Lecture 1: Introduction

COS 429: Computer Vision

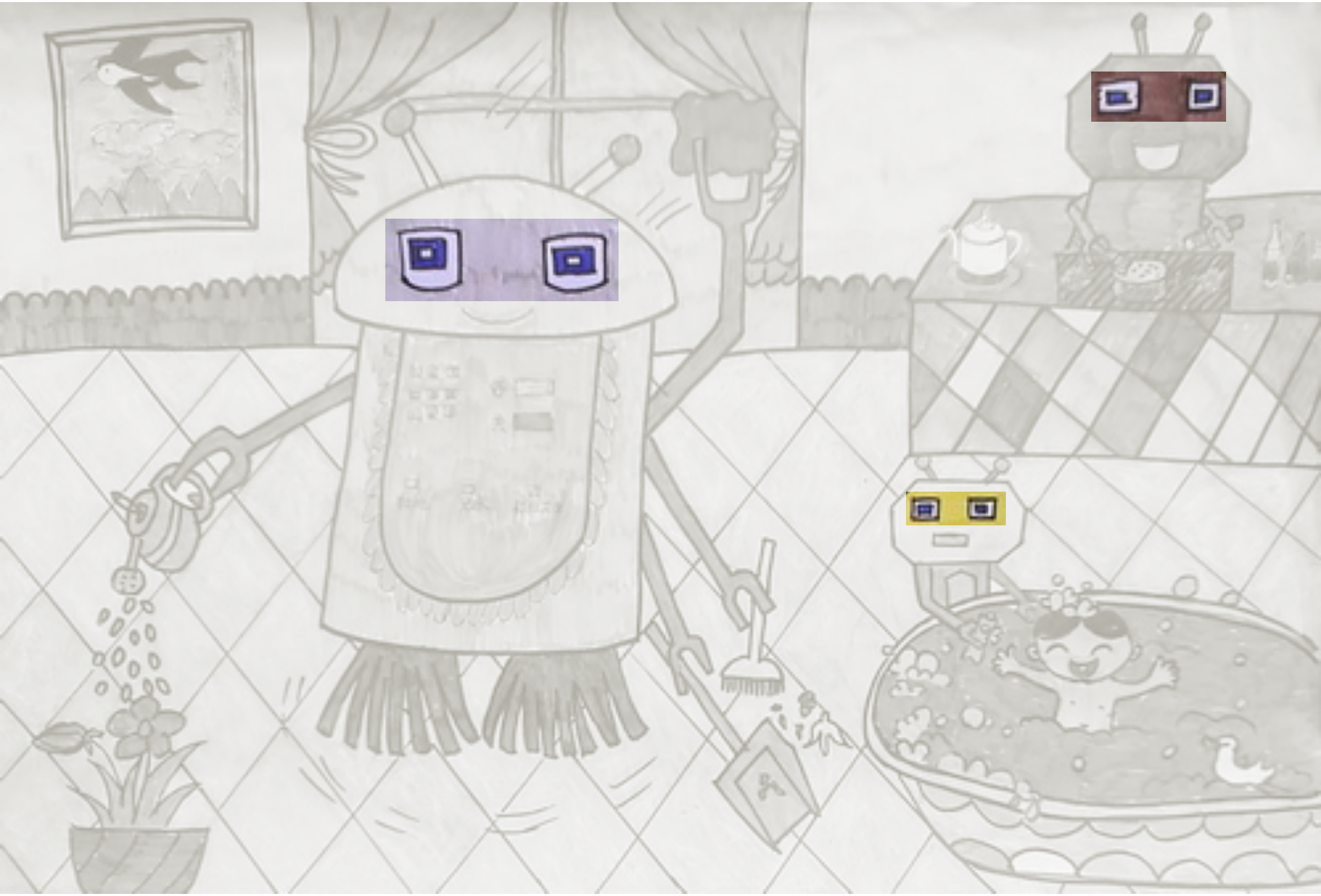
http://cs.princeton.edu/~cos429/

Slide credits: Fei-Fei Li, Deva Ramanan, Szymon Rusinkiewicz

What does the future world look like?

Slide credit: Fei-Fei Li

Y.Z. Dai, 7 year old



Slide credit: Fei-Fei Li

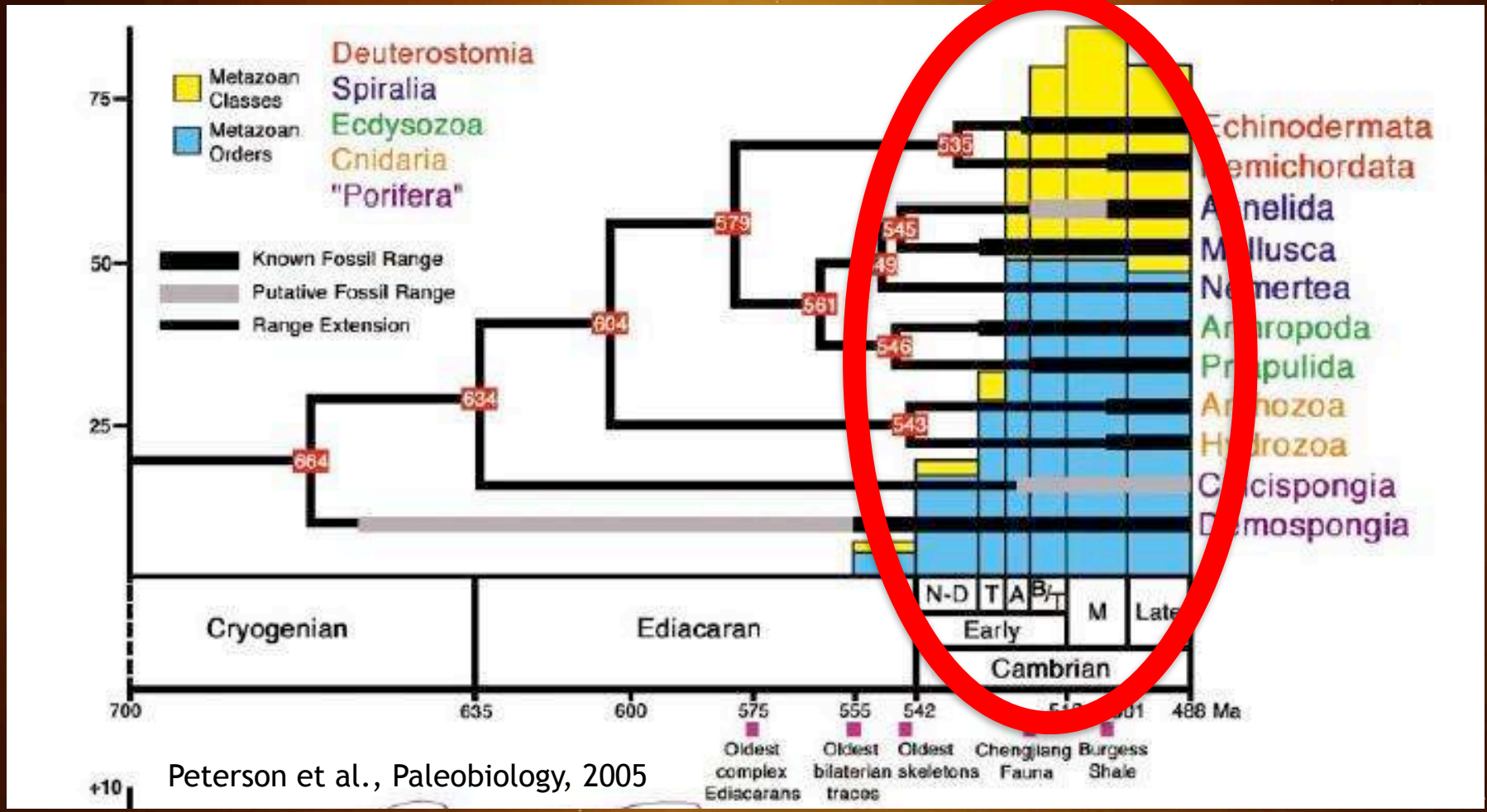
Y.Z. Dai, 7 year old

Slide credit: Fei-Fei Li

Y.Z. Dai, 7 year old

Slide credit: Fei-Fei Li

Slide credit: Fei-Fei Li

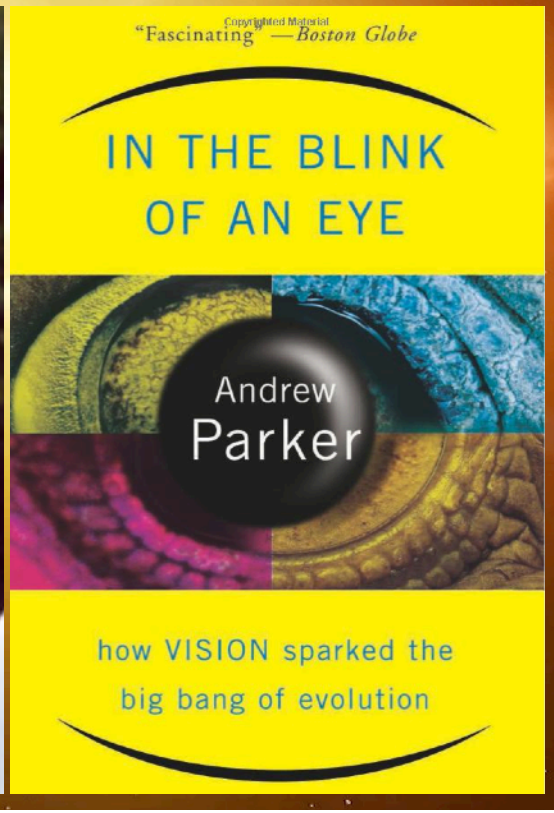


Peterson et al., Paleobiology, 2005

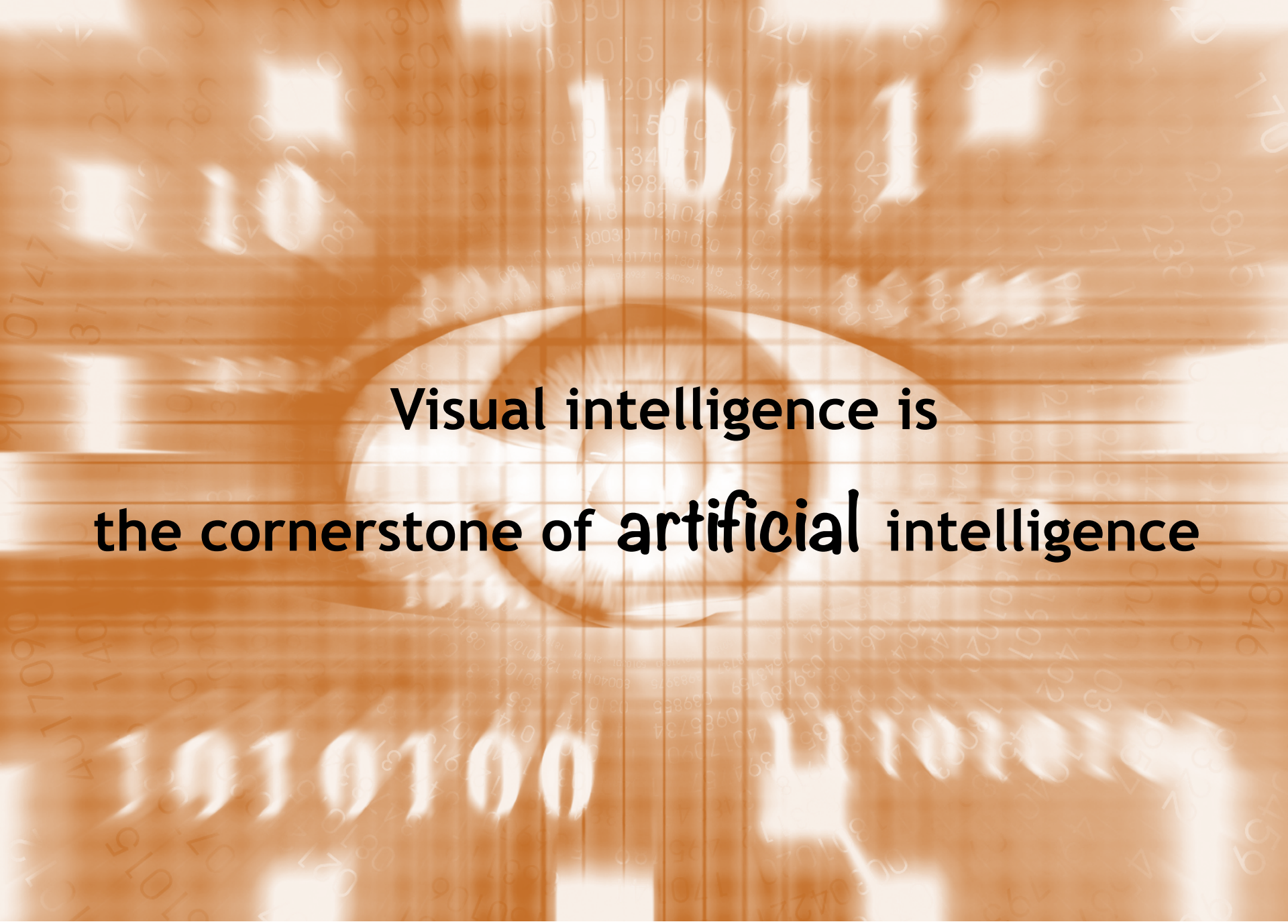
Slide credit: Fei-Fei Li

"The Cambrian Explosion is triggered by the sudden evolution of vision," which set off an evolutionary arms race where animals either evolved or died.

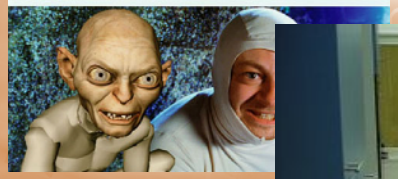
---- Andrew Parker, zoologist



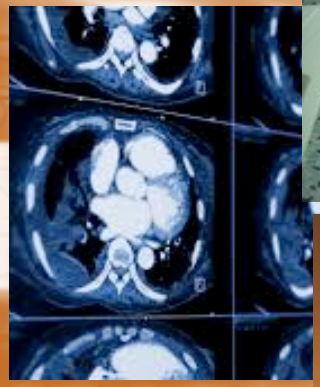
Slide credit: Fei-Fei Li



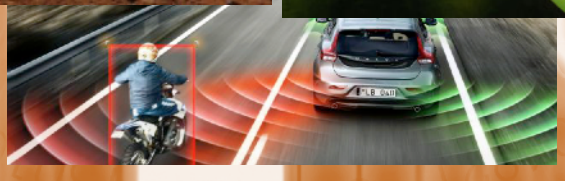
Slide credit: Fei-Fei Li







Slide credit: Fei-Fei Li

Seeing is challenging

Challenge: illumination

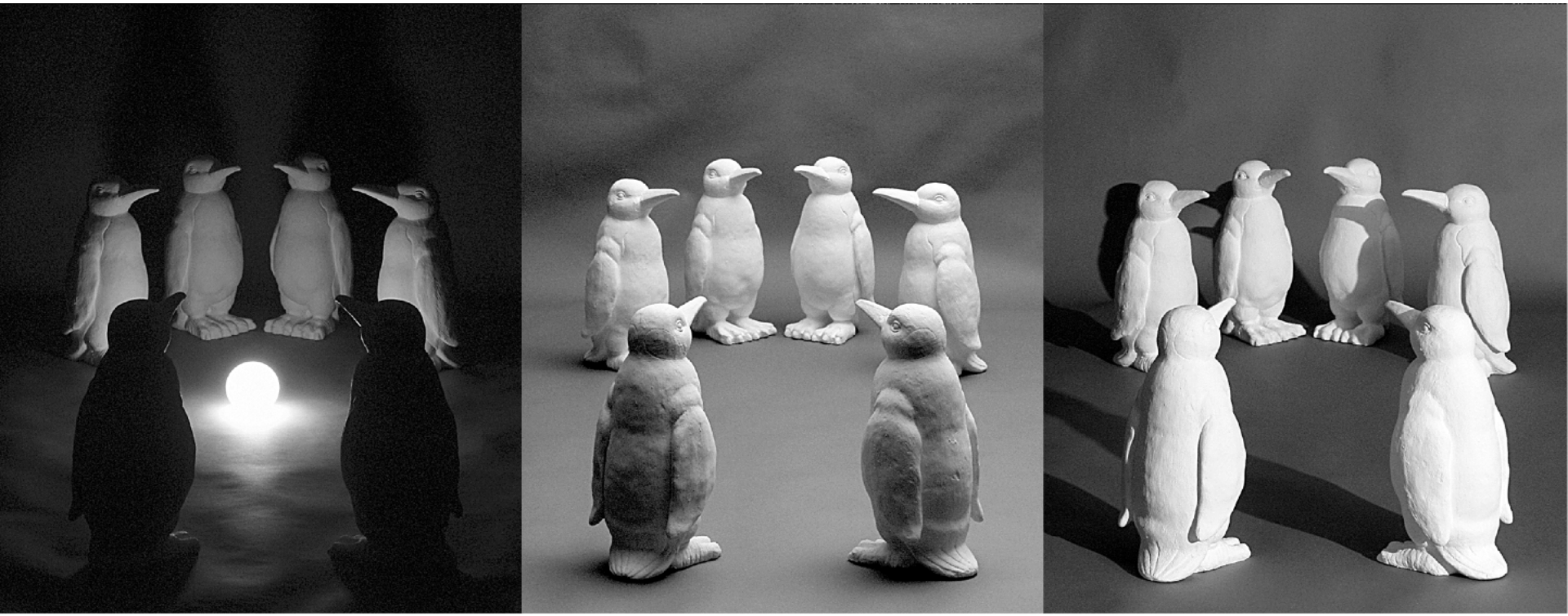
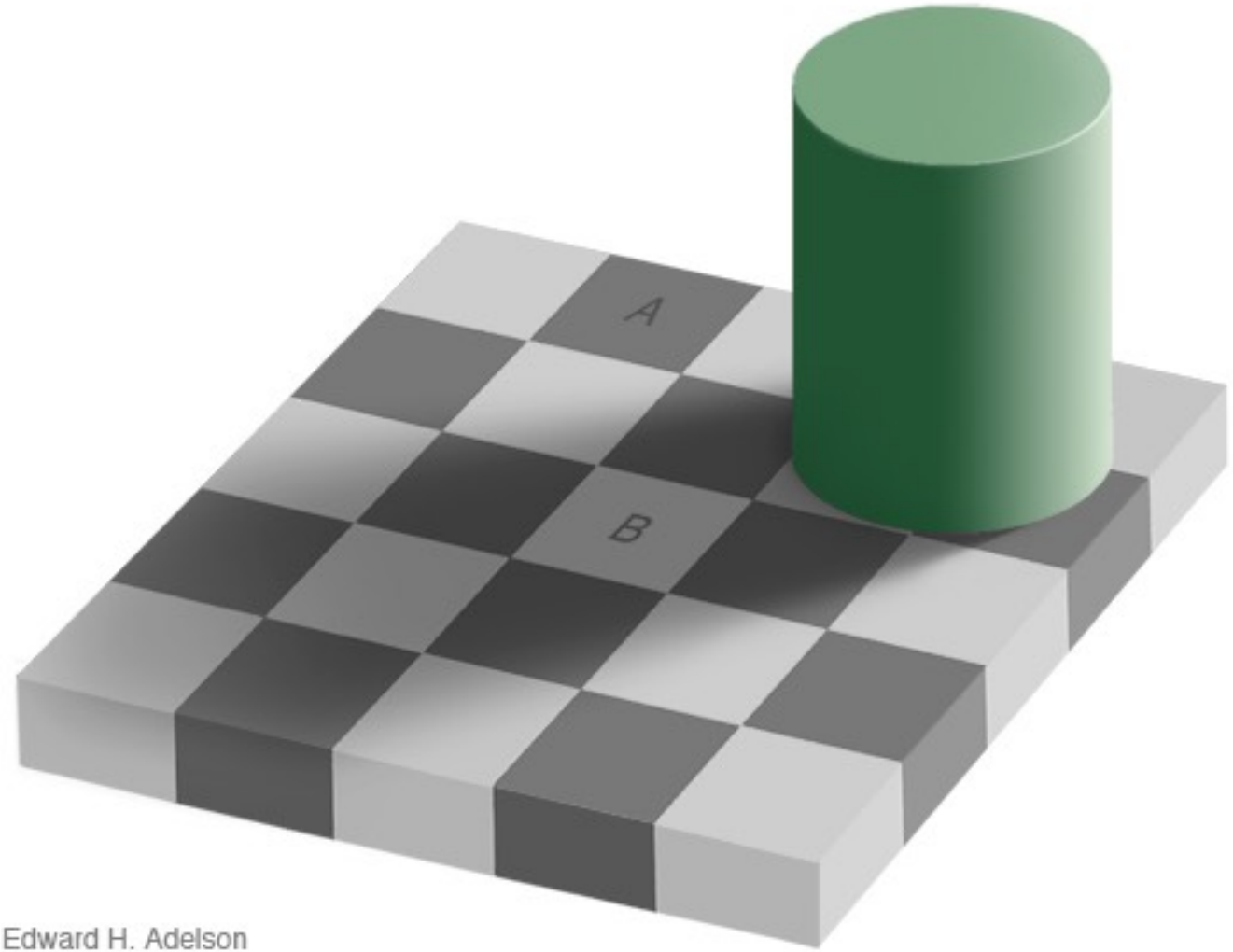


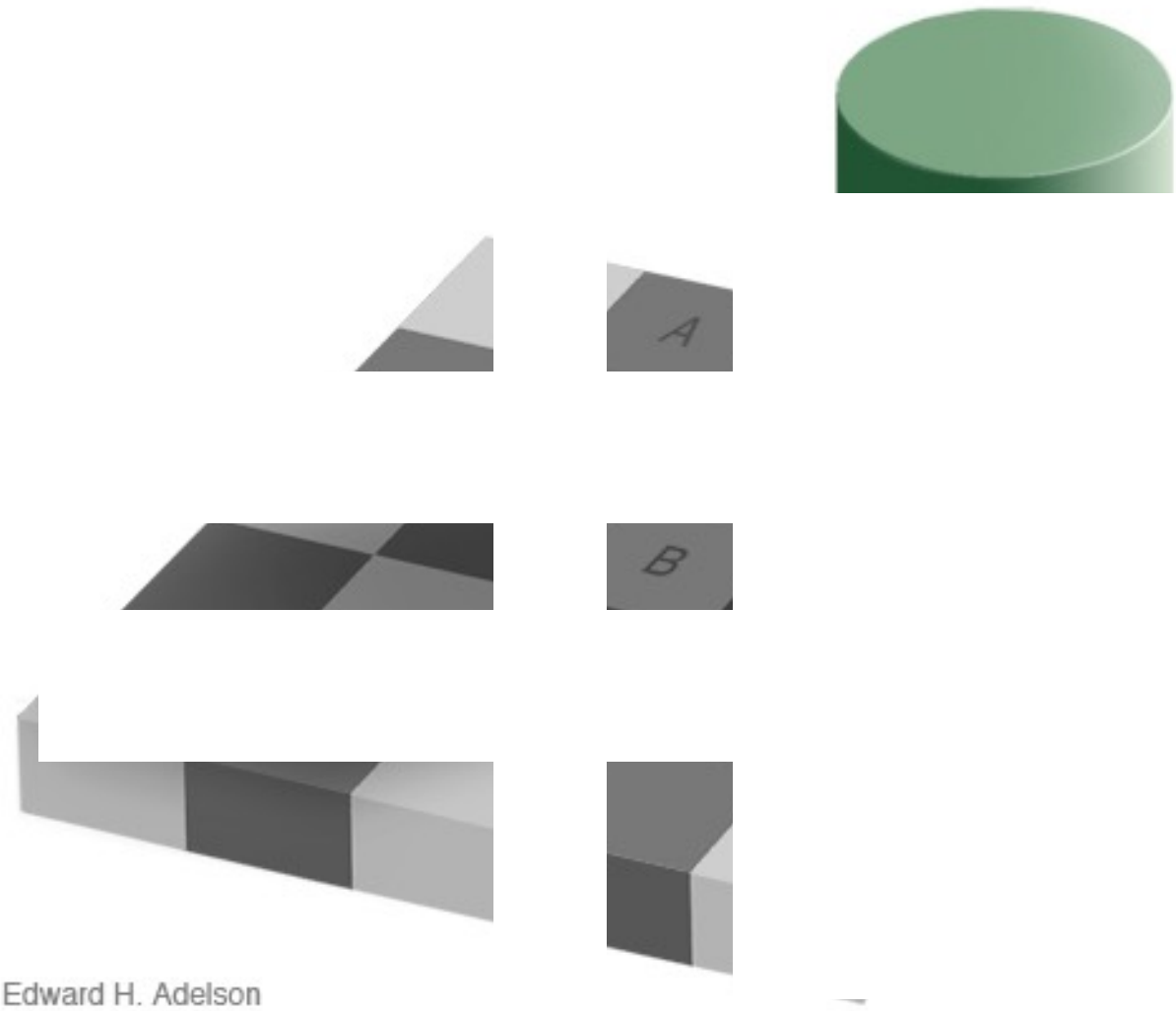
image credit: J. Koenderink

Challenge: light and shadow

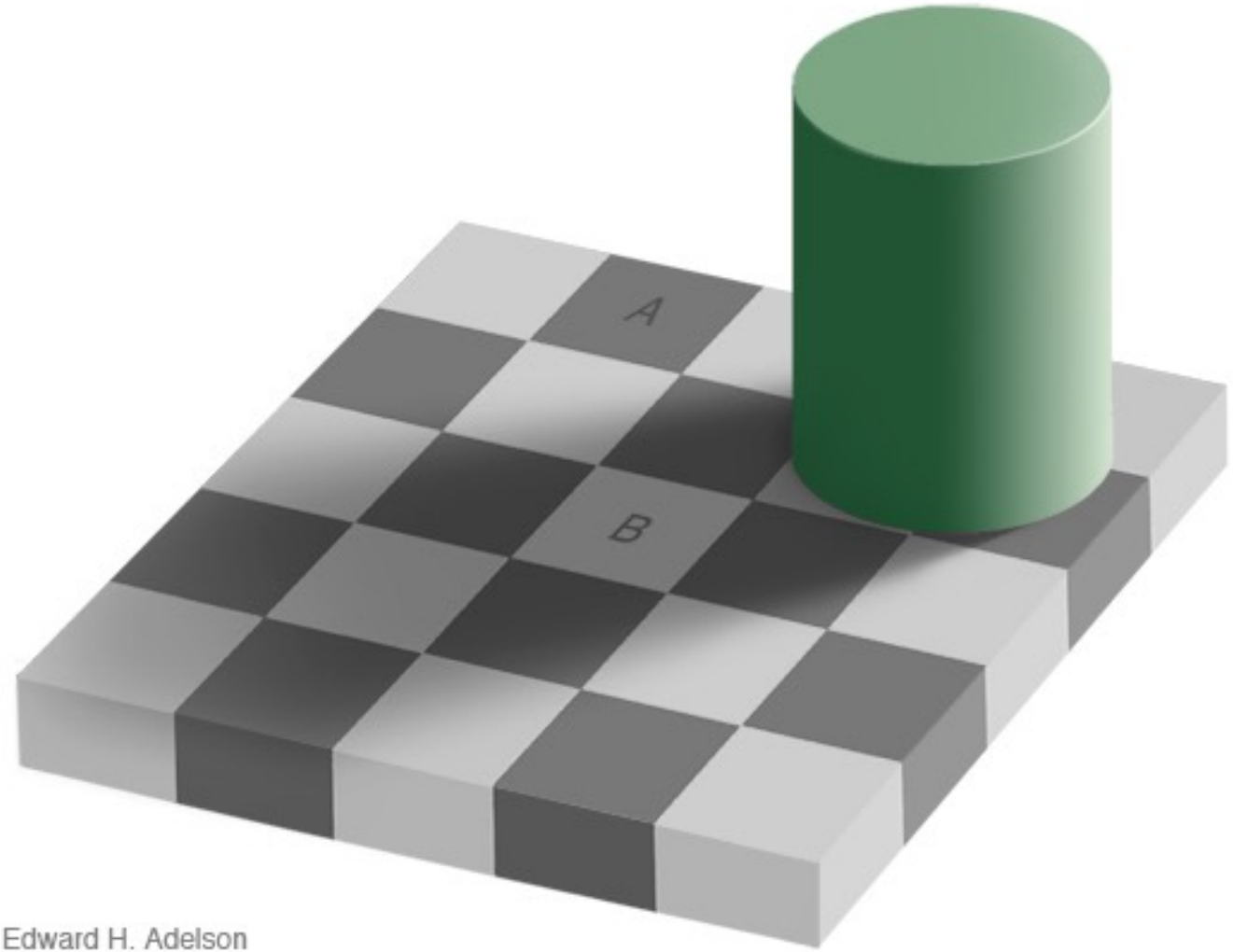


Edward H. Adelson

Challenge: light and shadow

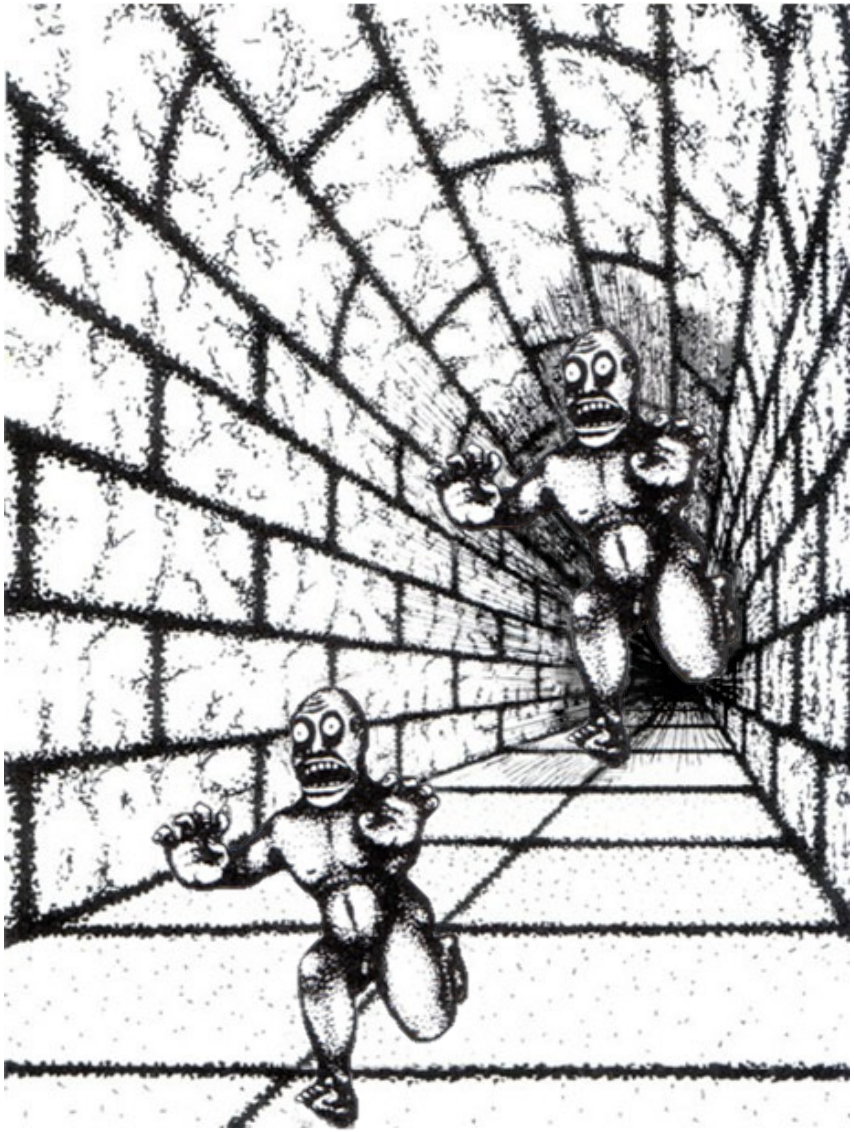


Challenge: light and shadow

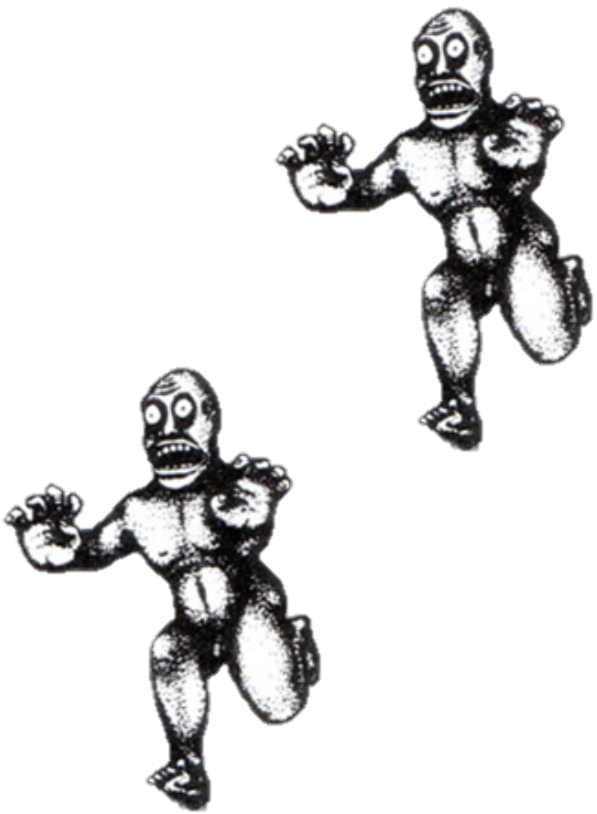


Edward H. Adelson

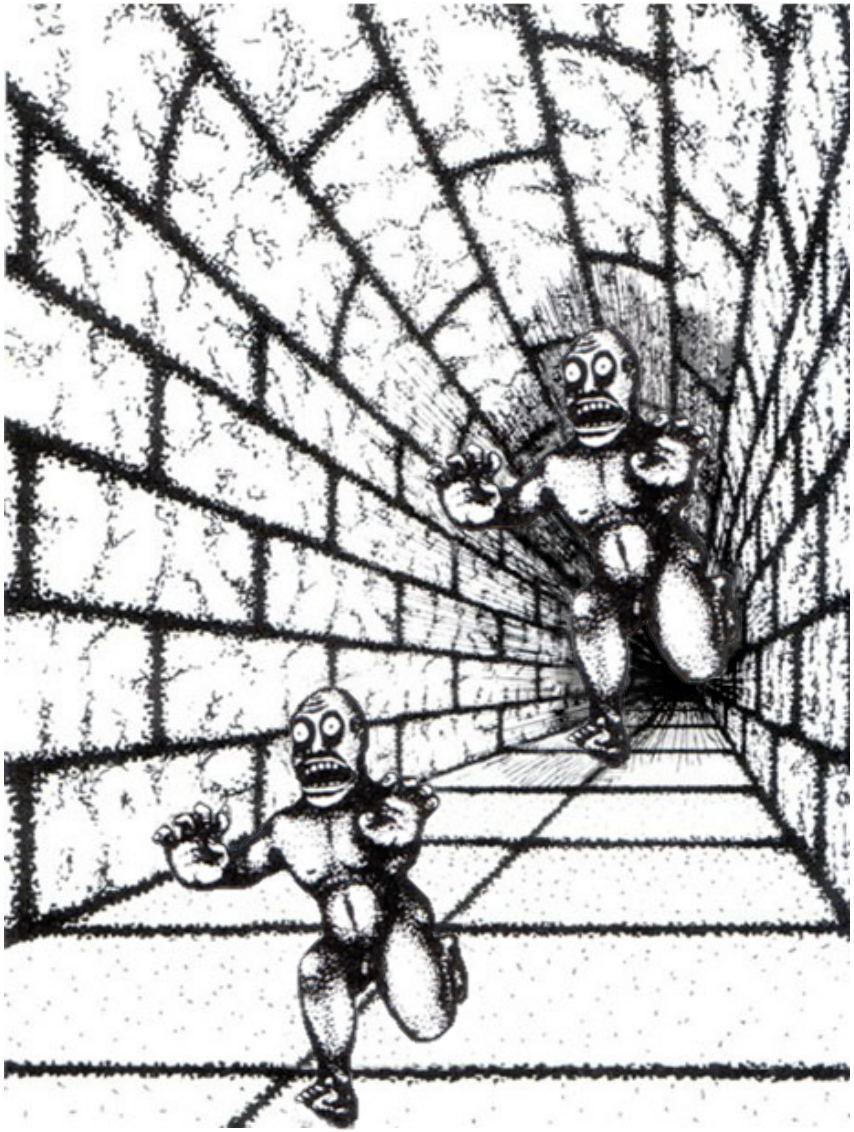
Challenge: scale



Challenge: scale



Challenge: scale



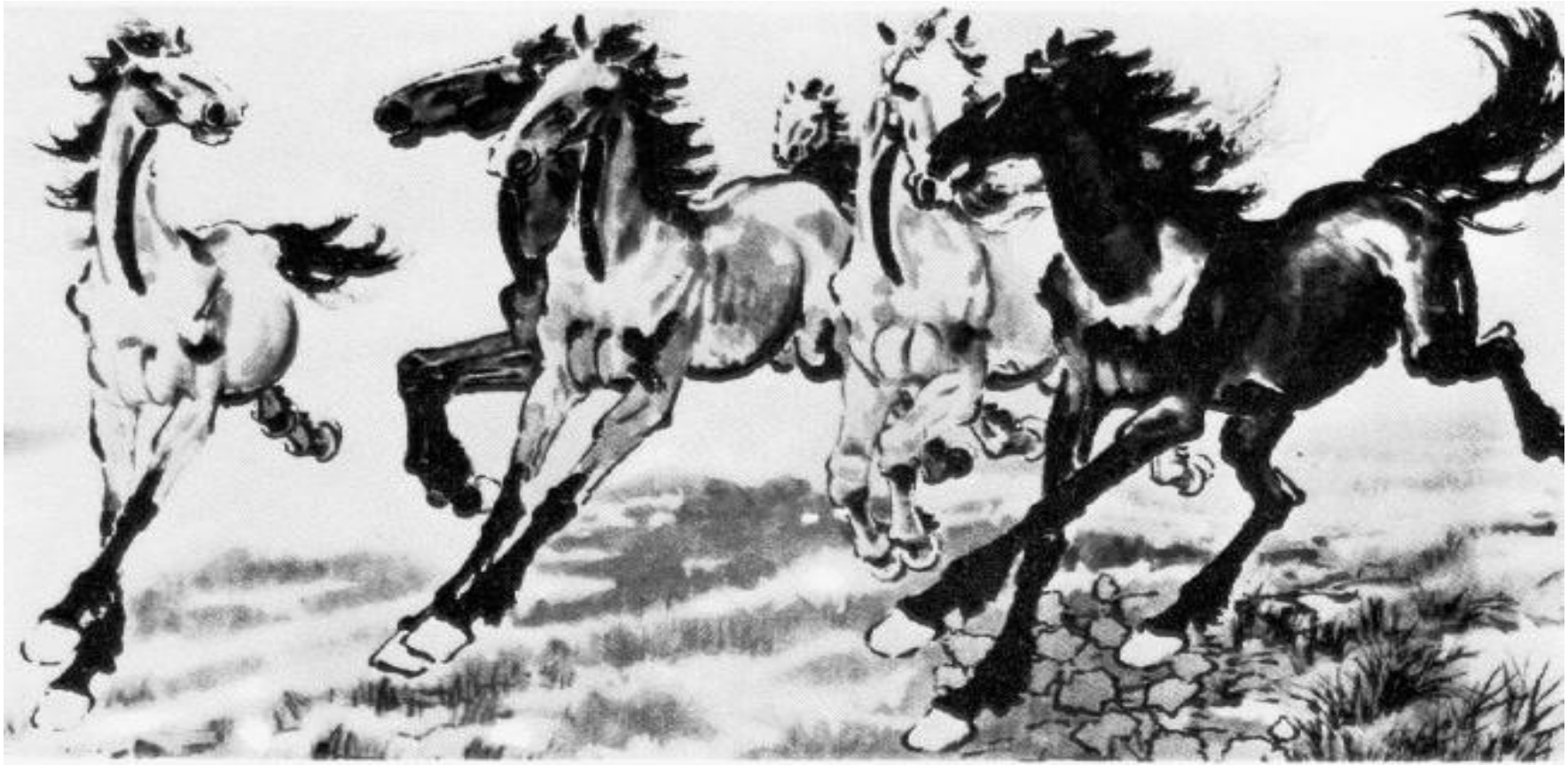
Challenge: occlusion and clutter

Image source: National Geographic

Challenge: intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

Challenge: deformation



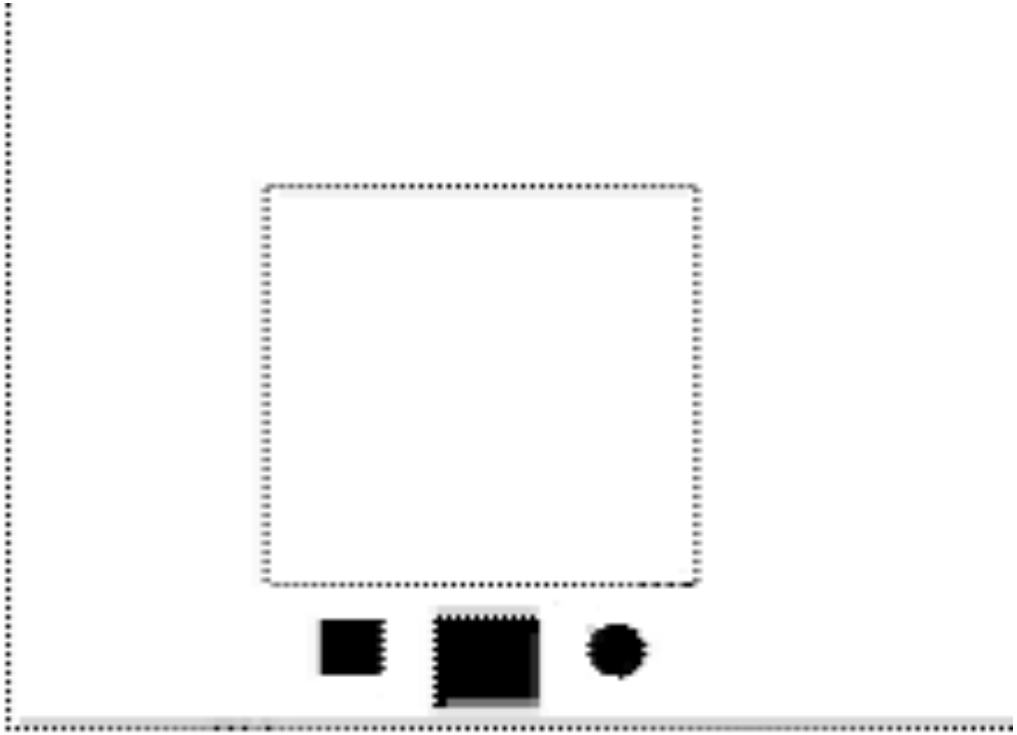
Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

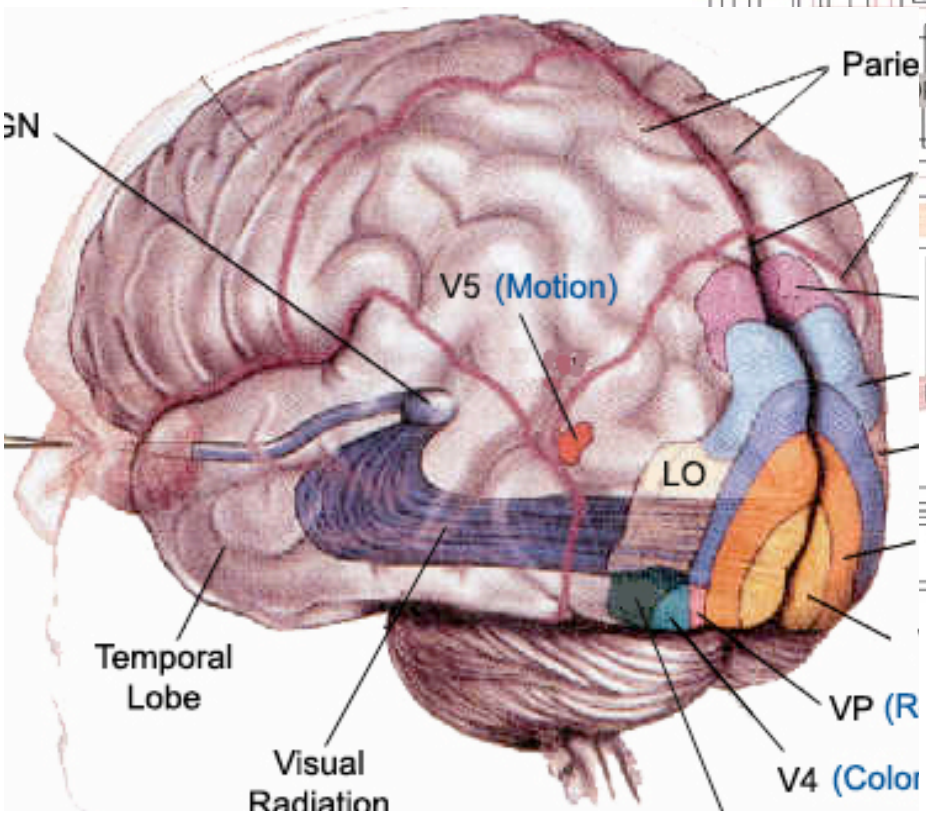
Challenge: motion

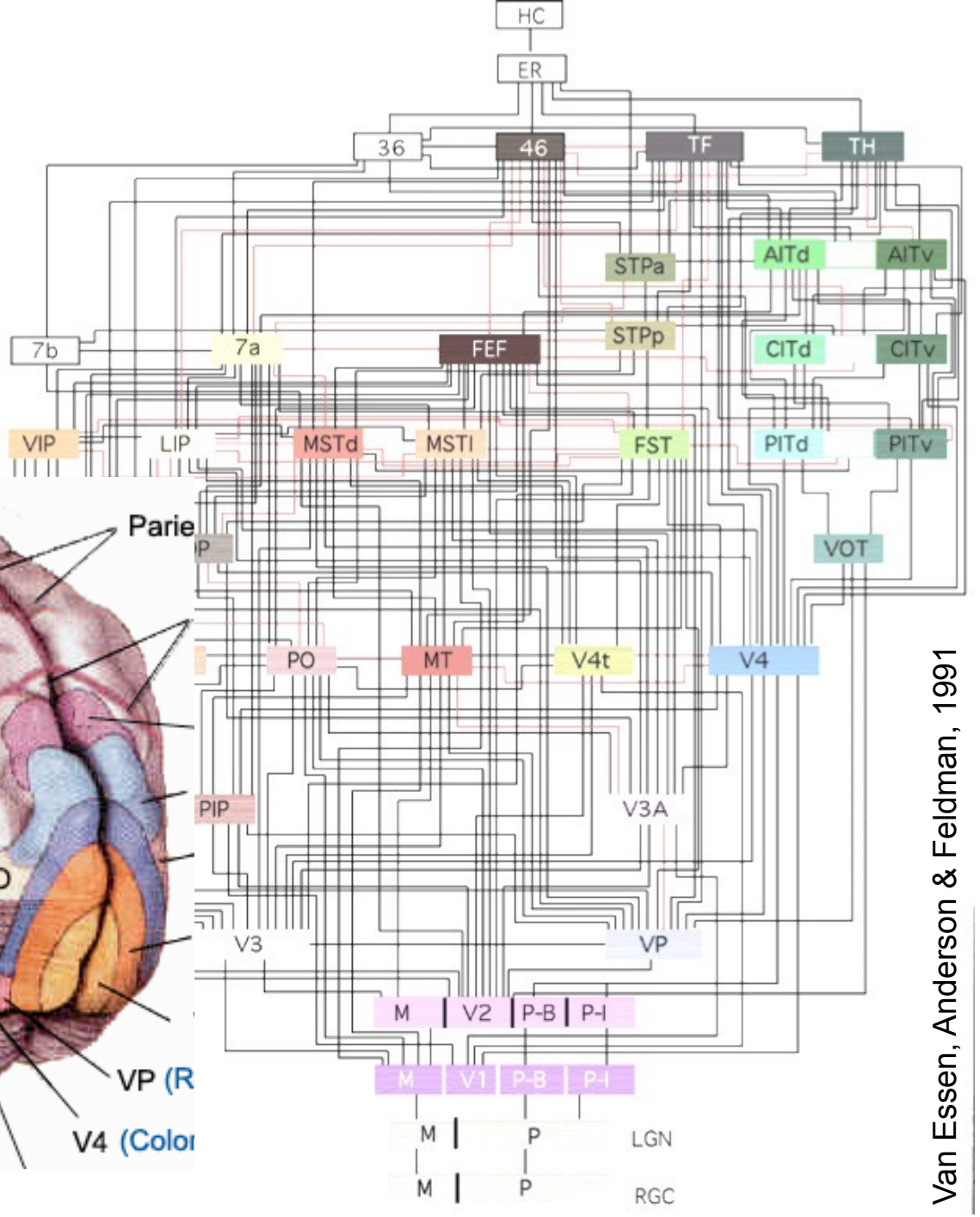
Challenge: perspective

(Aside) Challenge: intentions and goals



Slide credit: Deva Ramanan



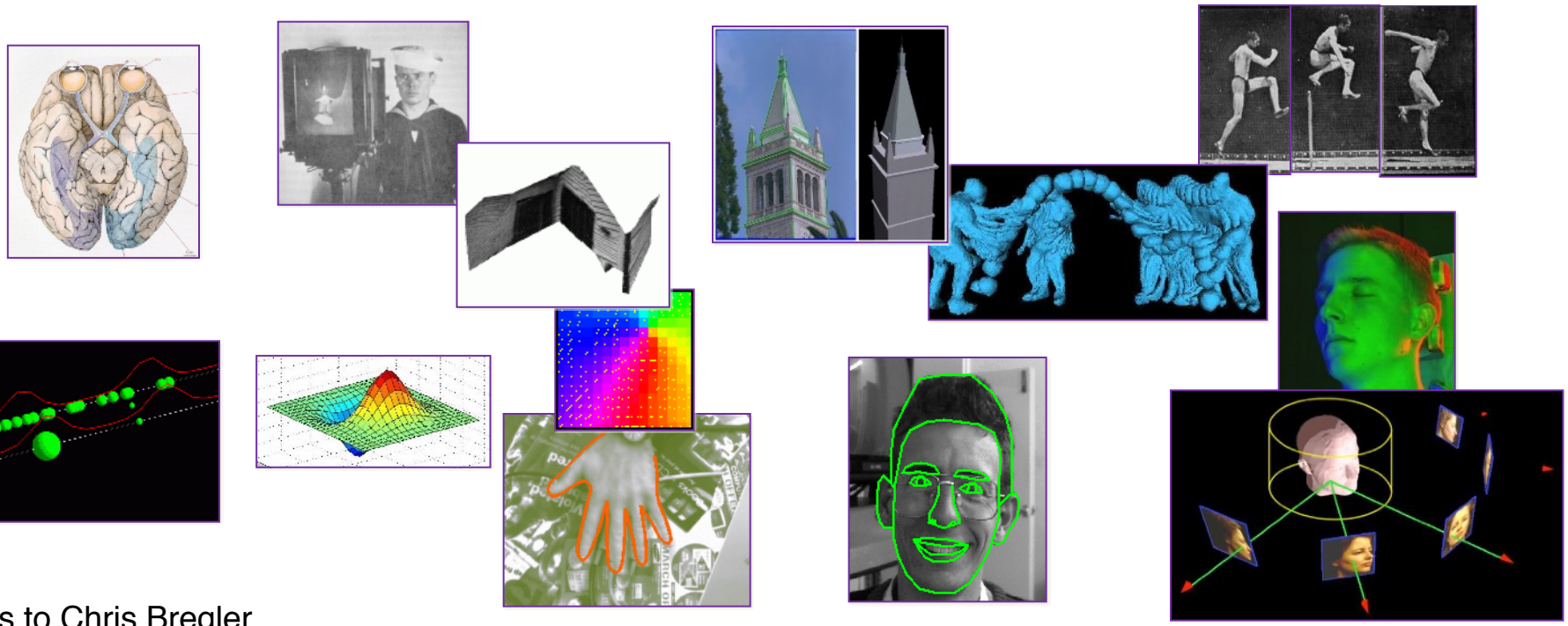


Van Essen, Anderson & Feldman, 1991

What is computer vision?

Computer vision

- Input: images or video
- Output: description of the world
  - Also: measuring, classifying, interpreting visual information



Thanks to Chris Bregler

One Pixel

- Amount of light recorded by a photoreceptor

"Is this the object's color?
Illumination? Noise?
I can't tell!"

Slide credit: Szymon Rusinkiewicz

Low-Level or “Early” Vision

- Local image/shape properties

“There’s an edge!”

Slide credit: Szymon Rusinkiewicz

Mid-Level Vision

- Grouping and segmentation

"There's an object and a background!"

Slide credit: Szymon Rusinkiewicz

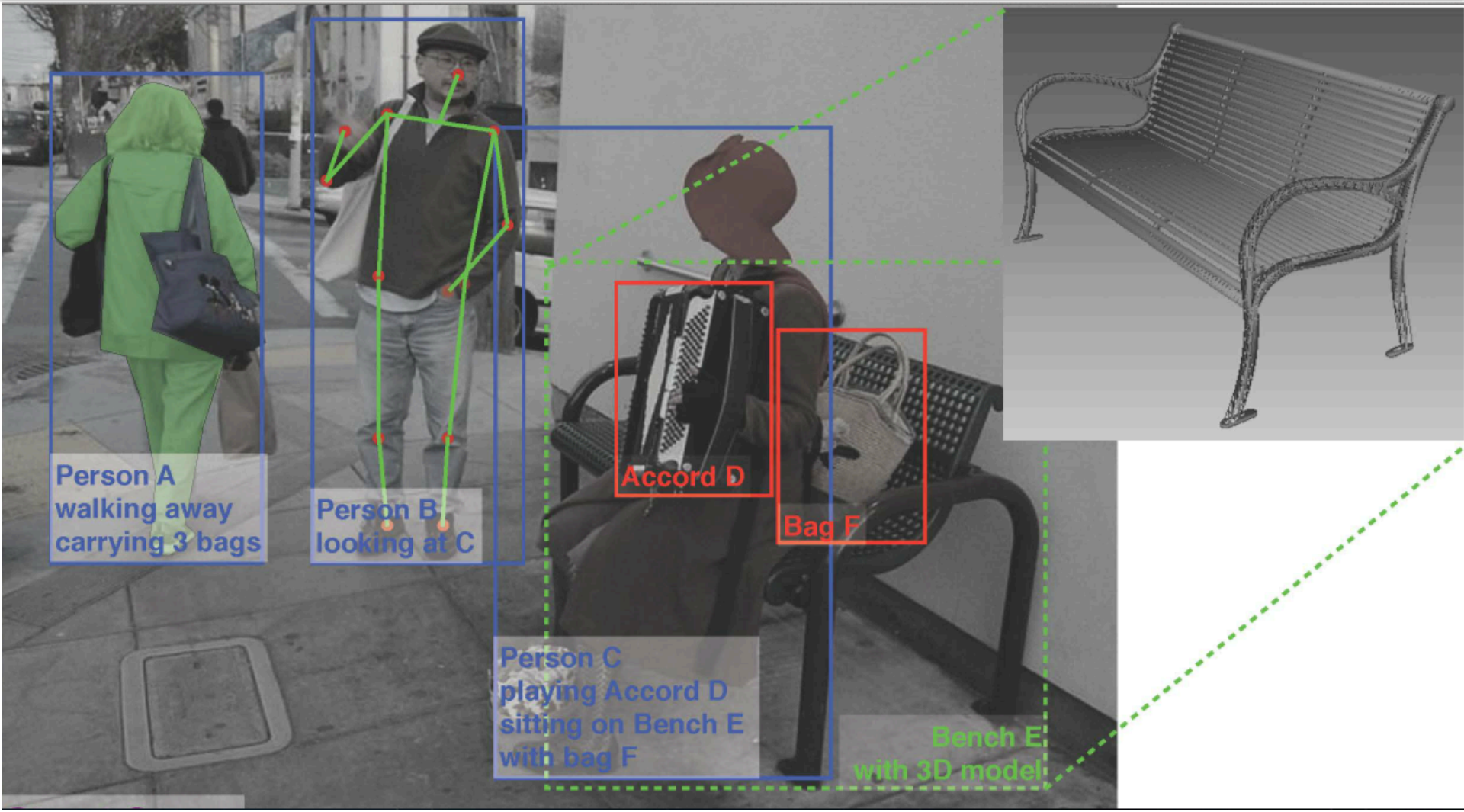
High-Level Vision

"It's a chair! It's in a room!"

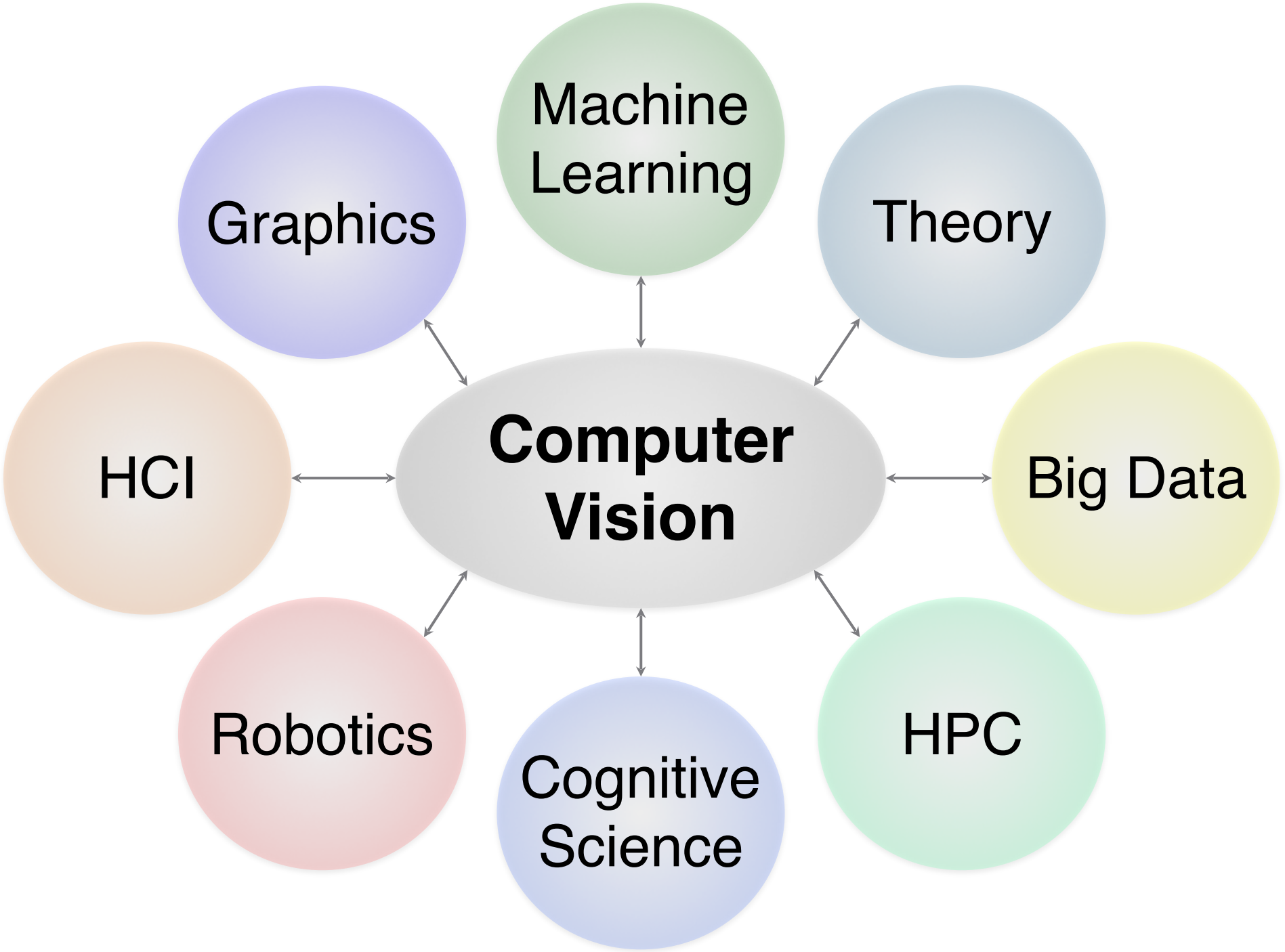
- Recognition
- Classification

Slide credit: Szymon Rusinkiewicz

What We Would Like to Infer…



Will person B put some money into Person C's tip bag?


Machine Learning
Graphics
Theory
HCI
Computer Vision
Big Data
Robotics
Cognitive Science
HPC

Computer vision success stories

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

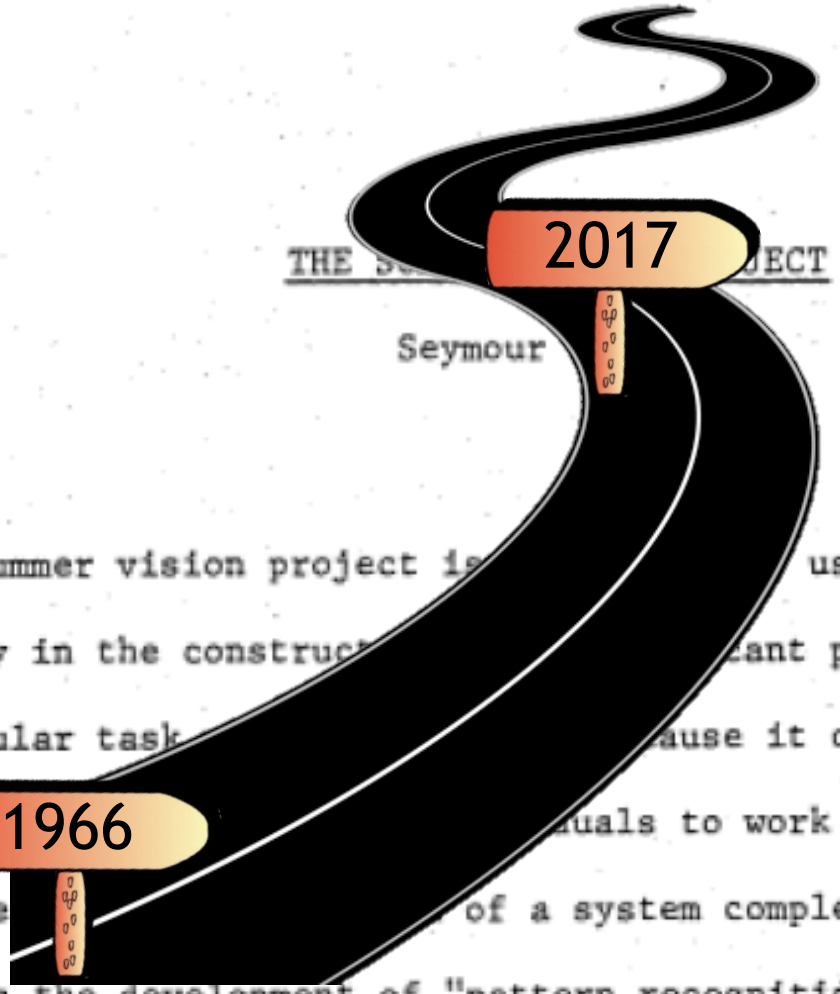
Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.



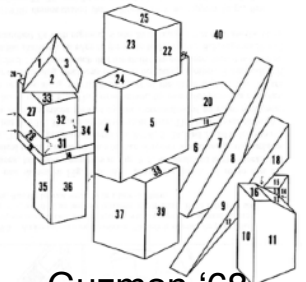
THE S JECT

Seymour

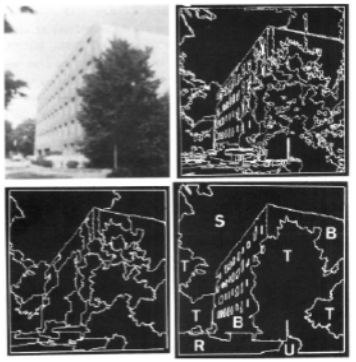
The summer vision project is use our summer workers effectively in the construct tant part of a visual system. The particular task ause it can be segmented into sub-prob uals to work independently and yet participate of a system complex enough to be a real landmark in the development of "pattern recognition".

A Brief History of Computer Vision

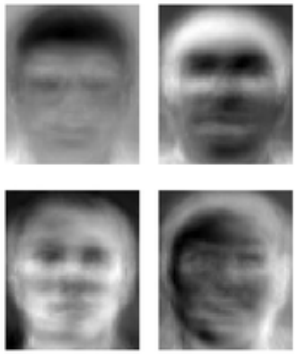
- 1966: Marvin Minsky assigns computer vision as an undergrad summer project
- 1960s: interpretation of synthetic worlds
- 1970s: interpretation of carefully selected images
- 1980s: NNs come and go; shift towards geometry and increased mathematical rigor
- 1990s: face recognition; statistical analysis
- 2000s: broader recognition; large annotated datasets available; video processing



Guzman ‘68

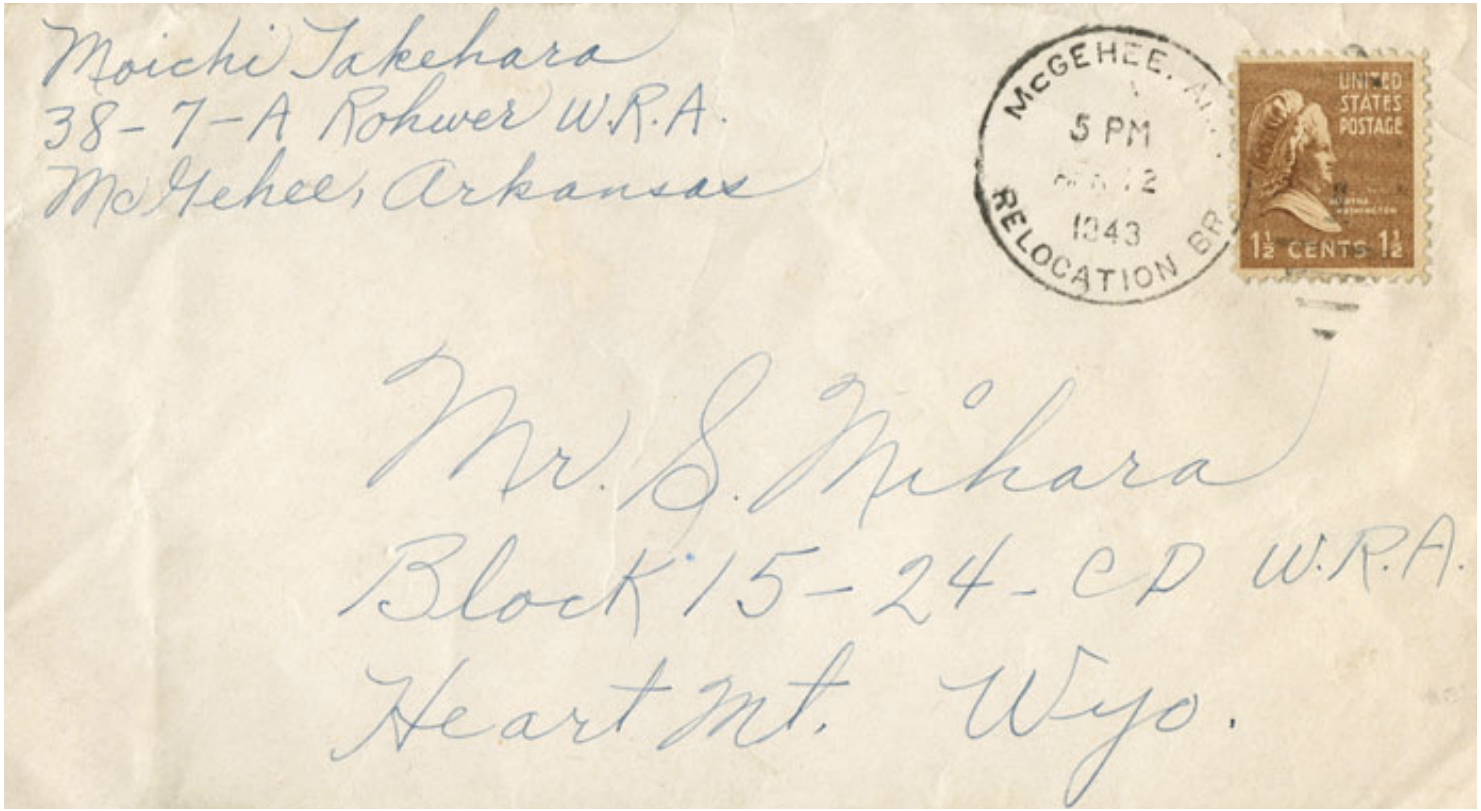


Ohta Kanade ‘78



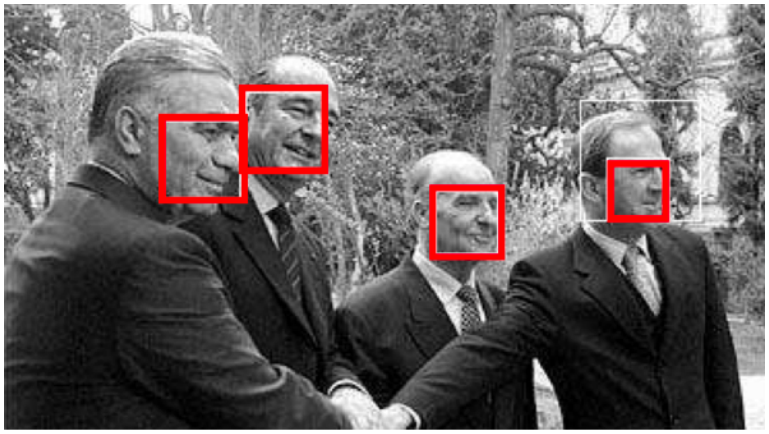
Turk and Pentland ‘91

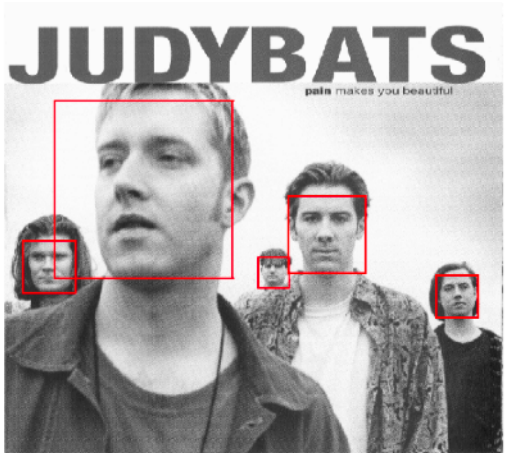
Sorting our mail

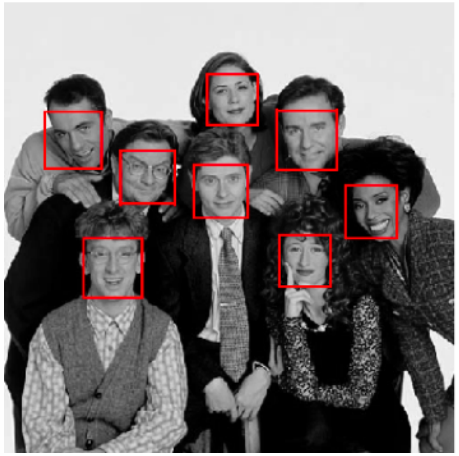


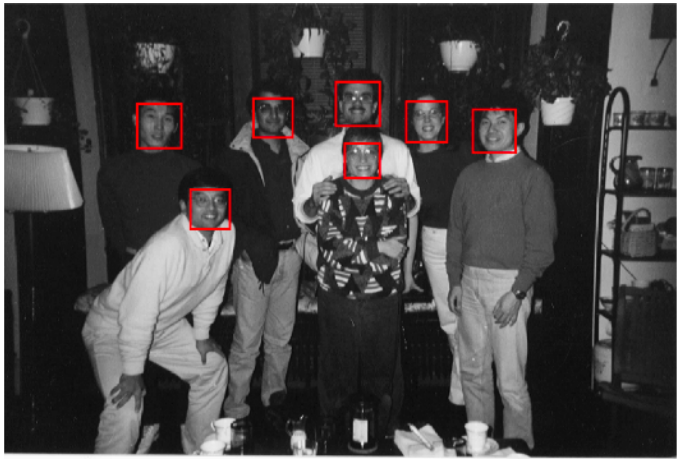
Depositing checks

Detecting (frontal) faces





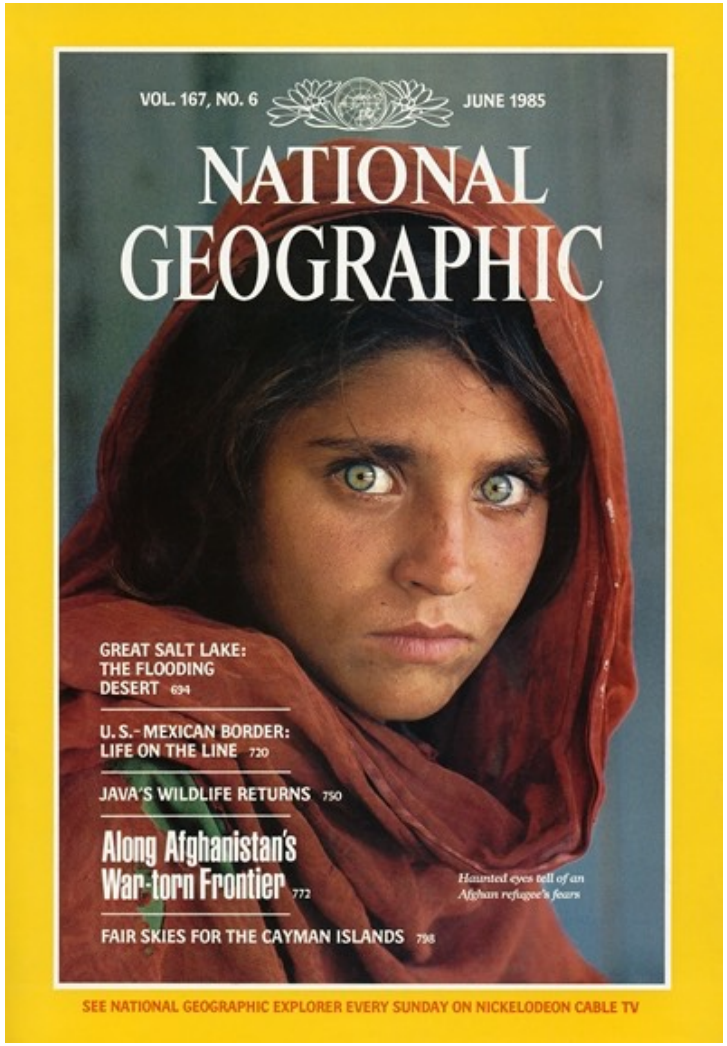




FinePix S6000fd, by Fujifilm, 2006

Viola & Jones, 2001

Face Recognition

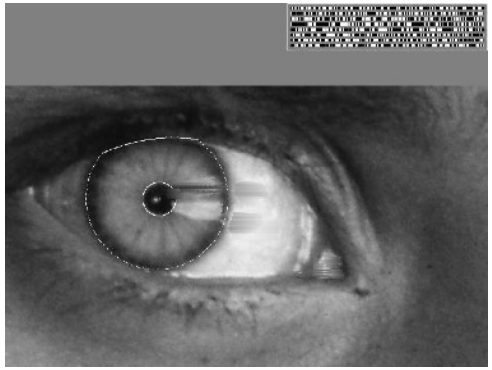


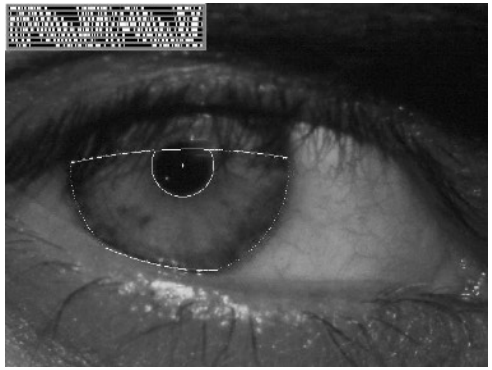
Who is she?

Slide credit: Deva Ramanan

Vision-Based Biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





Slide credit: Deva Ramanan

Vision-based security

iPhone X

Secure Authentication

Your face is now your password. Face ID is a secure new way to unlock, authenticate, and pay.

Sports

Sportvision first down line

Nice explanation on www.howstuffworks.com

Slide credit: Deva Ramanan

3D Shape Capture for Special Effects

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Slide credit: Deva Ramanan

3D Maps

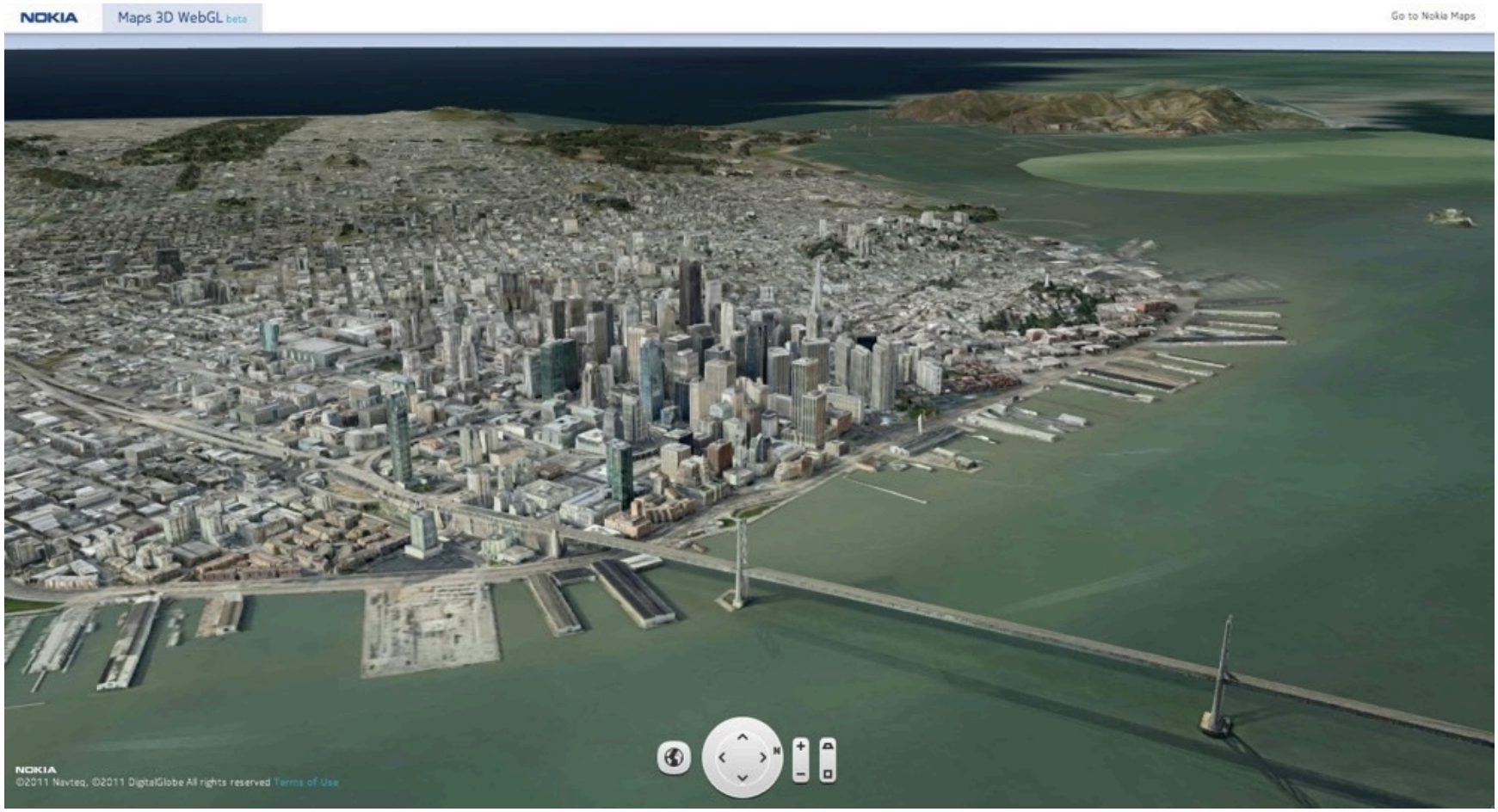


Image from Nokia's Maps 3D WebGL
(see also: Google Maps GL, Google Earth)

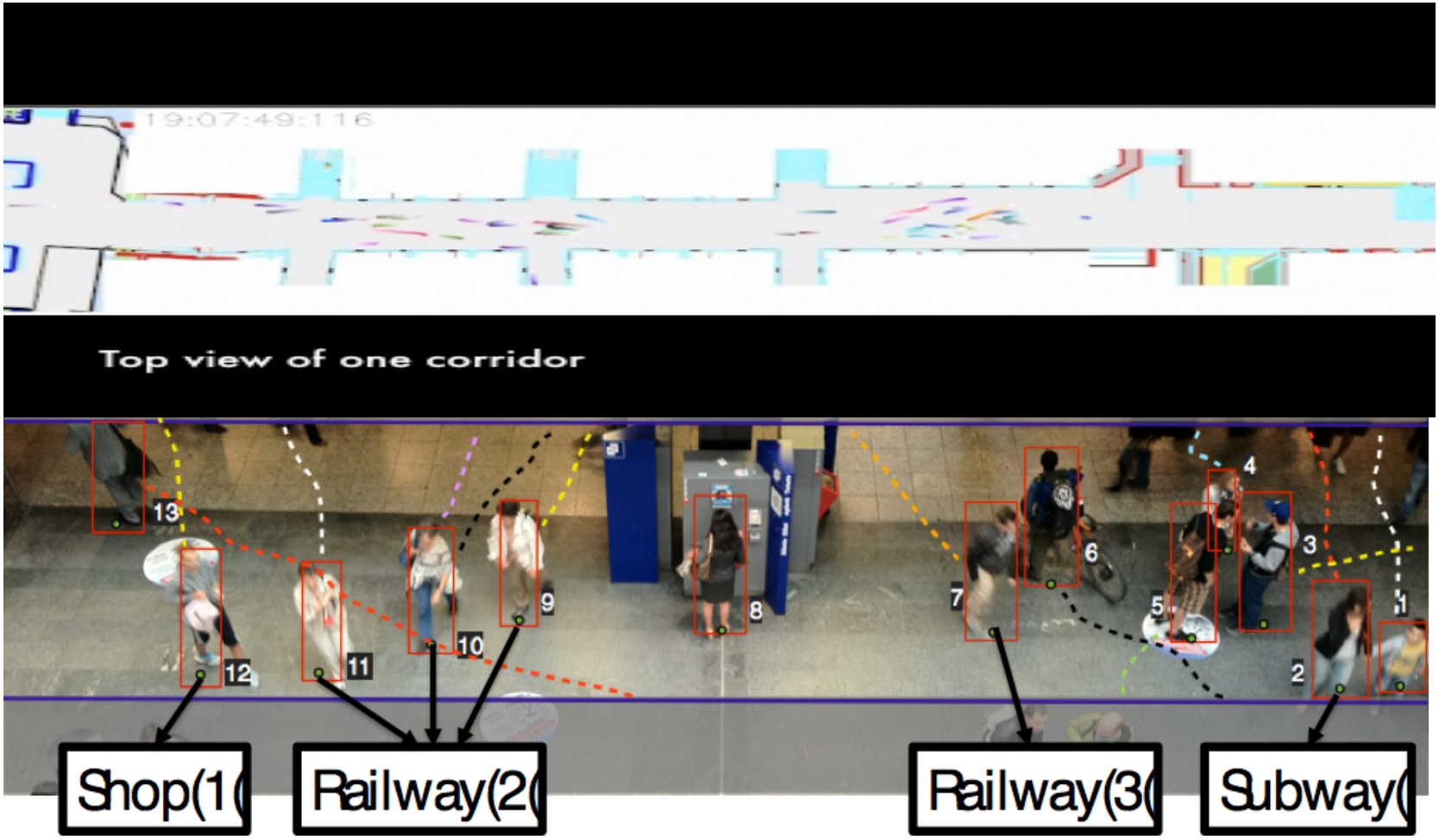
Slide credit: Deva Ramanan

Photo tourism

Reconstructing the 4D world
(UWashington/Microsoft)

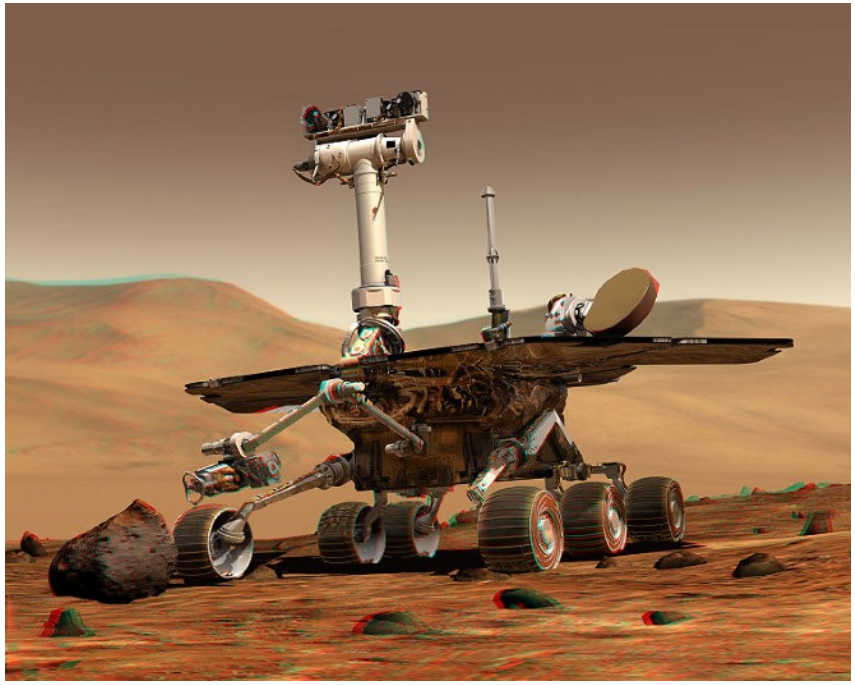
Slide credit: Deva Ramanan

Understanding traffic patterns

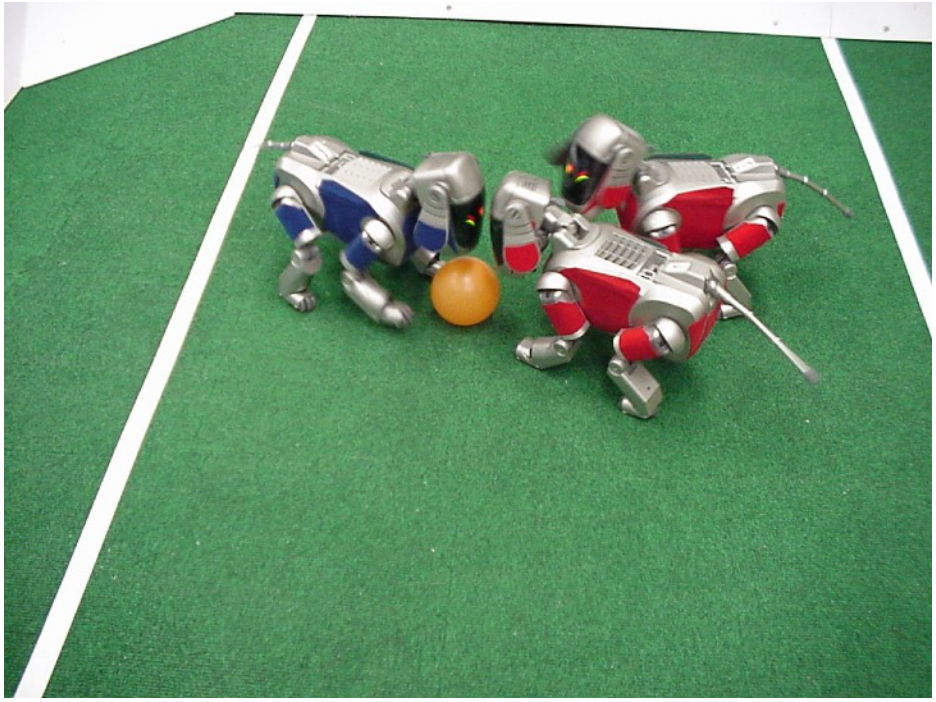


Alahi & Fei-Fei, 2014

Robotics

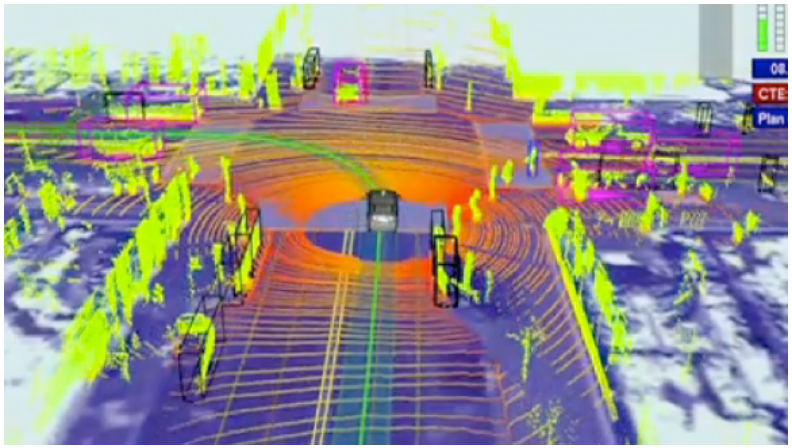


NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

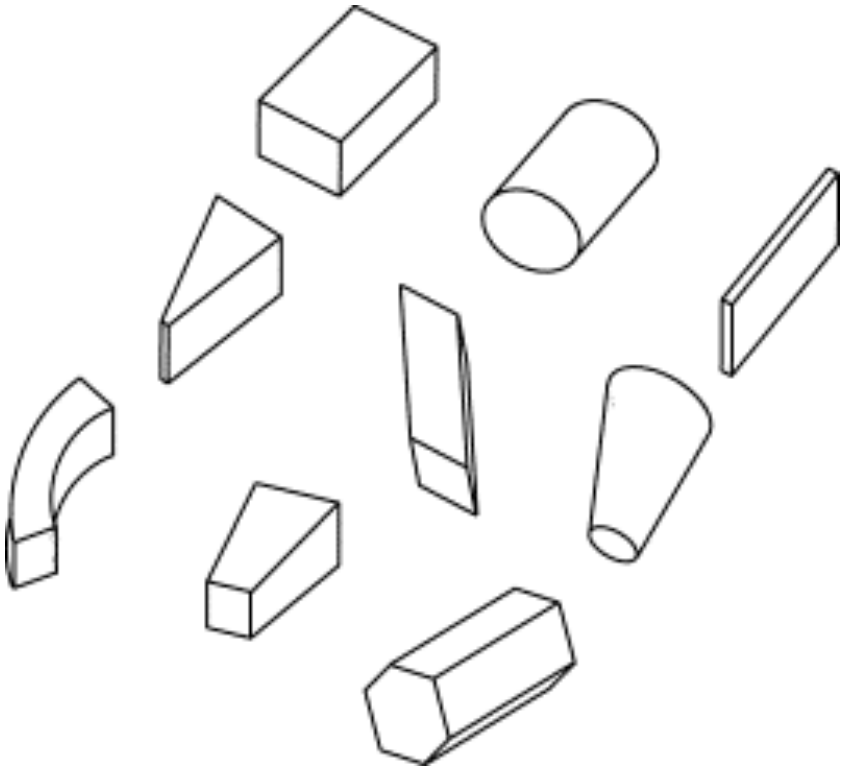


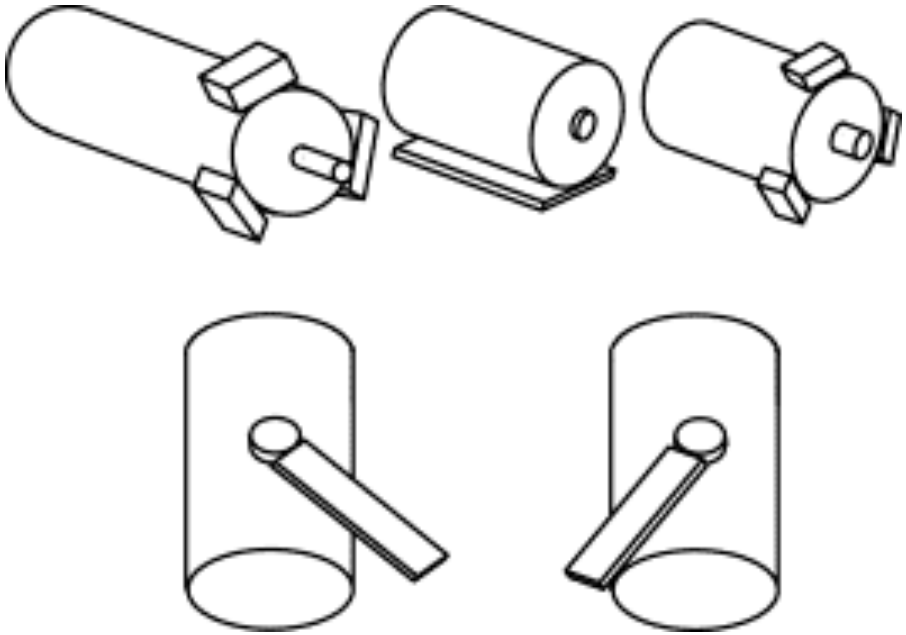
http://www.robocup.org/

Self-Driving Cars



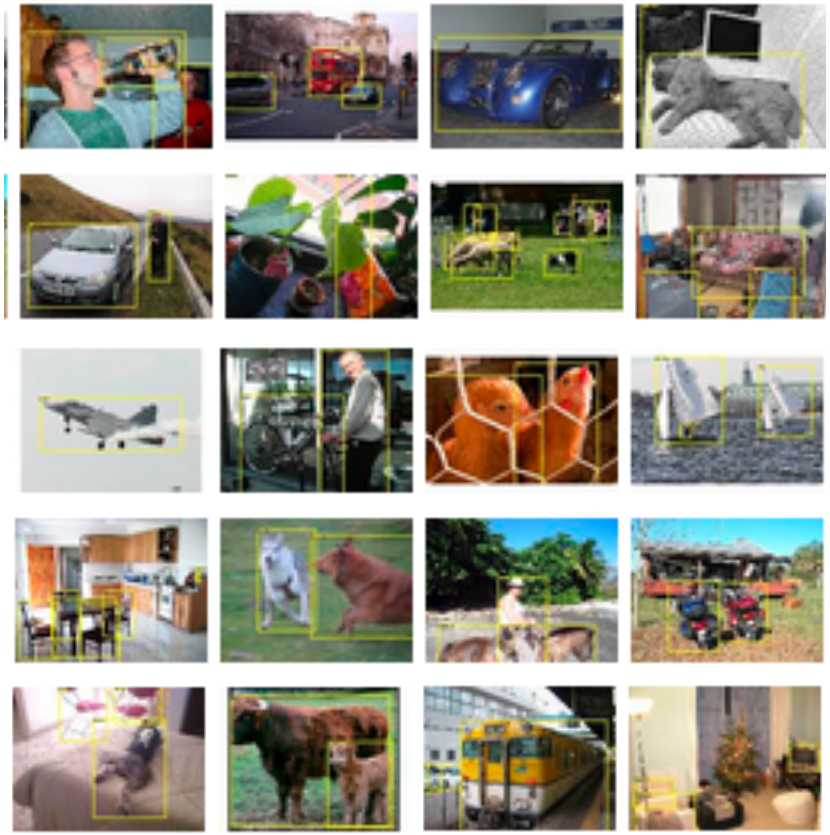
Object recognition: 1979



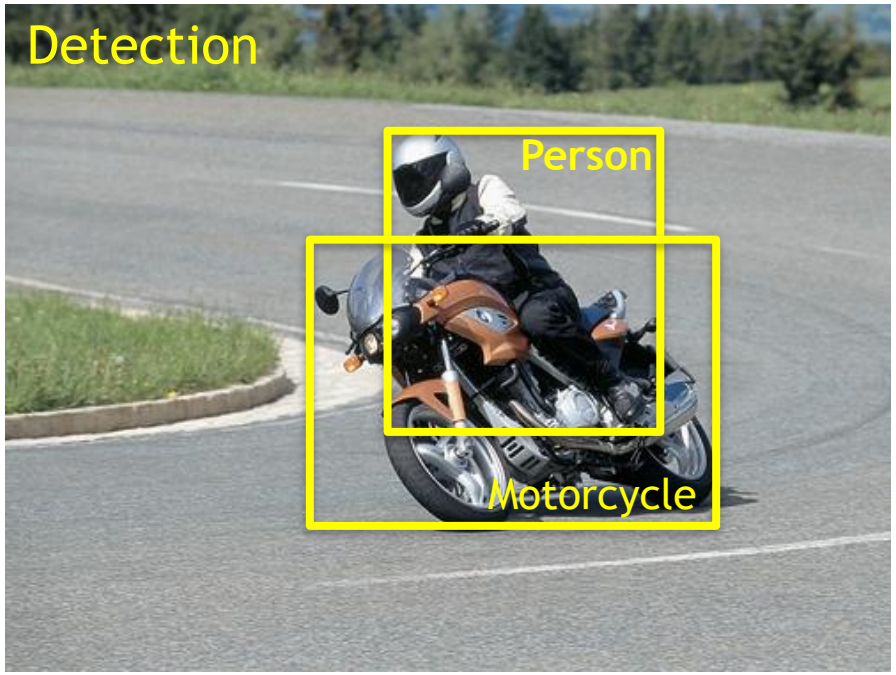


“Generalized Cylinders”, by Brooks & Binford, 1979

Object recognition: 2010

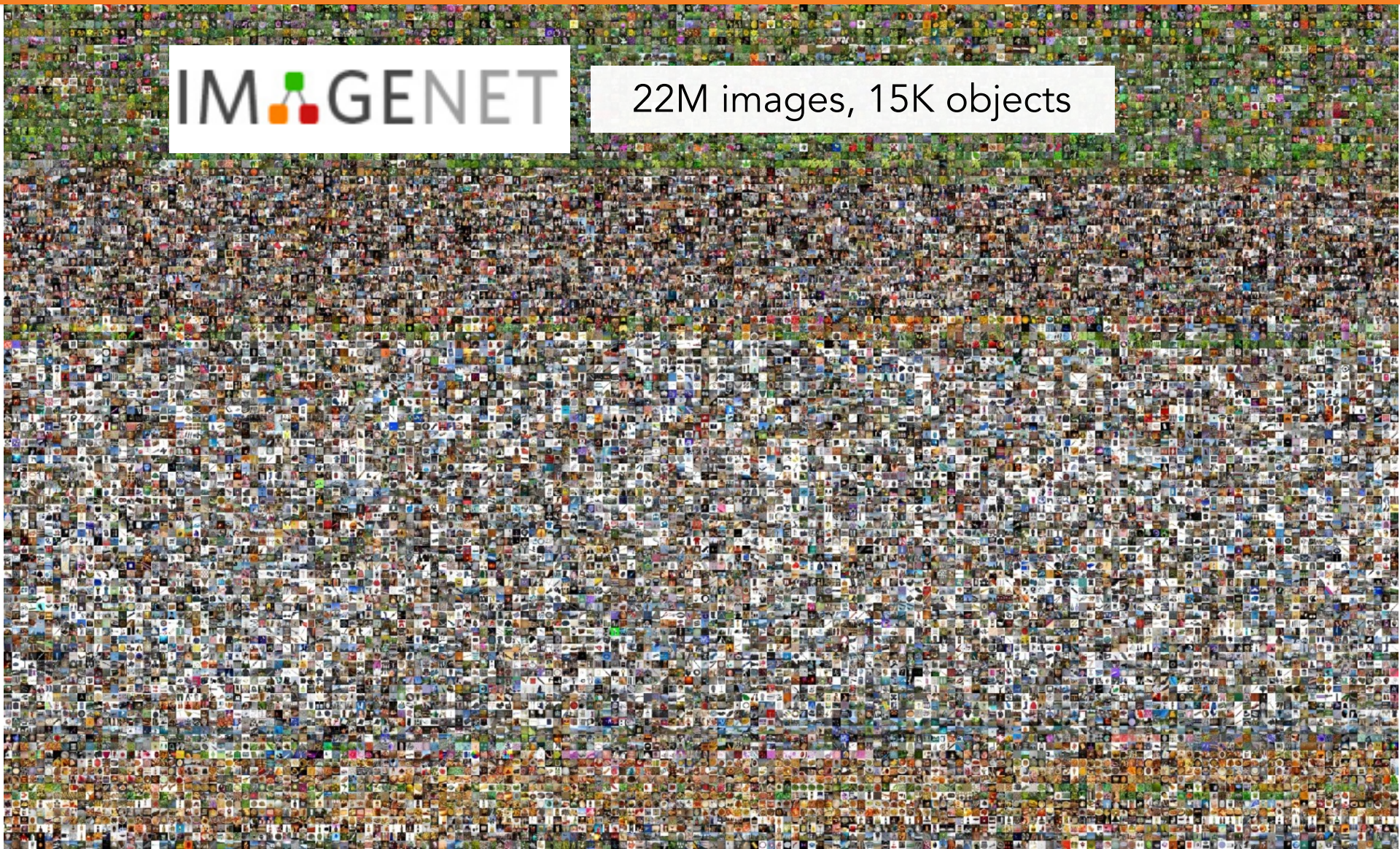


Classification: person, motorcycle



Everingham et al. IJCV 2010.

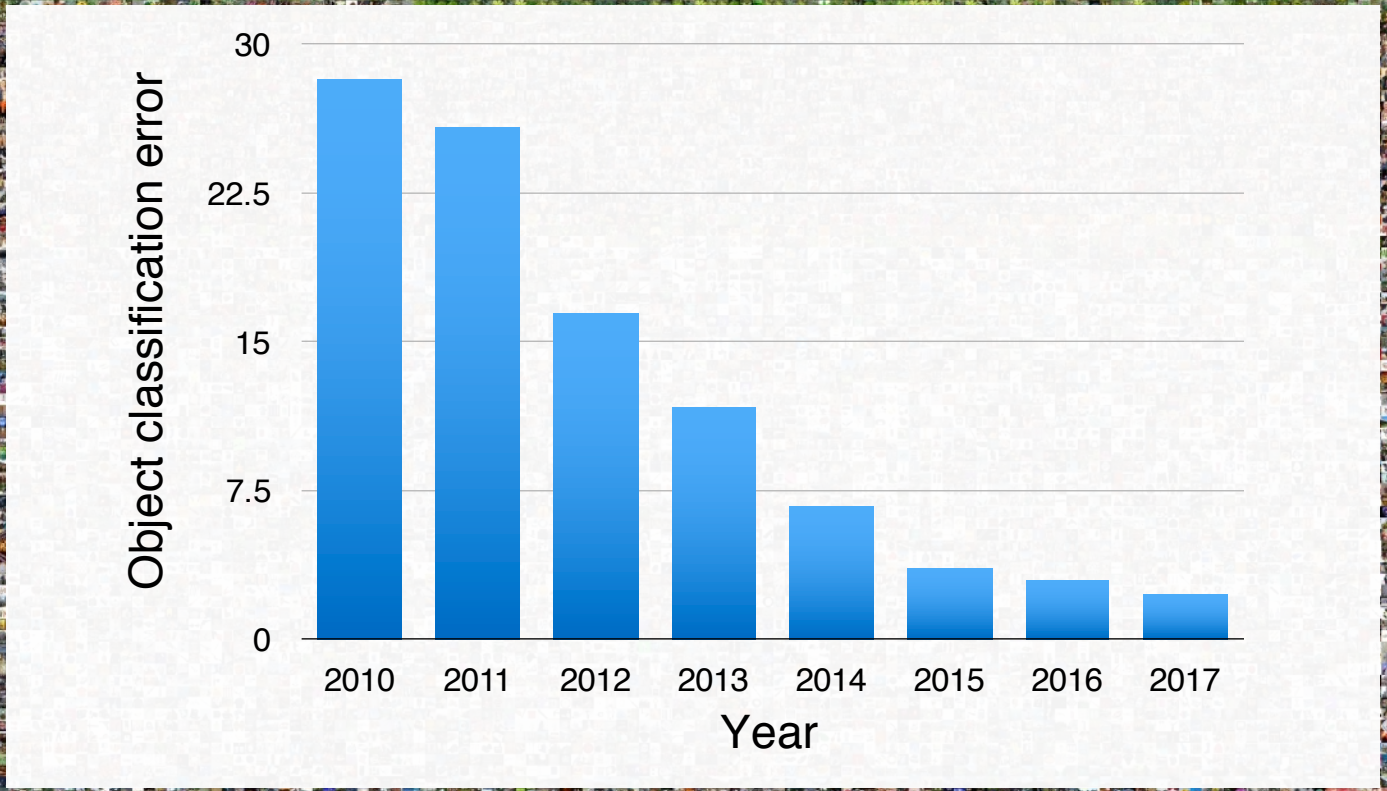
Object recognition: now



Object recognition: now

IMAGENET

(1.2M images, 1000 objects)

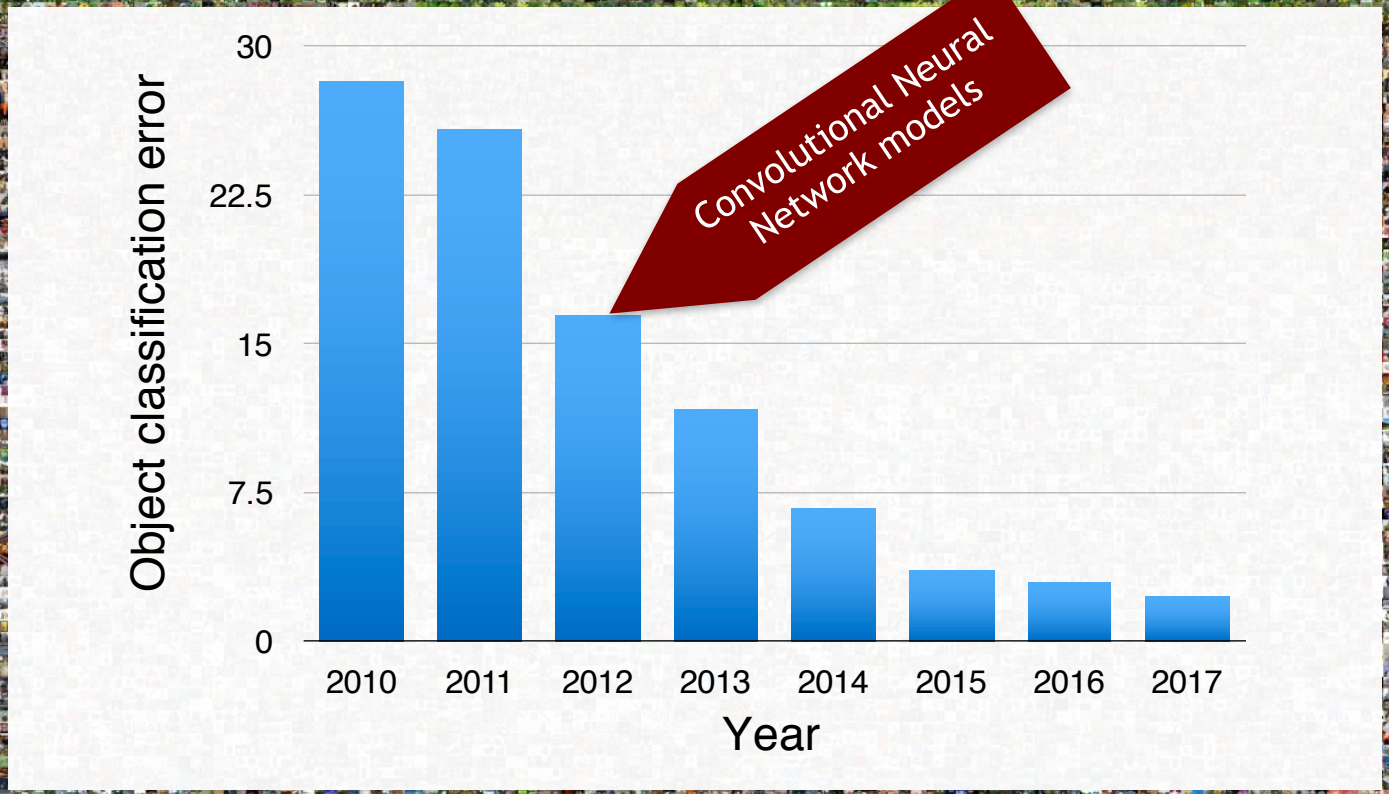


http://image-net.org

Object recognition: now

IMAGENET

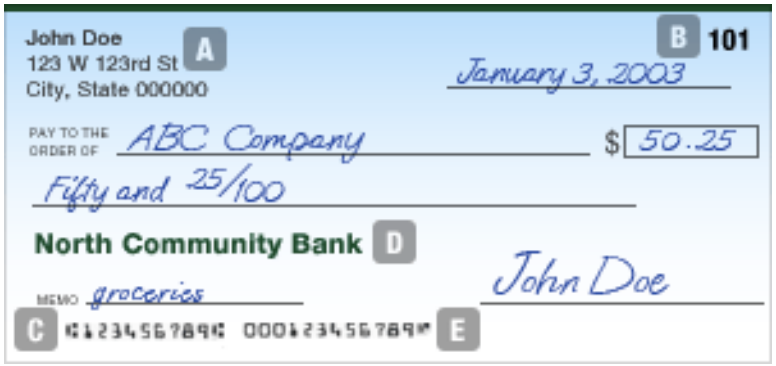
22M images, 15K objects



http://image-net.org

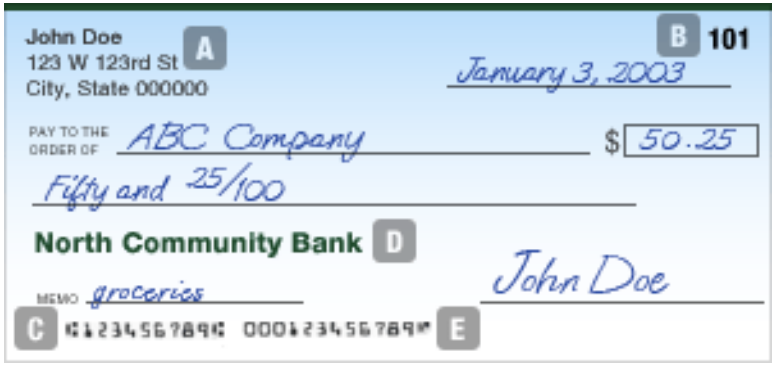
Although haven’t quite solved everything

Unconstrained environments



VS

Unconstrained environments



VS

VS

Pushing the limits with new tasks

Action recognition in videos

Approach	mAP	Approach	mAP
Random [43]	5.9	RGB++	15.6
C3D [54]	10.9	Two-Stream++	16.8
AlexNet [20]	11.3	Two-Stream+LSTM	17.8
IDT [57]	17.2	Two-Stream Extended	18.6
Two-Stream [44]	14.3	Ours (RGB Only)	18.3
		Ours	**22.4**

Table 1. Video classification results on Charades [43]. The left shows the published baselines from [43] and the right show additional new baselines. Our proposed approach outperforms all competing methods on this dataset.

[Sigurdsson CVPR 2017]

Pitfalls of looking at "overall accuracy"

Object Segmentation



SEGM: Standard Dev Challenge2015 Challenge2016

Small objects

Large objects

Copy to Clipboard | Export to CSV | Search:

	AP	AP^{50}	AP^{75}	AP^{S}	AP^{M}	AP^{L}	AR^{1}	AR^{10}	AR^{100}	AR^{S}	AR^{M}	AR^{L}	date
MSRA	0.370	0.592	0.39	0.162	0.40	0.549	.310	0.469	0.489	0.281	0.553	0.673	2016-09-16
G-RMI	0.333	0.564	0.34	0.145	0.35	0.499	.291	0.444	0.466	0.262	0.517	0.646	2016-09-17
MSRA_2015	0.279	0.512	0.27	0.086	0.30	0.453	.254	0.371	0.380	0.166	0.433	0.578	2015-11-27
FAIRCNN	0.250	0.454	0.24	0.072	0.28	0.390	.238	0.366	0.385	0.170	0.467	0.535	2015-11-26
1026	0.167	0.315	0.15	0.018	0.17	0.302	.172	0.240	0.246	0.045	0.282	0.403	2015-11-27

Show 25 entries

Previous 1 Next

http://cocodataset.org/dataset.htm#detections-leaderboard, Aug 28, 2017

Pitfalls of looking at "overall accuracy"

Visual Question Answering

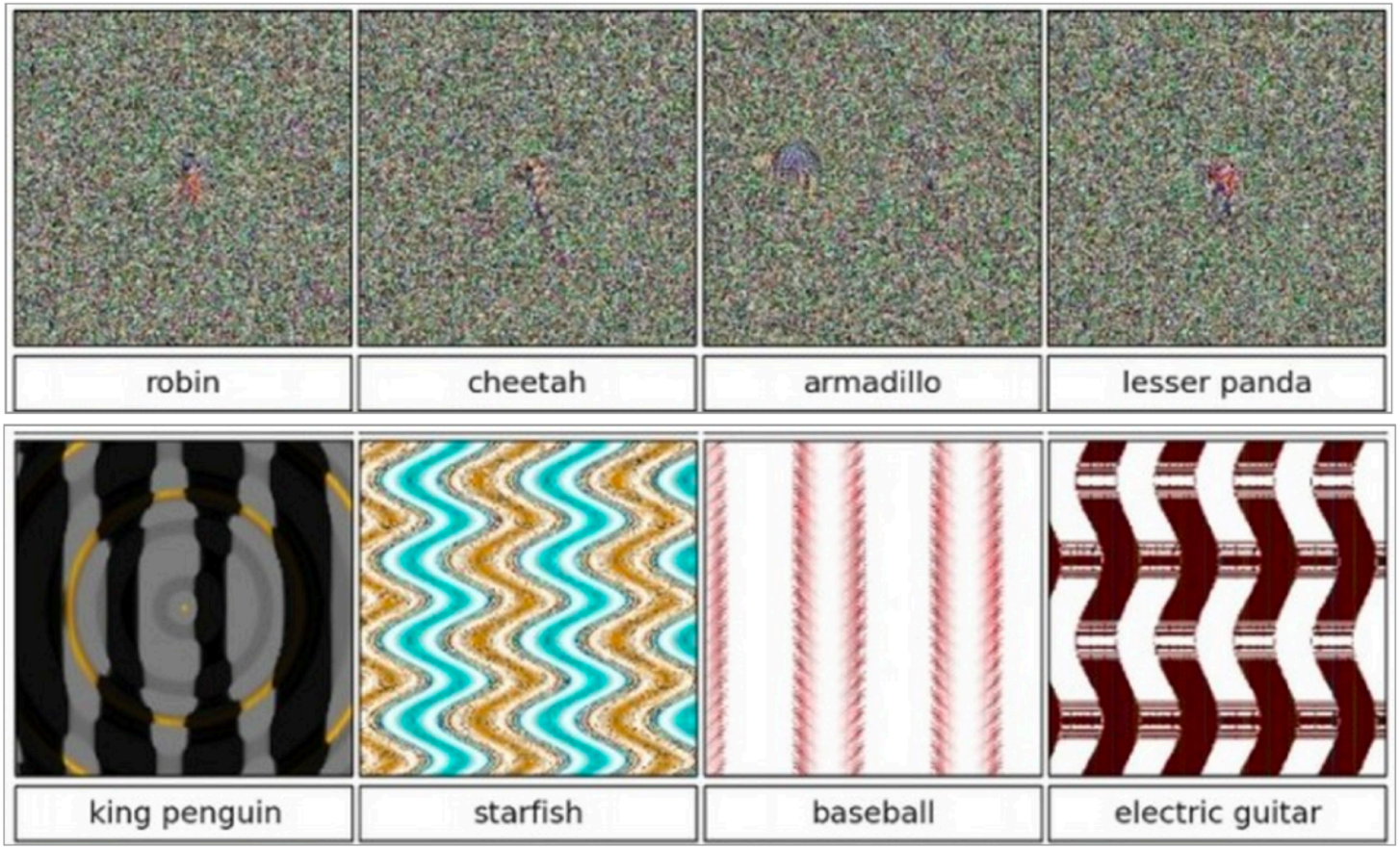
What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

	By Answer Type			Overall
	Yes/No	Number	Other	
Adelaide-Teney ACRV MSR	85.18	47.35	59.95	69
HDU-USYD-UNCC	84.28	45.36	59.42	68.16
DLAIT	82.92	46.42	60.17	68.07
LV_NUS	81.71	48.55	59.79	67.62

http://www.visualqa.org/roe_2017.html, results as of 07/2017

Surprising failures

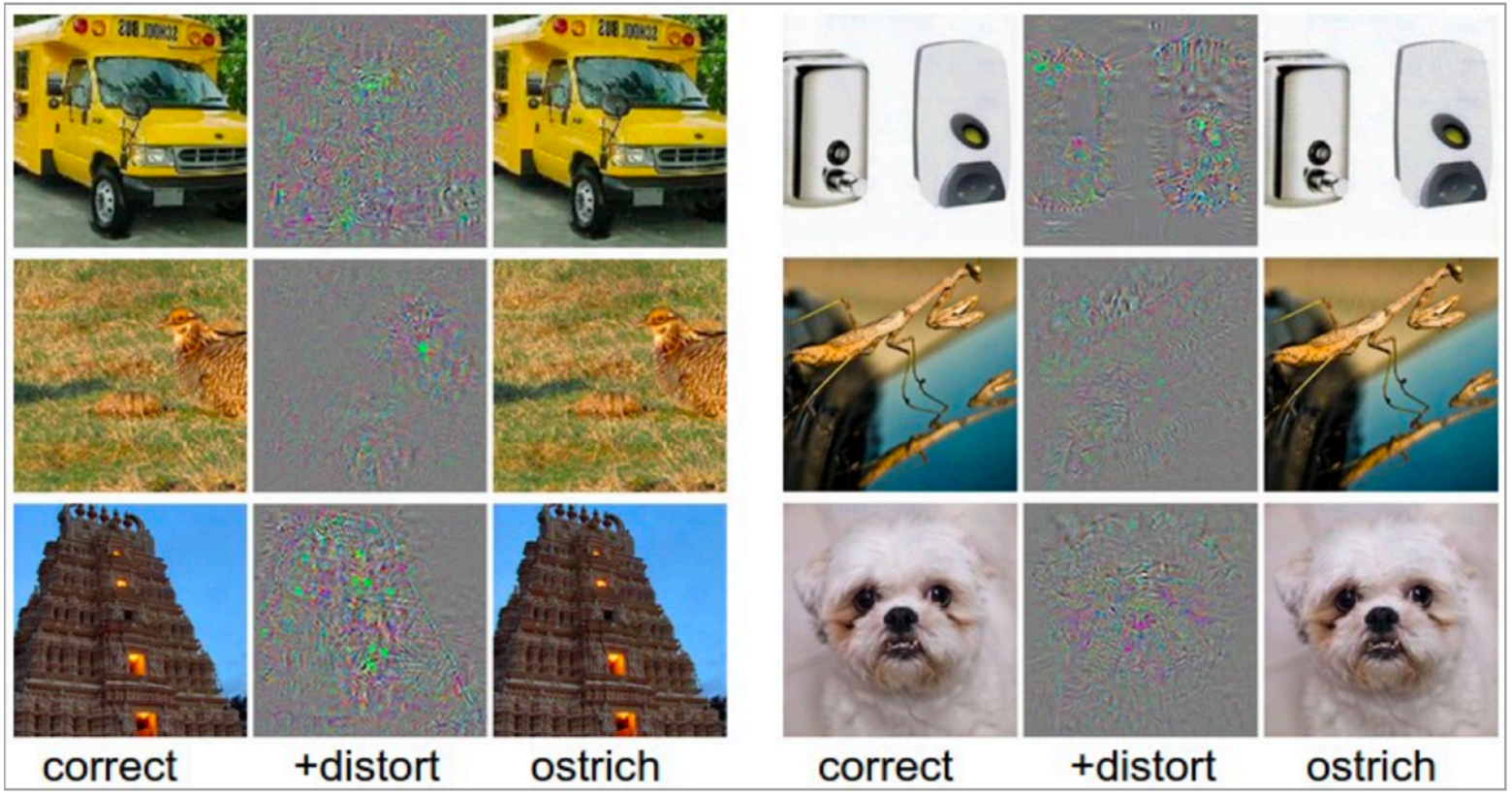


These images are classified with >99.6% confidence as the shown class by a Convolutional Network.

[Nguyen, Yosinski, Clune CVPR 2015]

http://karpathy.github.io/2015/03/30/breaking-convnets/

Surprising failures



Take a correctly classified image (left image in both columns), and add a tiny distortion (middle) to fool the ConvNet with the resulting image (right).

[Szegedy et al. Intriguing properties of neural networks]

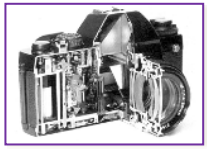
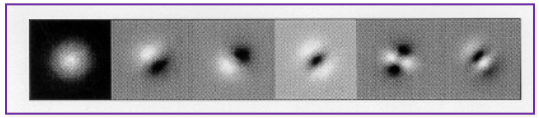
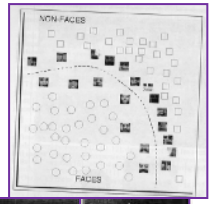
http://karpathy.github.io/2015/03/30/breaking-convnets/

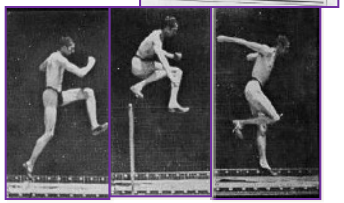
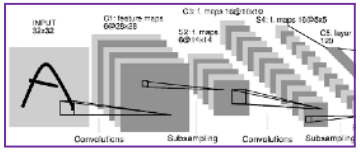
(aside) Assignment 0

- See if you can fool a modern image classification model

Our plan in COS 429

Course Outline

- Image formation and capture 
- Filtering and feature detection 
- Segmentation and clustering 
- Recognition and classification 

- Motion estimation and tracking 
- 3D shape reconstruction 
- Convolutional neural nets / deep learning 

Course Staff

Instructor: Prof. Olga Russakovsky

Guest lecturer: Dr. Andras Ferencz, MobilEye

TAs: Kyle Genova and Riley Simmons-Edler

Q&A

- We will use piazza for Q&A. Please direct all non-private questions there.

- Feel free to answer each others' questions (we will monitor and endorse students' answers) but keep in mind collaboration policy

Course Mechanics

- Recommended book:

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

  - Also available online: http://szeliski.org/Book/

- Assigned papers / other readings

Course Mechanics

- 56%: 4 written / programming assignments
  - Individual or with a partner (but same partner on at most 2 assignments)
  - 4 free late days, at most 2 per assignment
- 20%: Midterm
  - Thu, Oct 26th, **no exceptions**
- 24%: Final project
  - Small groups of 1-3 people
  - Start working soon and do a cool project!
  - Milestone due Dec 15, poster session in January
  - Writeup due on Dean's date

MATLAB

- The assignments use the MATLAB language.
- Assignment 0 will walk you through the basics
  - Highly recommended, by Sept 21st
- Install on your computer:
  - `http://www.princeton.edu/software/licenses/software/matlab/`
- Matlab workshop offered by School of Engineering
  - Sept 18-19, registration required
  - `https://www.tfaforms.com/4634935`

Next class: image formation and capture