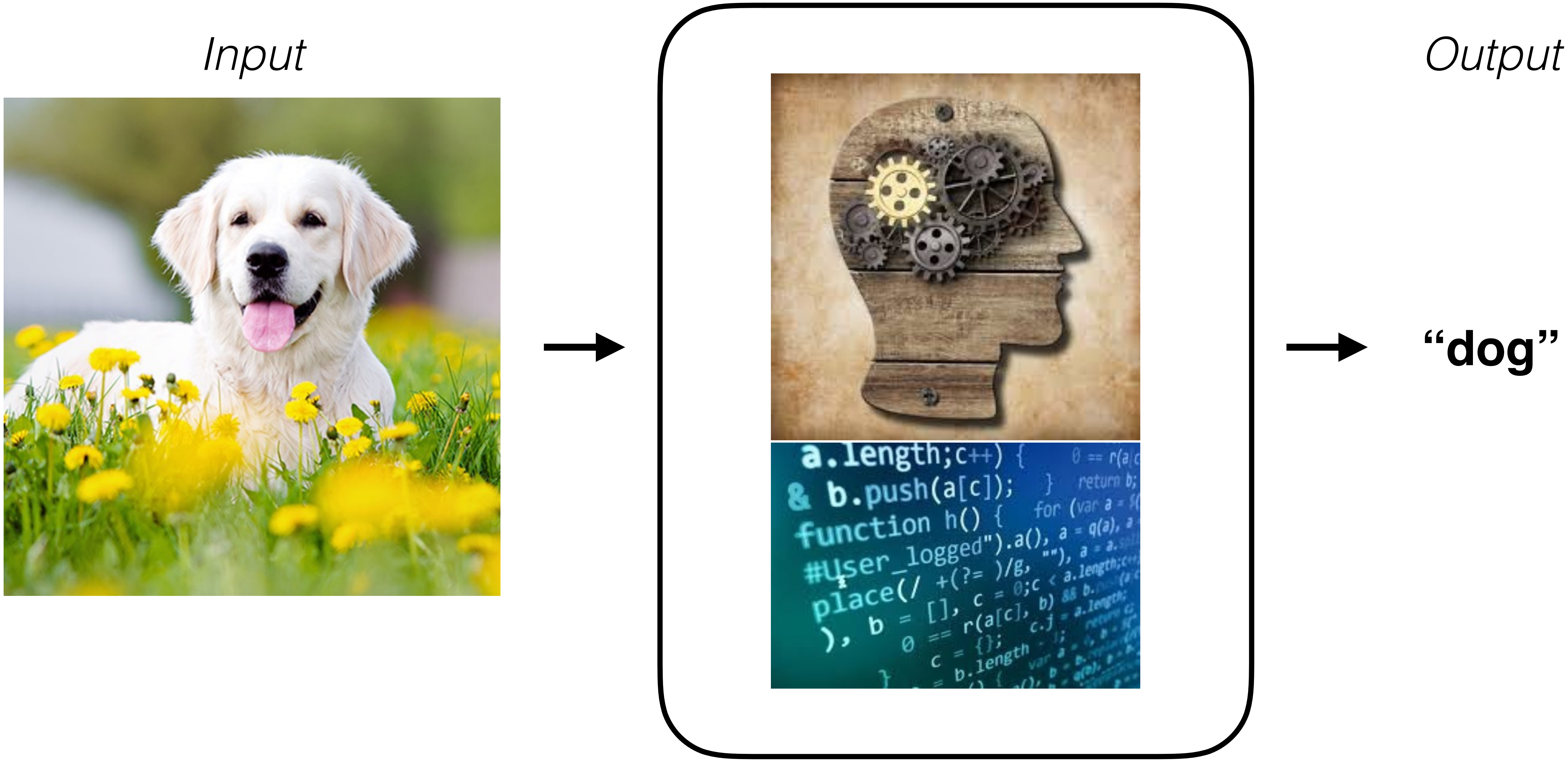


… but truly occurs in the brain!



Functional specialization:
V1: primary visual cortex
V4: color, V3/VP: recognition
MT/V5: motion

Teaching machines to see



Input

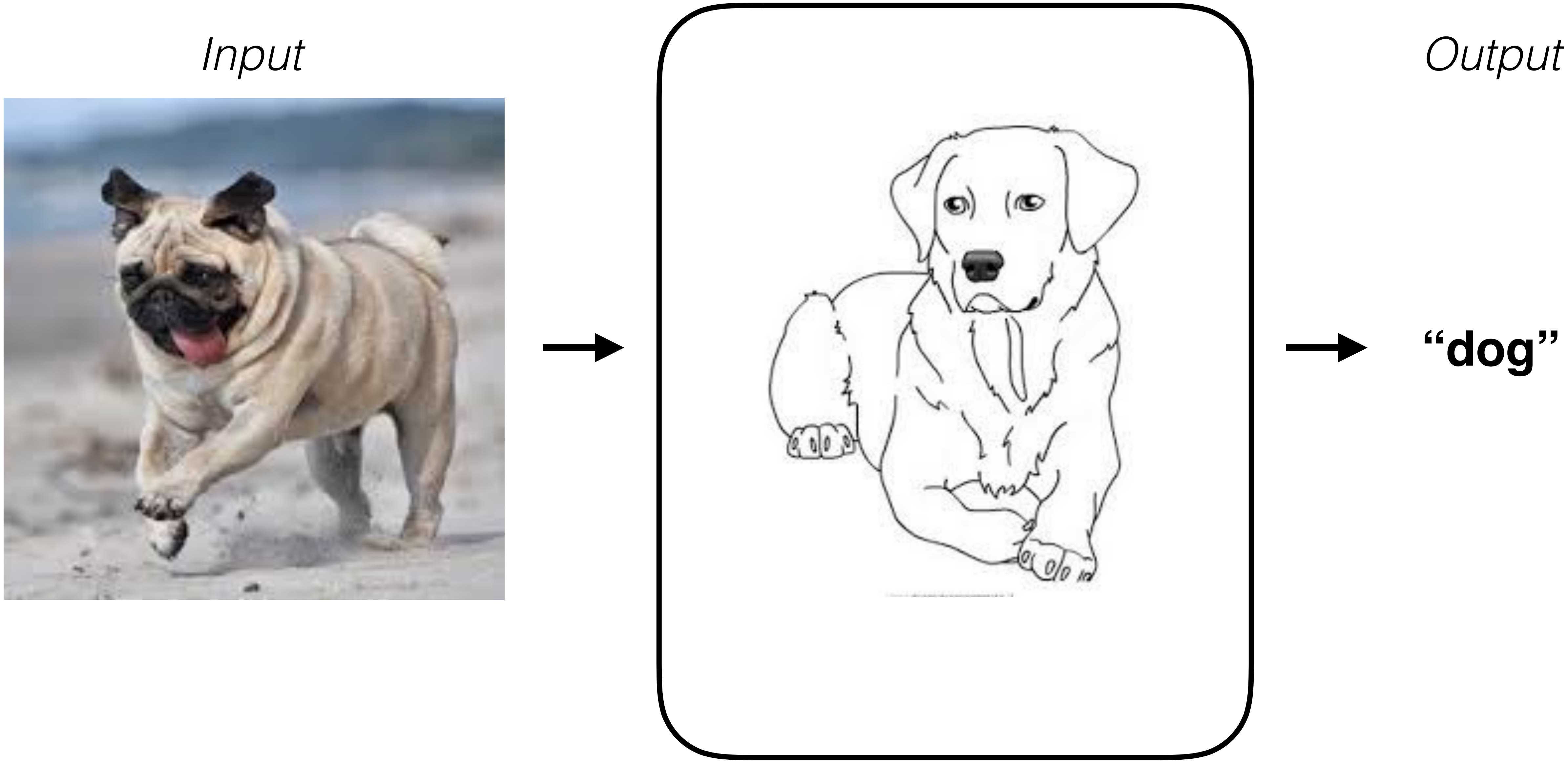
→

→

Output

“dog”

Teaching machines to see

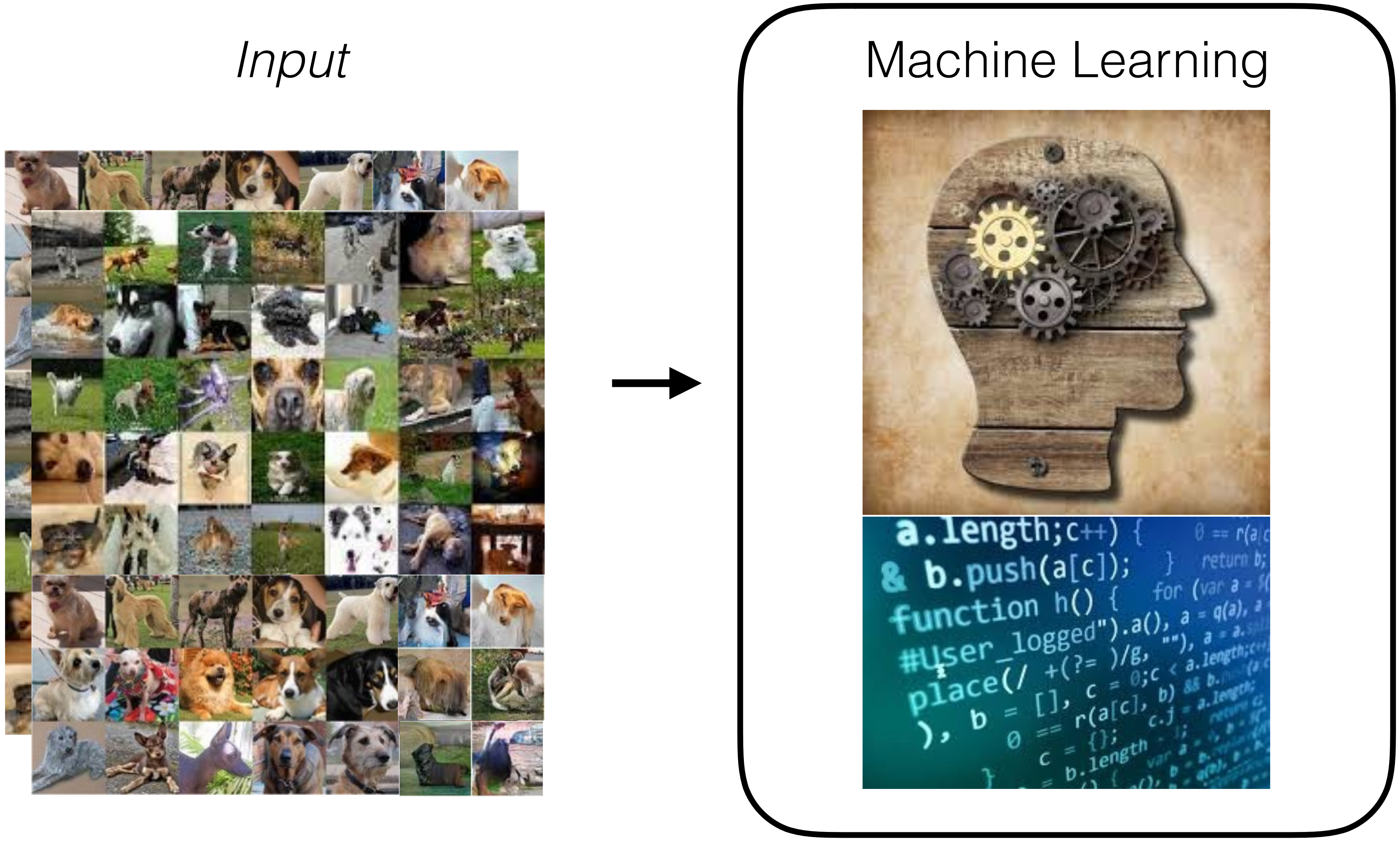


Input

Output

“dog”

Teaching machines to see



Input → Machine Learning → *Output*

“car”

“dog”

“cat”

“person”

“tree”

“chair”

“Bike”

“sky”

“computer”

…

Teaching machines to see

Input

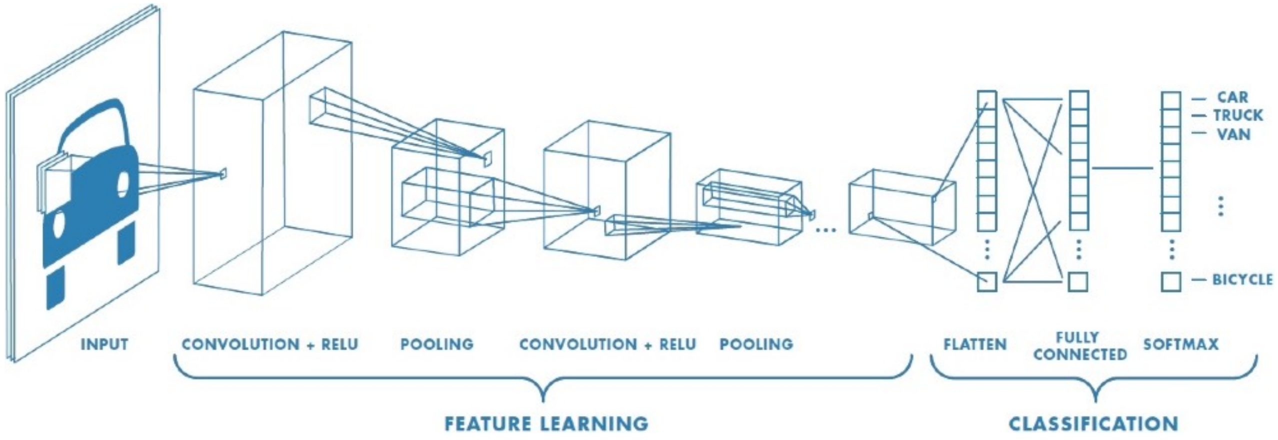
...

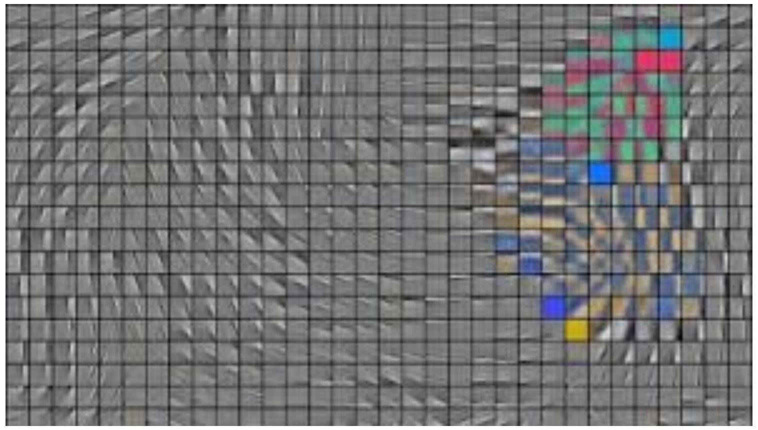
15M images, 22K classes

NVIDIA

Deep Learning

Convolutional Neural Network





Output

"car"

"dog"

"cat"

"person"

"tree"

"chair"

"Bike"

"sky"

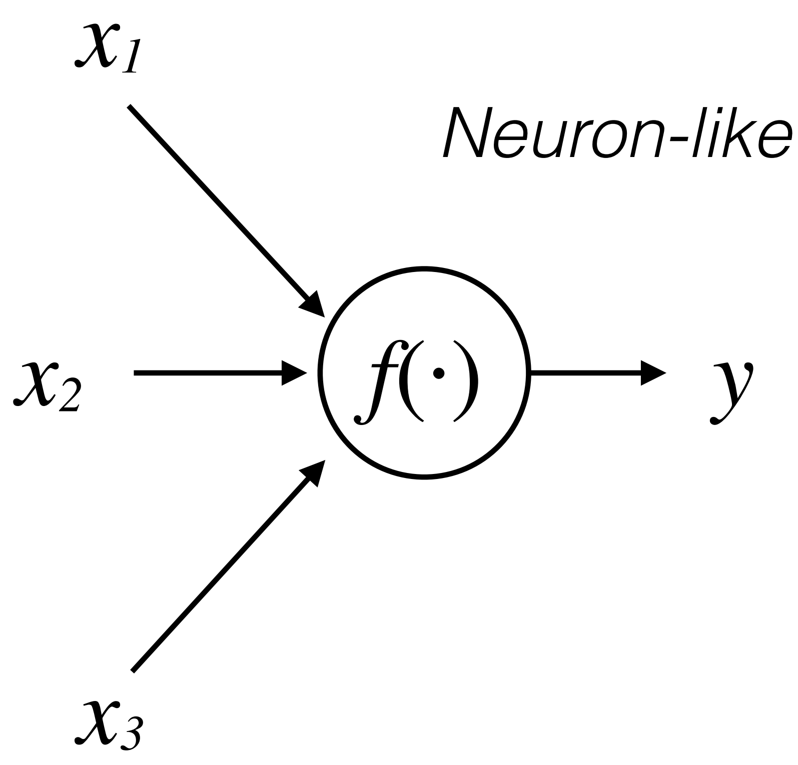
"computer"

…

Teaching machines to see

Deep Learning

Convolutional Neural Network

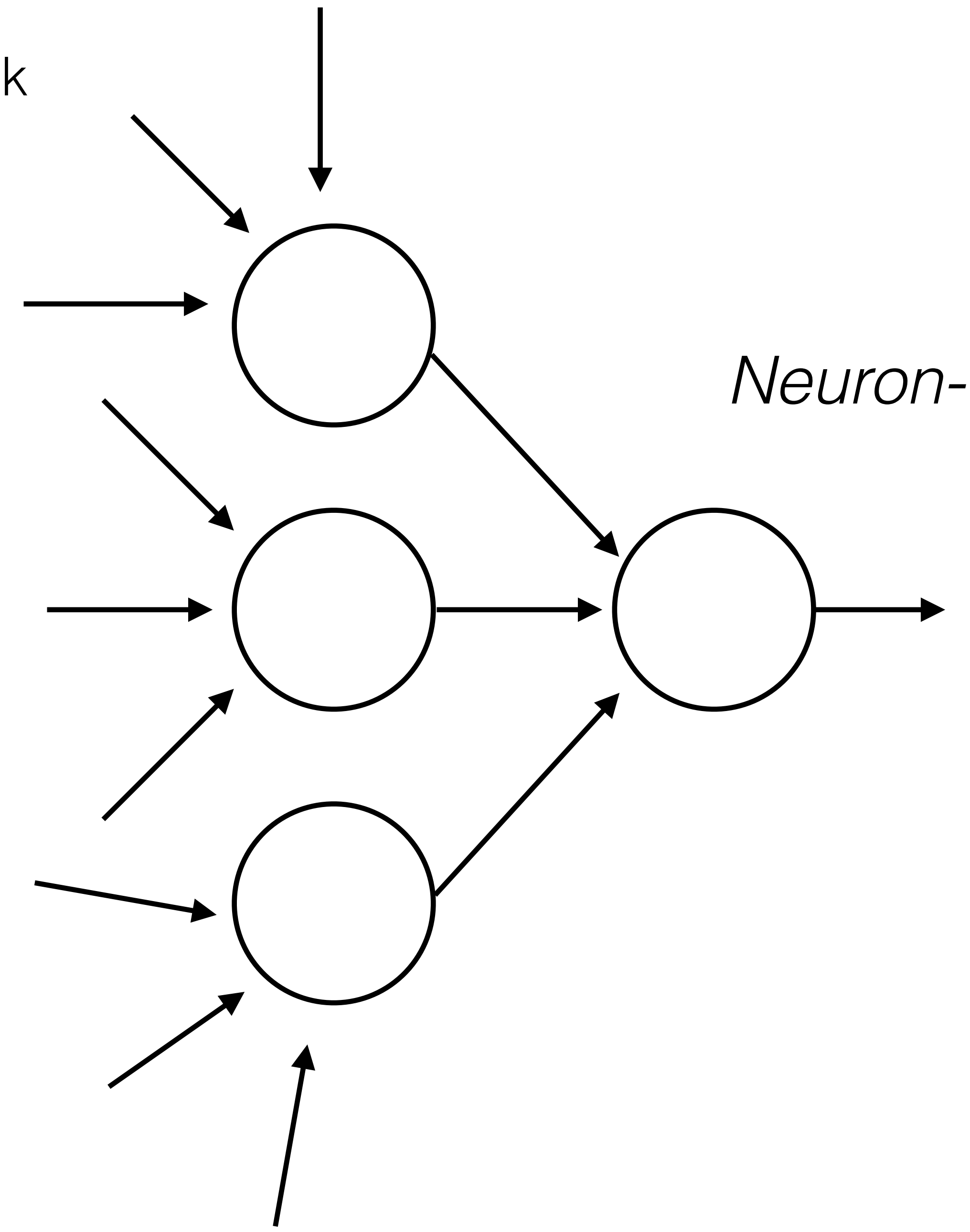


Neuron-like unit

Teaching machines to see

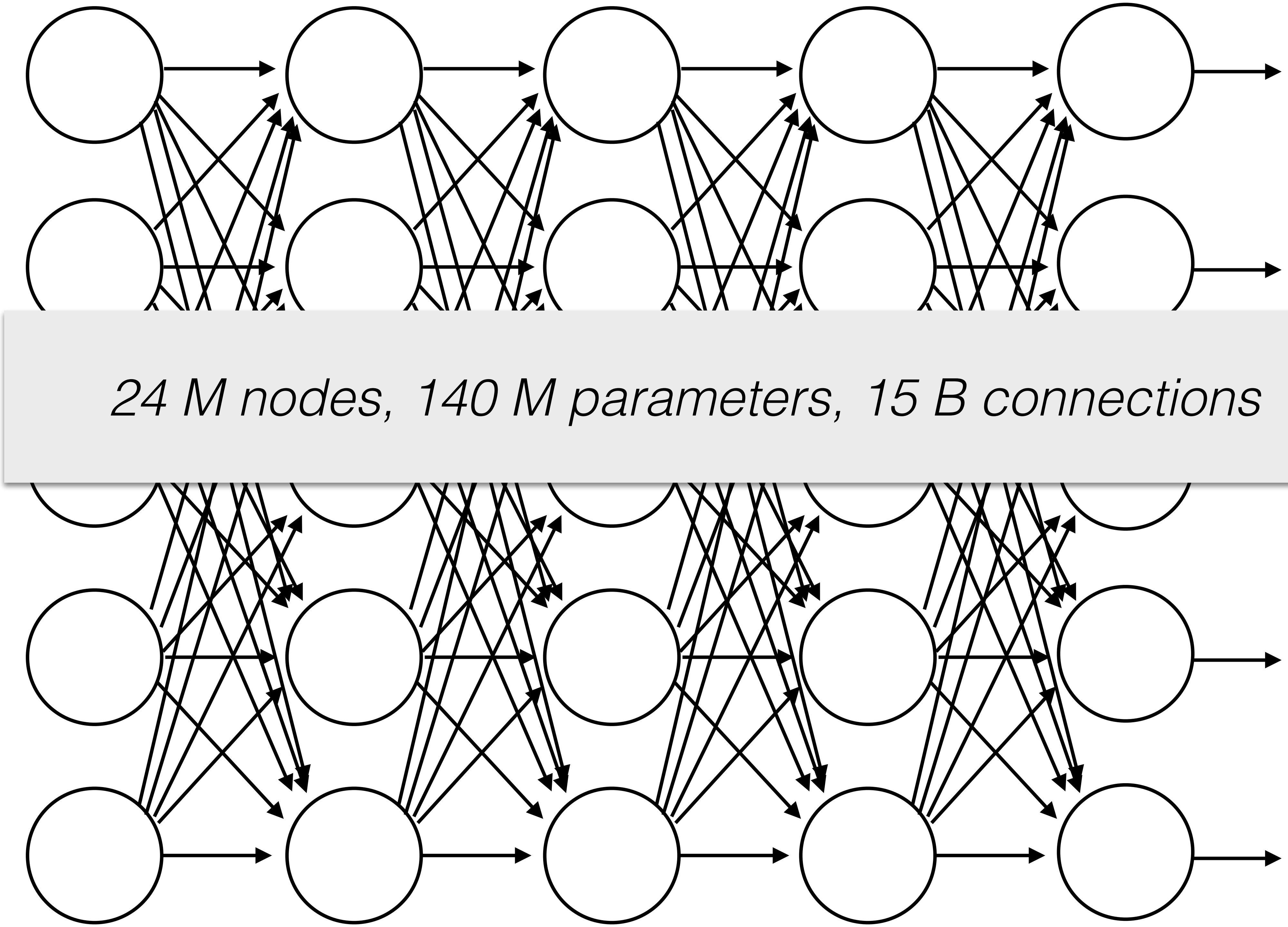
Deep Learning

Convolutional Neural Network



Neuron-like unit

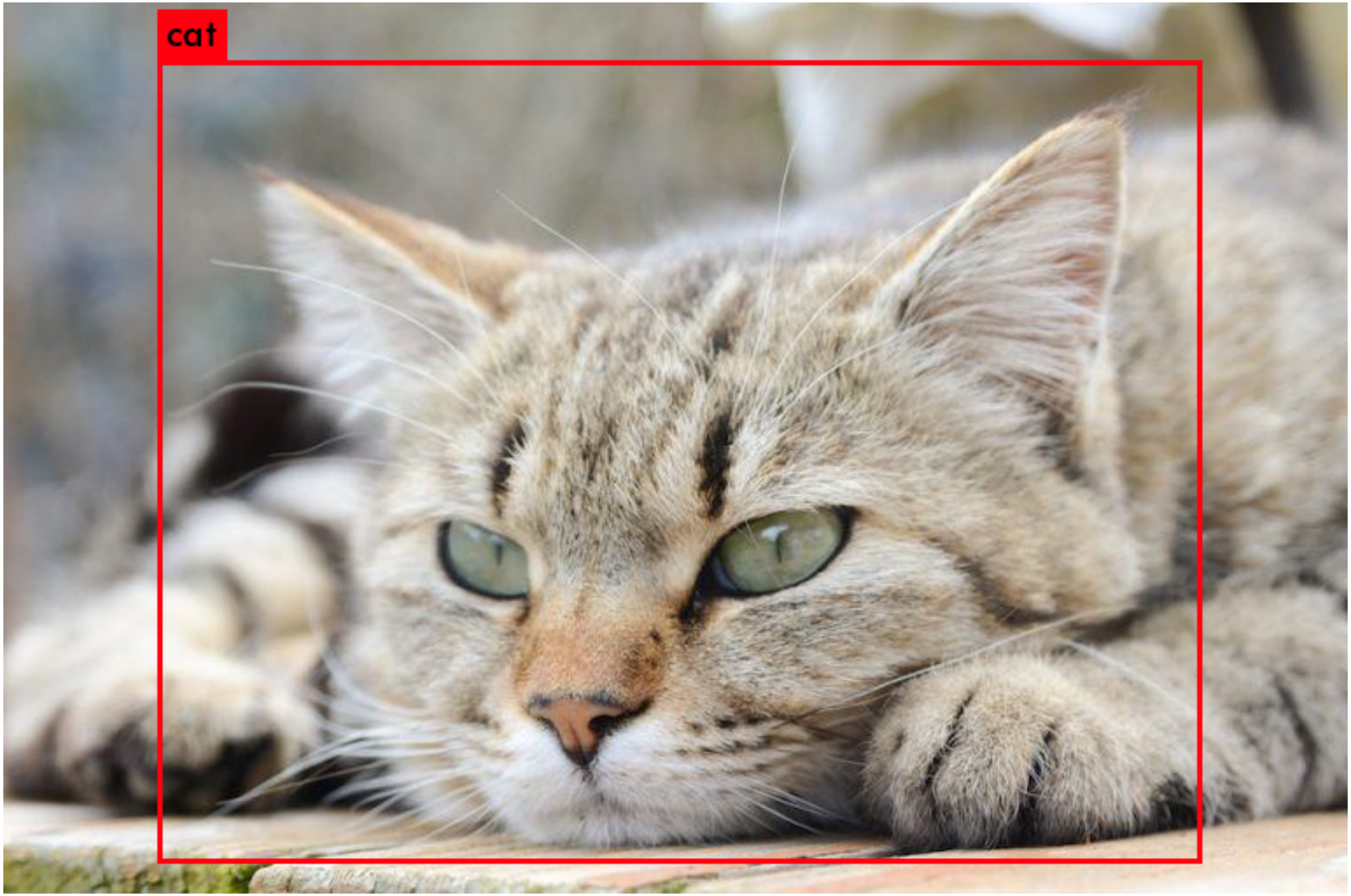
Teaching machines to see



24 M nodes, 140 M parameters, 15 B connections

Lamberto Ballan: *"Teaching machines to see: a quest to visual intelligence"*

Teaching machines to see



Contact

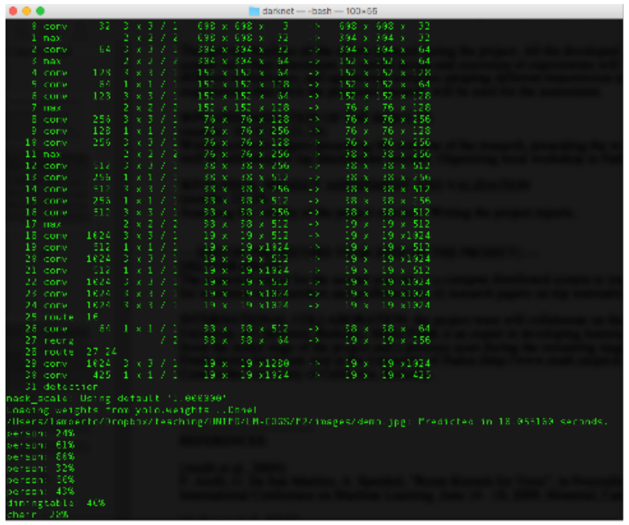
✉ lamberto.ballan@unipd.it

⌂ www.math.unipd.it/~ballan

@ twitter.com/lambertoballan

Thanks!

Live demo…



```
    0 conv     32  3 x 3 / 1   608 x 608 x   3   ->   608 x 608 x  32
    1 max          2 x 2 / 2   608 x 608 x  32   ->   304 x 304 x  32
    2 conv     64  3 x 3 / 1   304 x 304 x  32   ->   304 x 304 x  64
    3 max          2 x 2 / 2   304 x 304 x  64   ->   152 x 152 x  64
    4 conv    128  3 x 3 / 1   152 x 152 x  64   ->   152 x 152 x 128
    5 conv     64  1 x 1 / 1   152 x 152 x 128   ->   152 x 152 x  64
    6 conv    128  3 x 3 / 1   152 x 152 x  64   ->   152 x 152 x 128
    7 max          2 x 2 / 2   152 x 152 x 128   ->    76 x  76 x 128
    8 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256
    9 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128
   10 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256
   11 max          2 x 2 / 2    76 x  76 x 256   ->    38 x  38 x 256
   12 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512
   13 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256
   14 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512
   15 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256
   16 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512
   17 max          2 x 2 / 2    38 x  38 x 512   ->    19 x  19 x 512
   18 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024
   19 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512
   20 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024
   21 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512
   22 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024
   23 conv   1024  3 x 3 / 1    19 x  19 x1024   ->    19 x  19 x1024
   24 conv   1024  3 x 3 / 1    19 x  19 x1024   ->    19 x  19 x1024
   25 route  16
   26 conv     64  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x  64
   27 reorg              / 2    38 x  38 x  64   ->    19 x  19 x 256
   28 route  27 24
   29 conv   1024  3 x 3 / 1    19 x  19 x1280   ->    19 x  19 x1024
   30 conv    425  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 425
   31 detection
mask_scale: Using default '1.000000'
Loading weights from yolo.weights...Done!
/Users/lamberto/Dropbox/teaching/UNIPD/LM-DSSS/F2/images/demo.jpg: Predicted in 18.055150 seconds.
person: 24%
person: 61%
person: 88%
person: 32%
person: 56%
person: 43%
diningtable: 40%
chair: 22%
```