Introduction and overview

370: Intro to Computer Vision

Subhransu Maji

Jan 30, 2025

College of
INFORMATION AND
COMPUTER SCIENCES

UMASS
AMHERST

Who are we?

Instructor

Subhransu Maji

Office hours: ?? @ CS 274

TA

Aaron Sun

Office hours: Wed 1-2pm @ CS207

TA

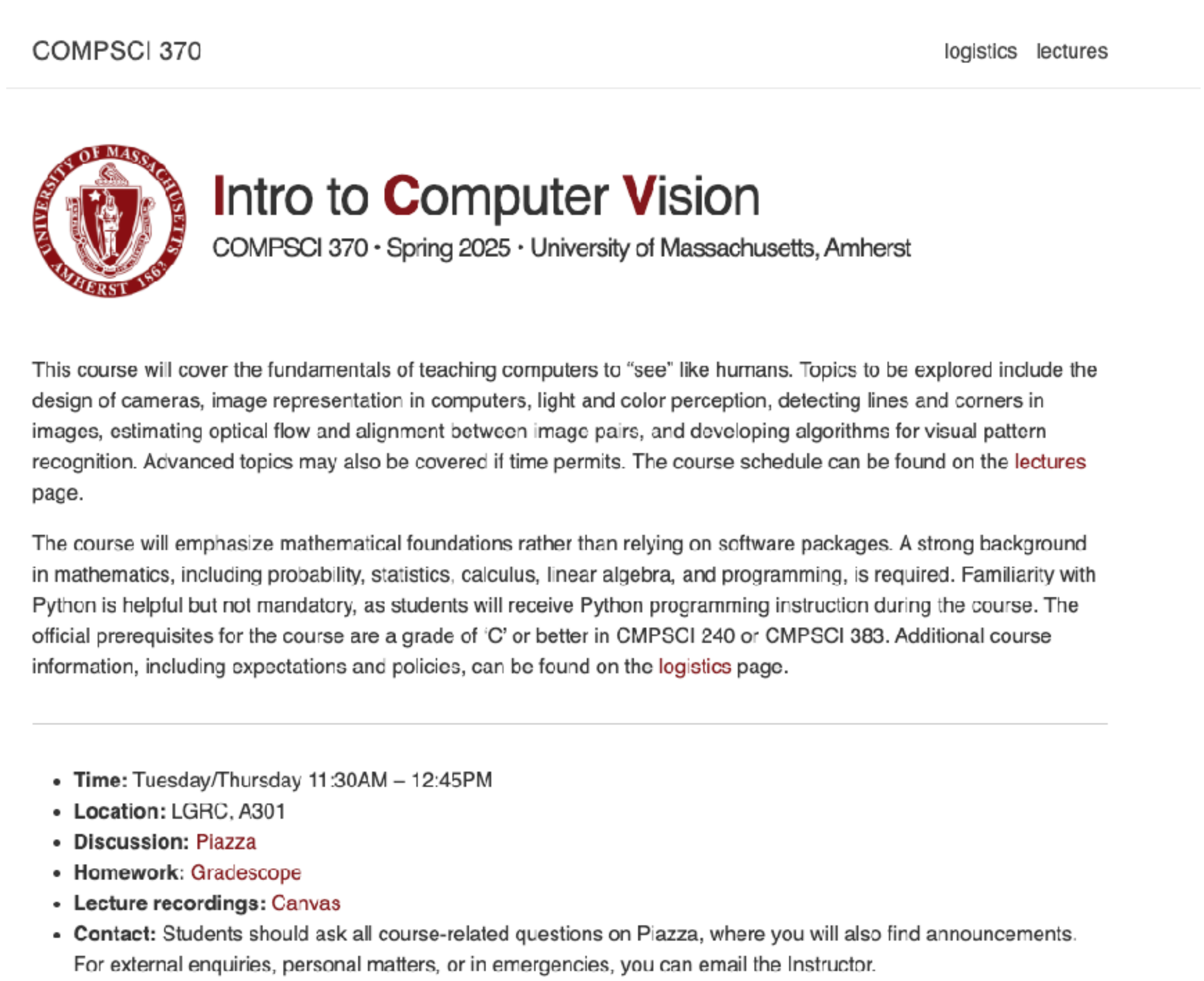
Frank Chiu

Office hours: Tue 2:30-3:30 @ CS207

Course info

Course website: https://cvl-umass.github.io/intro-cv-spring-2025

Read the course logistics and lectures



COMPSCI 370 logistics lectures

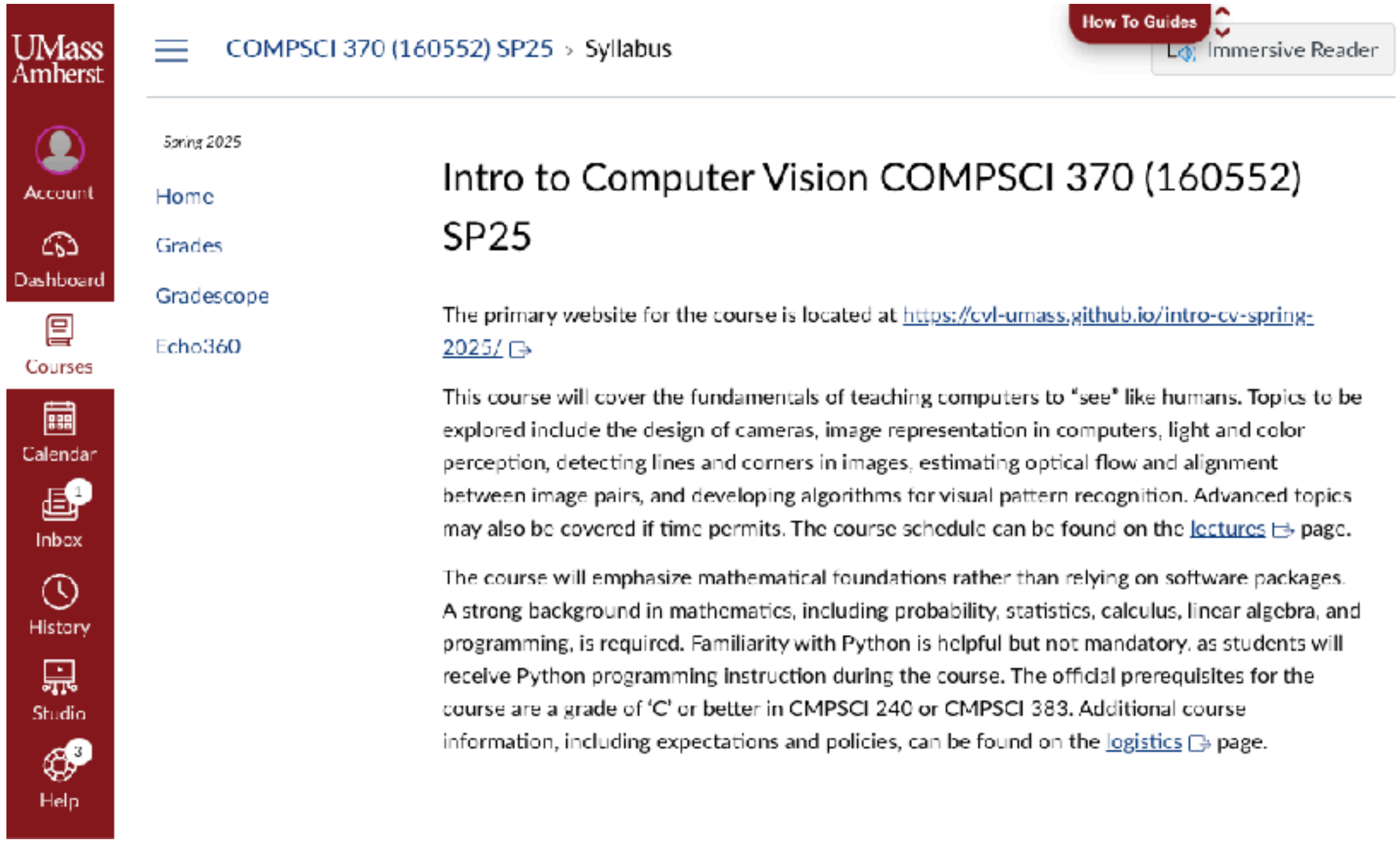
Intro to Computer Vision

COMPSCI 370 • Spring 2025 • University of Massachusetts, Amherst

This course will cover the fundamentals of teaching computers to "see" like humans. Topics to be explored include the design of cameras, image representation in computers, light and color perception, detecting lines and corners in images, estimating optical flow and alignment between image pairs, and developing algorithms for visual pattern recognition. Advanced topics may also be covered if time permits. The course schedule can be found on the lectures page.

The course will emphasize mathematical foundations rather than relying on software packages. A strong background in mathematics, including probability, statistics, calculus, linear algebra, and programming, is required. Familiarity with Python is helpful but not mandatory, as students will receive Python programming instruction during the course. The official prerequisites for the course are a grade of 'C' or better in CMPSCI 240 or CMPSCI 383. Additional course information, including expectations and policies, can be found on the logistics page.

- **Time:** Tuesday/Thursday 11:30AM – 12:45PM
- **Location:** LGRC, A301
- **Discussion:** Piazza
- **Homework:** Gradescope
- **Lecture recordings:** Canvas
- **Contact:** Students should ask all course-related questions on Piazza, where you will also find announcements. For external enquiries, personal matters, or in emergencies, you can email the Instructor.



Intro to Computer Vision COMPSCI 370 (160552) SP25

The primary website for the course is located at https://cvl-umass.github.io/intro-cv-spring-2025/

This course will cover the fundamentals of teaching computers to "see" like humans. Topics to be explored include the design of cameras, image representation in computers, light and color perception, detecting lines and corners in images, estimating optical flow and alignment between image pairs, and developing algorithms for visual pattern recognition. Advanced topics may also be covered if time permits. The course schedule can be found on the lectures page.

The course will emphasize mathematical foundations rather than relying on software packages. A strong background in mathematics, including probability, statistics, calculus, linear algebra, and programming, is required. Familiarity with Python is helpful but not mandatory, as students will receive Python programming instruction during the course. The official prerequisites for the course are a grade of 'C' or better in CMPSCI 240 or CMPSCI 383. Additional course information, including expectations and policies, can be found on the logistics page.

Course info

Course website: https://cvl-umass.github.io/intro-cv-spring-2025

Read the course logistics and lectures

COMPSCI 370 logistics lectures

Logistics

Textbooks
Required Background
Grading
Course policies
Accommodation Statement
Acknowledgements

Textbooks

The primary material for the class are lectures and readings listed on the lectures page. There is no required textbook for this class. Nevertheless, the following textbooks might be useful, even though they are aimed at a graduate audience.

- Computer Vision: A Modern Approach by David Forsyth and Jean Ponce (2nd ed.)
- Computer Vision: Algorithms and Applications, by Richard Szeliski (2nd ed.) (available online).

We will post links to sections of Szeliski's book for each lecture. These readings are not required, but they might be helpful especially if you want to dig deeper into specific topics.

Course info

Course website: https://cvl-umass.github.io/intro-cv-spring-2025

Read the course logistics and lectures

COMPSCI 370 logistics lectures

Class schedule

Date	Topic	Readings and Annoucements
1/30	**Introduction and logistics**	• Slides • Szeliski book, Chapter 1 (optional) • The speed of processing in the human visual system, Thorpe et al., 1996 (optional)
	Module 1: Image Formation	
2/4	**Light and color: I** - Spectral basis of light - Color perception in the human eye	• Szeliski book, Chapter 2 • Beau Lotto's TED talk • **Homework 1** released (due 2/22)

Requirements and grading

Homework: **45%**

- 5 in total
- Completed individually (use of AI not permitted)
- Roughly every two weeks

Midterm: **20%** (3/13 in class)

Final: **30%** (5/14 1-3pm,)

Class participation: **5%**

- In-class activities — low-stakes quizzes throughout the semester
- Echo 360:
  - Will available within a few hours
  - Unreliable — fails 10% of the time

Who should take this course?

Do you have all the pre-requisites?

- Math — good understanding of calculus, linear algebra and probability
- Programming — ability to program in Python

Teaching style

- Slides, notes, tutorials
- Optional readings — papers, articles, references to books

Still not sure?

- Email / drop by office hours for a chat

Waitlisted?

- Will decide on a case by case basis (mostly limited by space)

Course background

What is the course about?

- Physics and geometry of image formation
- Finding and exploiting patterns in visual data
- **It is hard, ad-hoc —** few theorems, but we rely on those from other areas

Why study vision?

- **You are in good company:** Euclid, Alhazen, da Vinci, Kepler, Galileo, Descartes, Newton, Huygens, Maxwell, Helmholtz, Mach, Herring, Cajal, Minkowski, Hubel, Wiesel, Wald
- **Broad applicability:** robotics, astronomy, ecology, medicine
- Open area, lots of room for new work

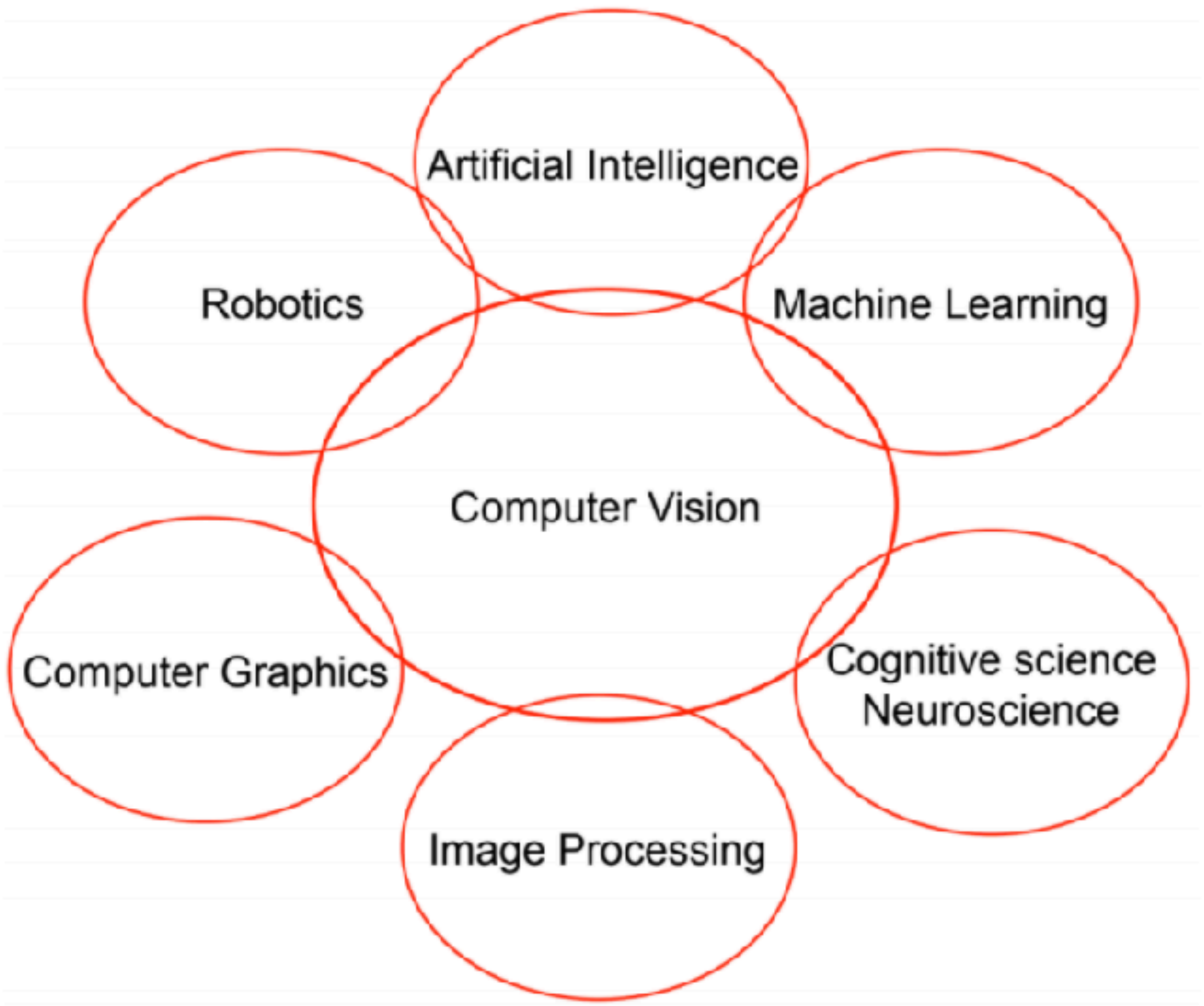


Figure credit: Kristen Grauman

Topics covered

Image

Formation
- Geometry & physics
- Design of cameras
- Color perception & phenomenon

7x lectures
1x homework

Processing
- Digital representation
- Signal processing
- Linear filtering
- Applications

7x lectures
2x homework

Understanding
- Alignment & matching
- Recognition
- Datasets & benchmarks
- Machine learning
- Advanced topics

9x lectures
2x homework

Not a zoo tour!

Not an introduction to tools!

You will learn how these techniques work and how to implement them

Course goals

By the end of the semester, you should be able to:

- Look at a problem and identify if CV is an appropriate solution
- If so, identify what types of algorithms might be applicable
- Apply those algorithms, conquer the world
- Consider taking other courses in AI

In order to get there, you will need to:

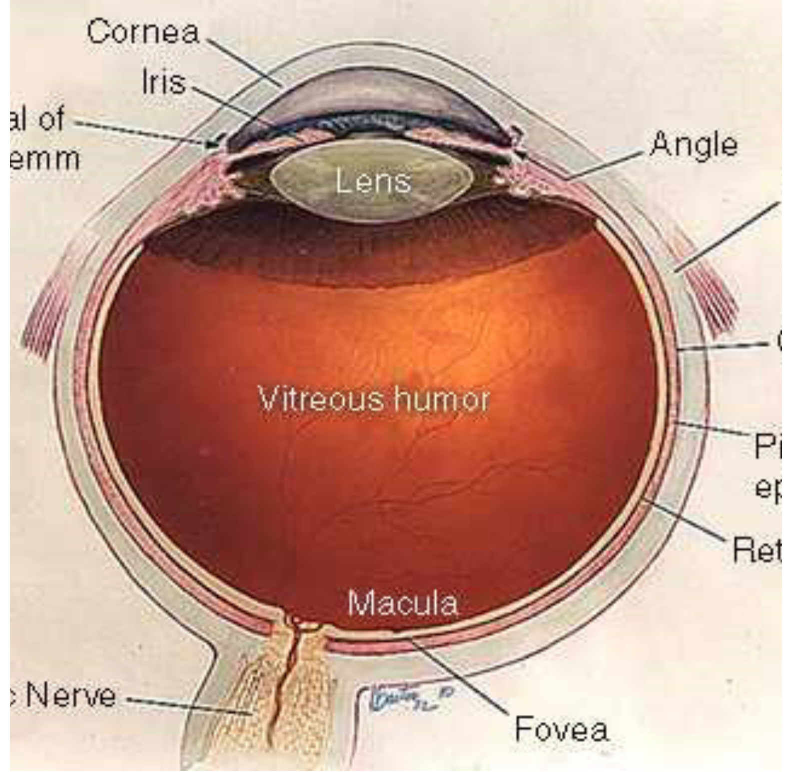
- Do a lot of math (calculus, linear algebra, probability)
- Do a fair amount of programming
- Work hard (this is a 3-unit course)

Now, on to some real content …

(but first, questions?)

Why vision? Light!

It is how we see other people,
navigate our environment,
communicate ideas, entertain,
and **measure** the world around us.

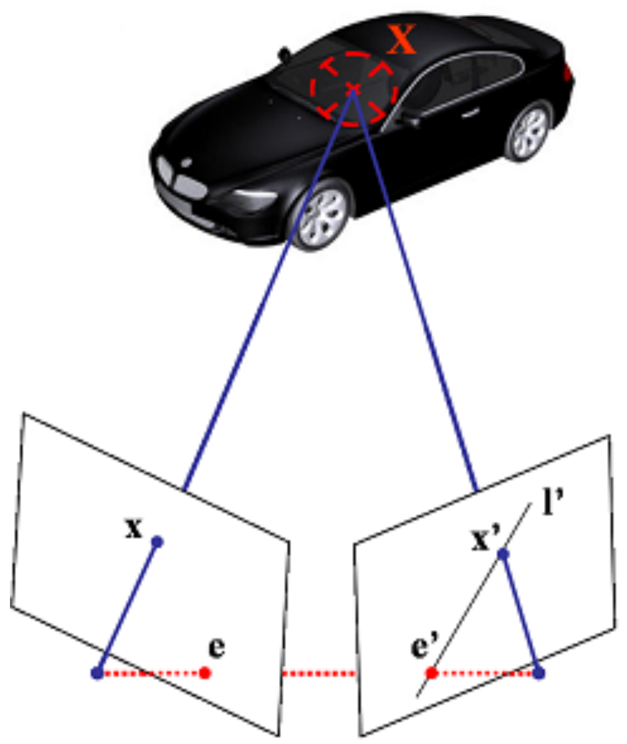


Source: A. Berg

Why is light good for measurement?

Microscopy

Surveillance



3D Analysis / Navigation

Remote Sensing

- Plentiful, sometimes free
- Interacts with many things, but not too many
- Goes generally straight over distance
- Very small → high spatial resolution
- Fast, but not too fast → time of flight sensors
- Easy to detect → cameras work, are cheap

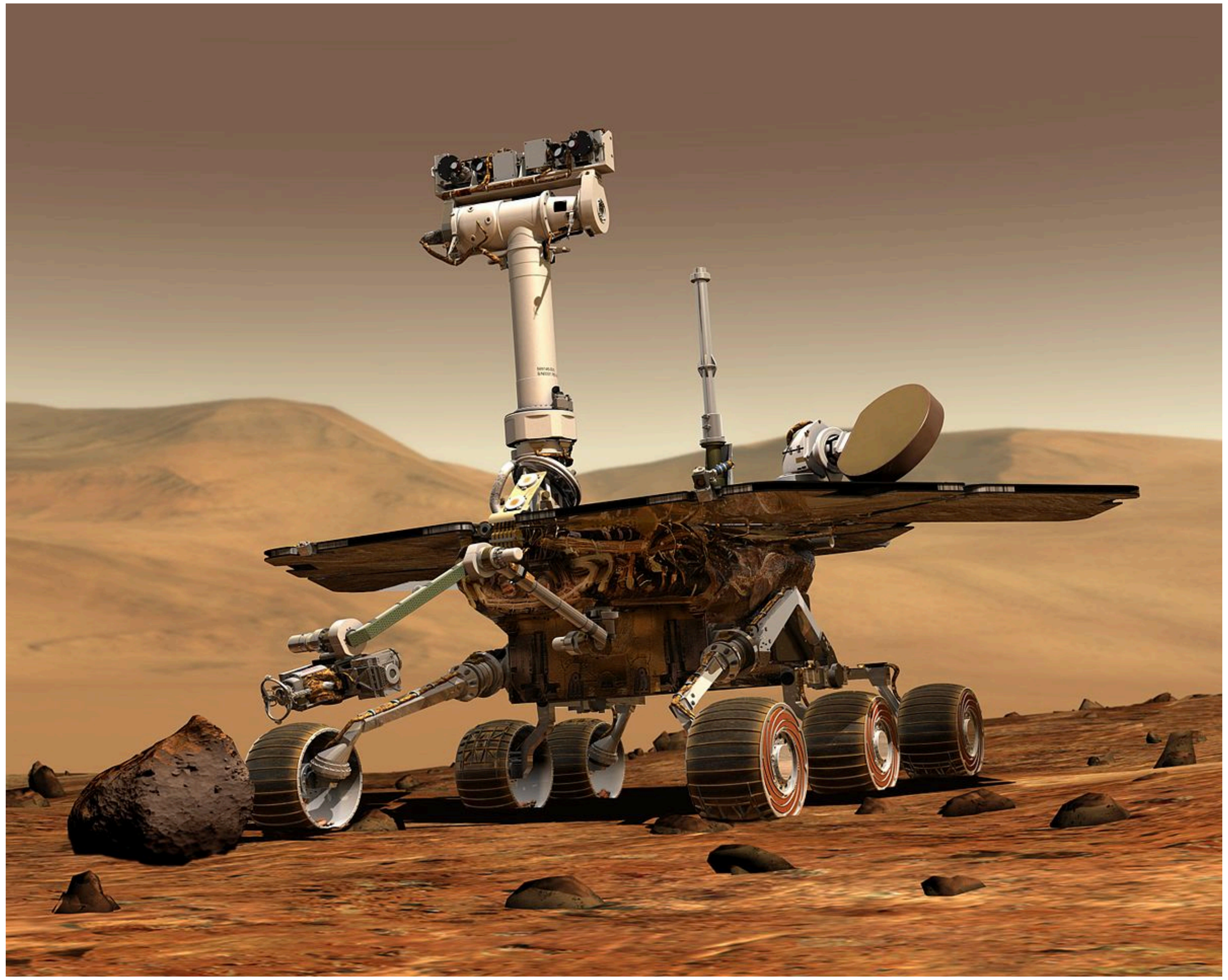
Source: Alex Berg

Goal of computer vision

Extract properties of the world from visual data
(i.e., measurements of light)

What properties to extract?

Example 1: Robotics



What properties to extract?

Example 2: Internet Vision


Facebook Users Are Uploading 350 Million New Photos Each Day

Cooper Smith

Sep. 18, 2013, 8:00 AM 23,351

Slide credit: B. Hariharan

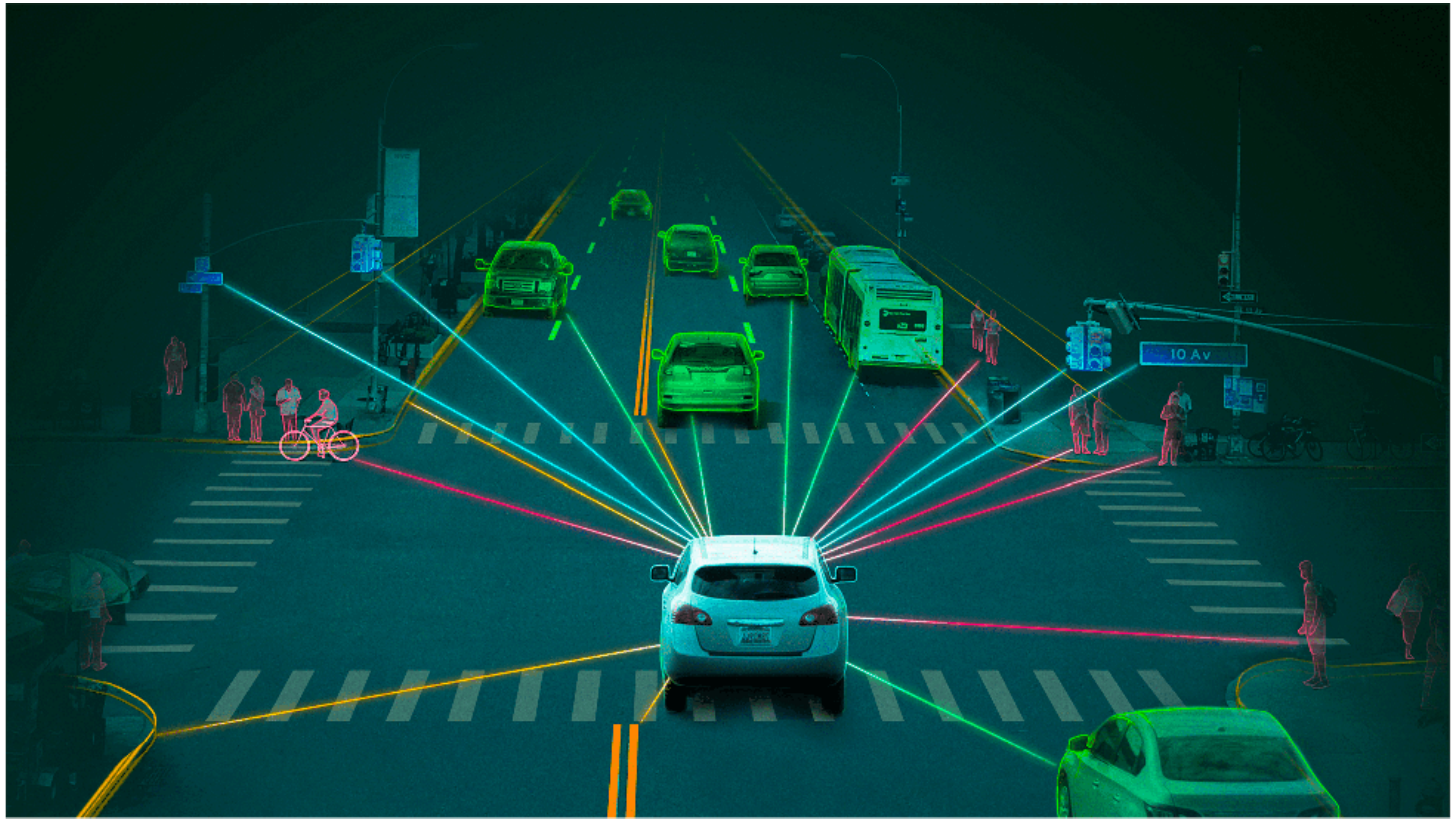
What properties to extract?

Example 3: Amazon Go

https://www.recode.net/2018/1/21/16914188/amazon-go-grocery-convenience-store-opening-seattle-dilip-kumar

What properties to extract?

Example 4: Autonomous driving



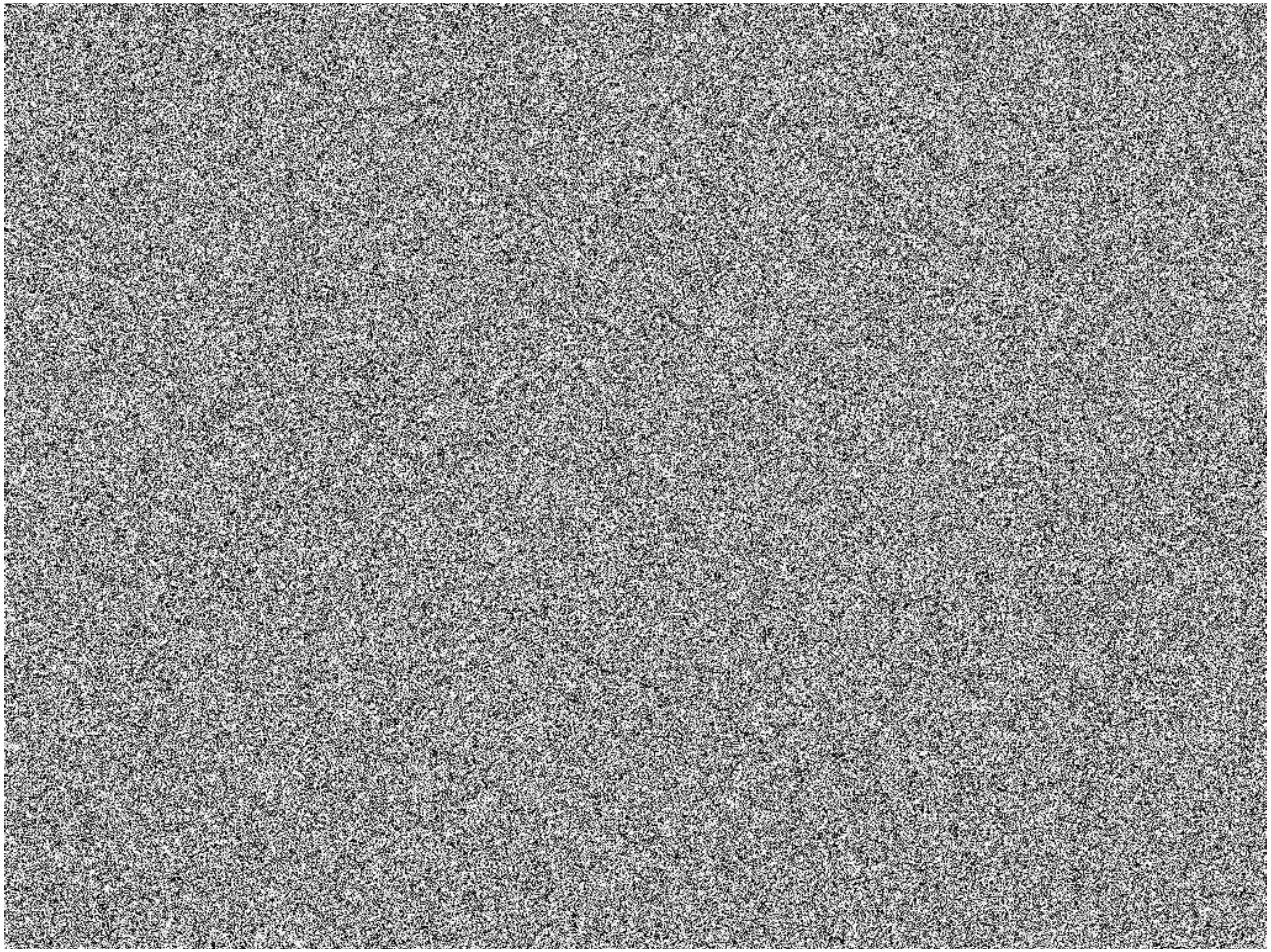
https://researchleap.com/research-in-autonomous-driving-a-historic-bibliometric-view-of-the-research-development-in-autonomous-driving/

Goal of computer vision

Extract **enough** information of the world from visual data to make **good decisions**
(i.e., measurements of light)

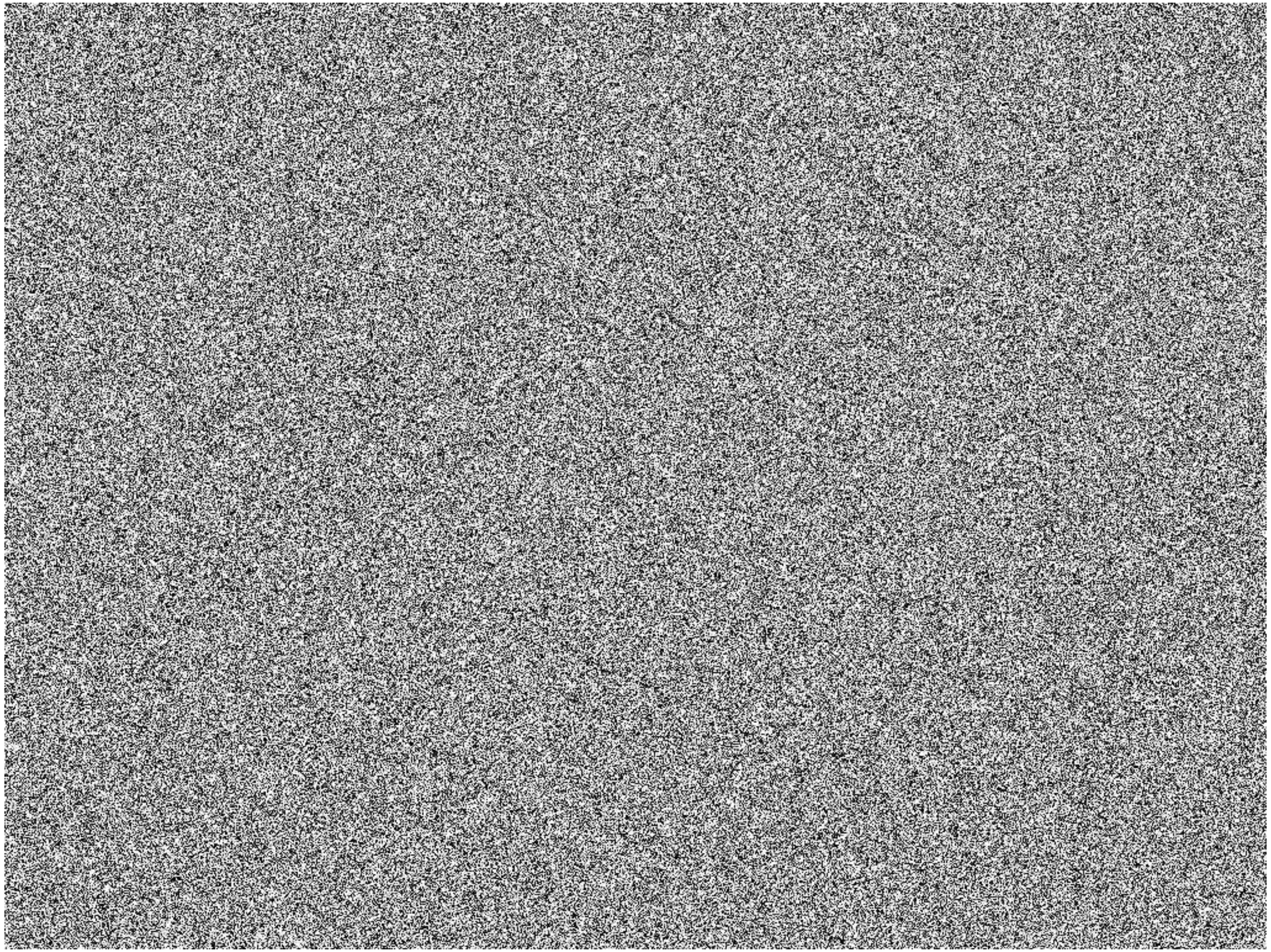
We are remarkably good at this!

Example #1



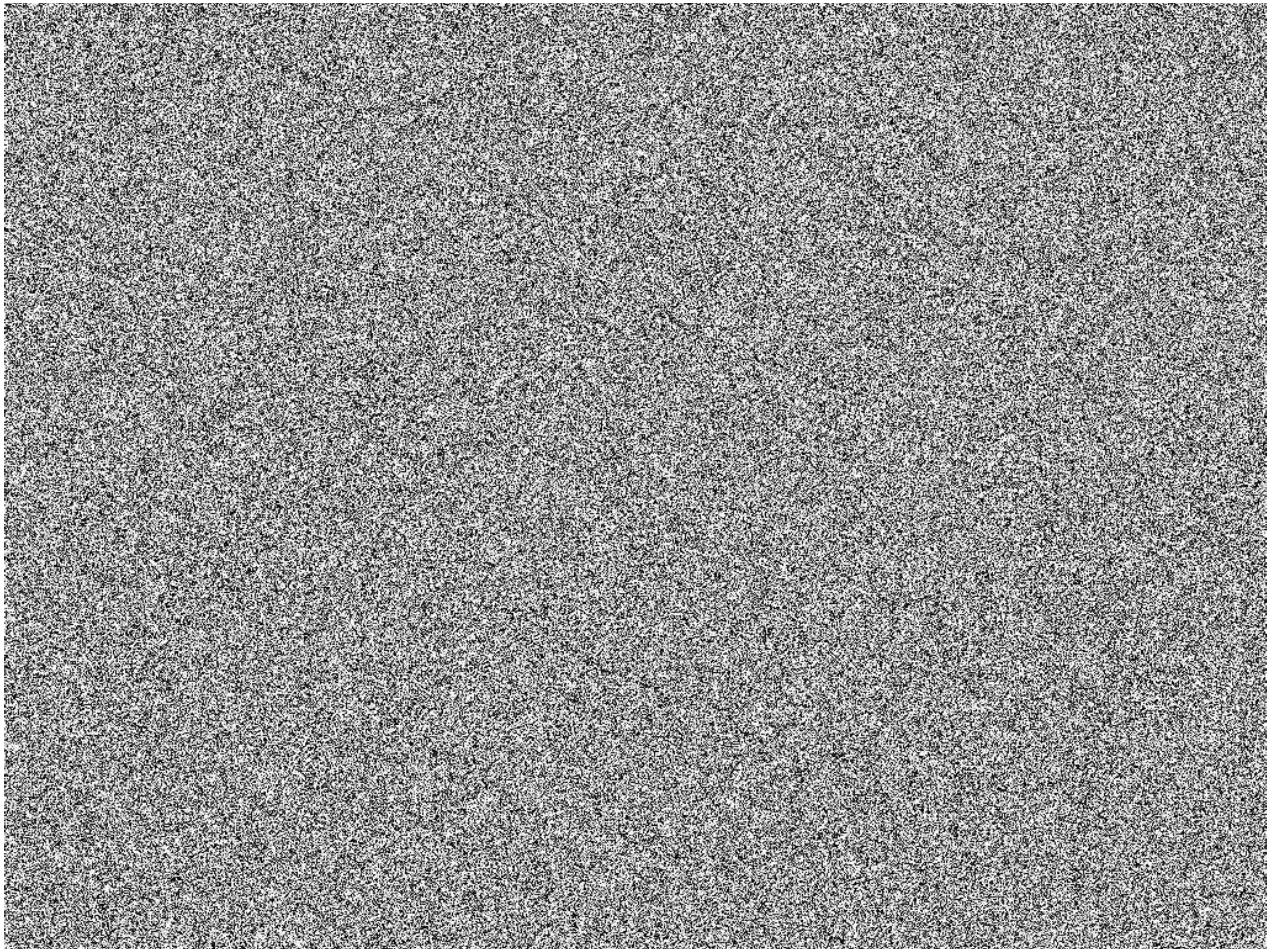
animal or not?

Example #2



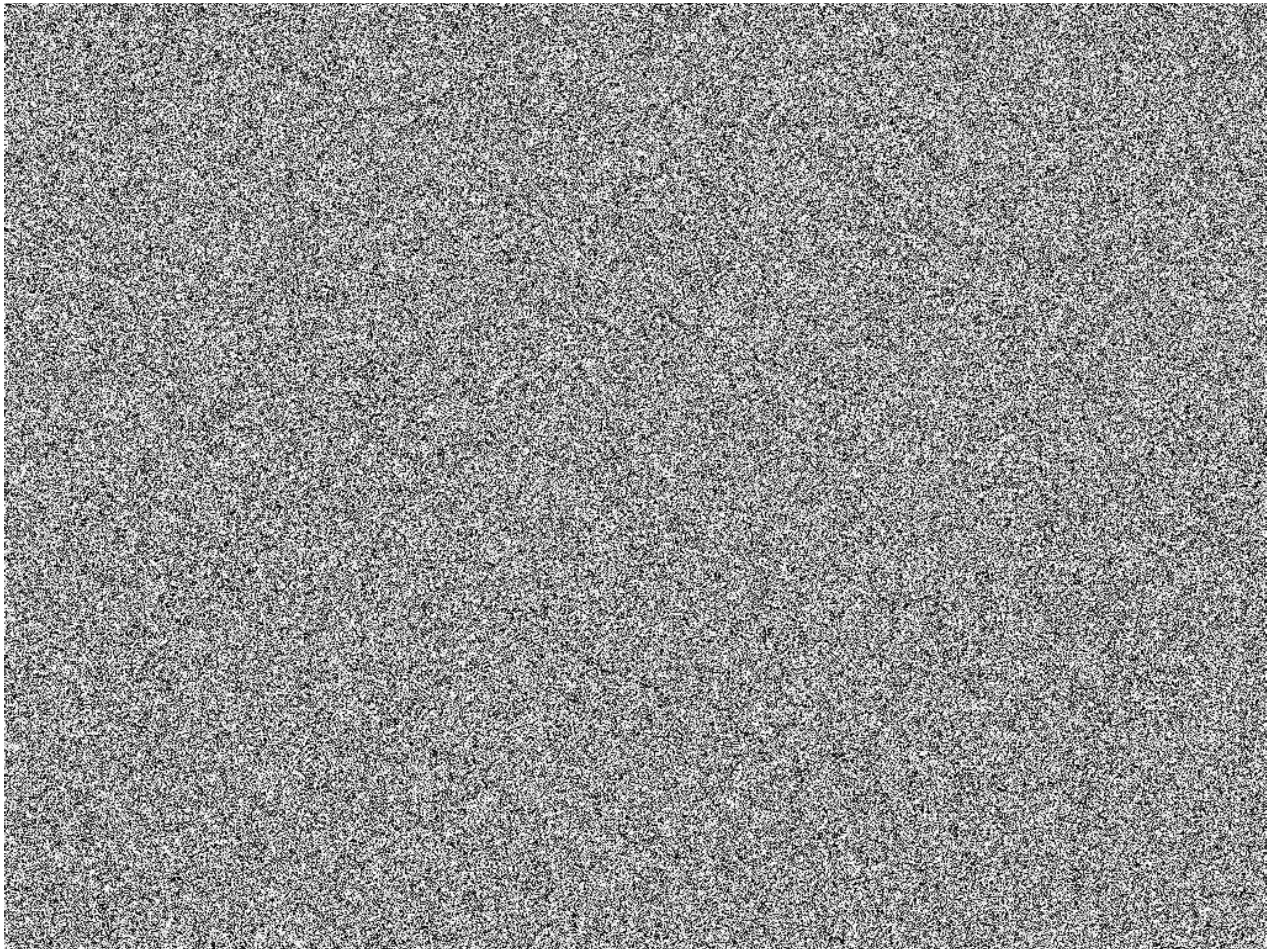
animal or not?

Example #3



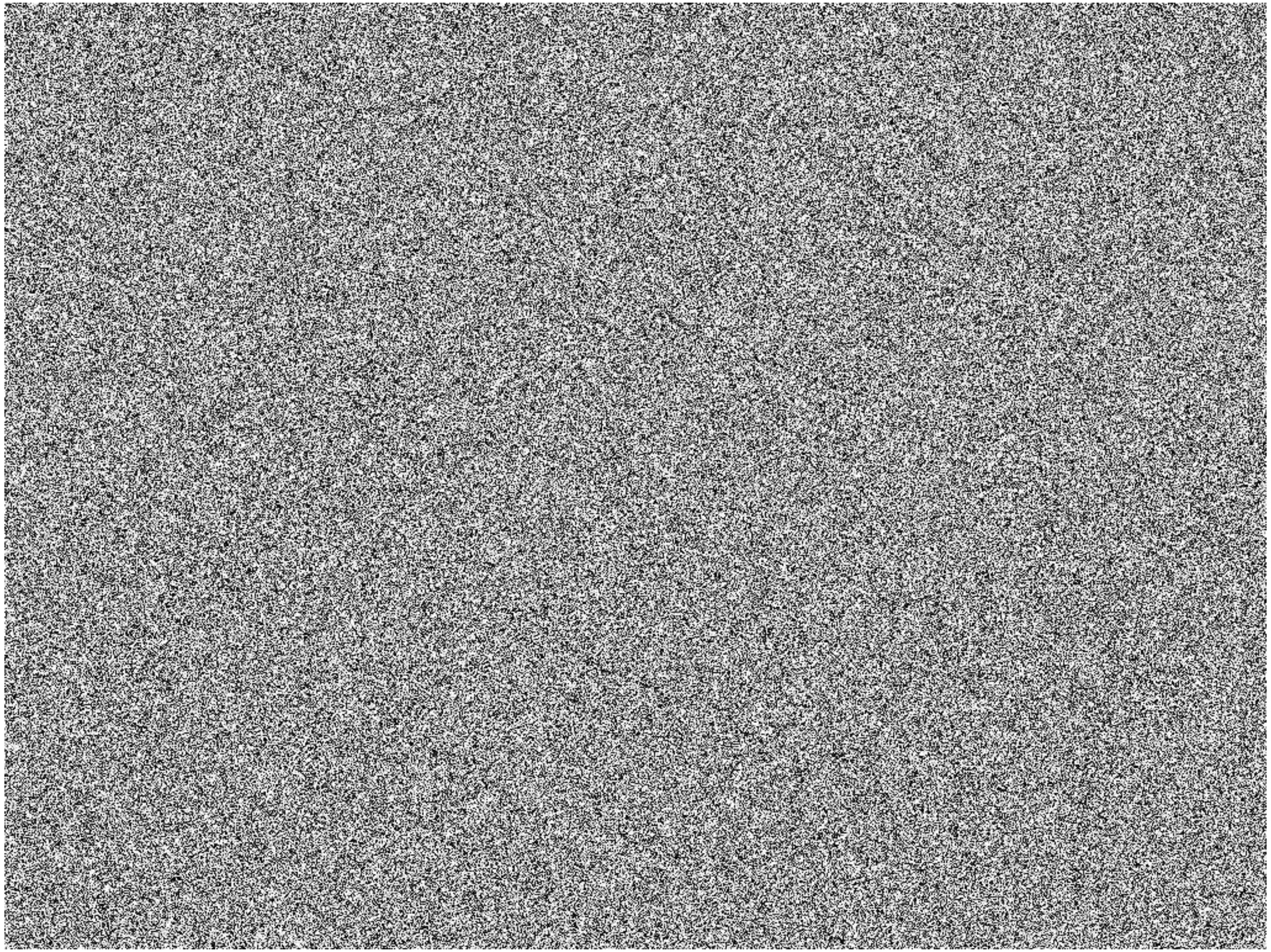
animal or not?

Example #4



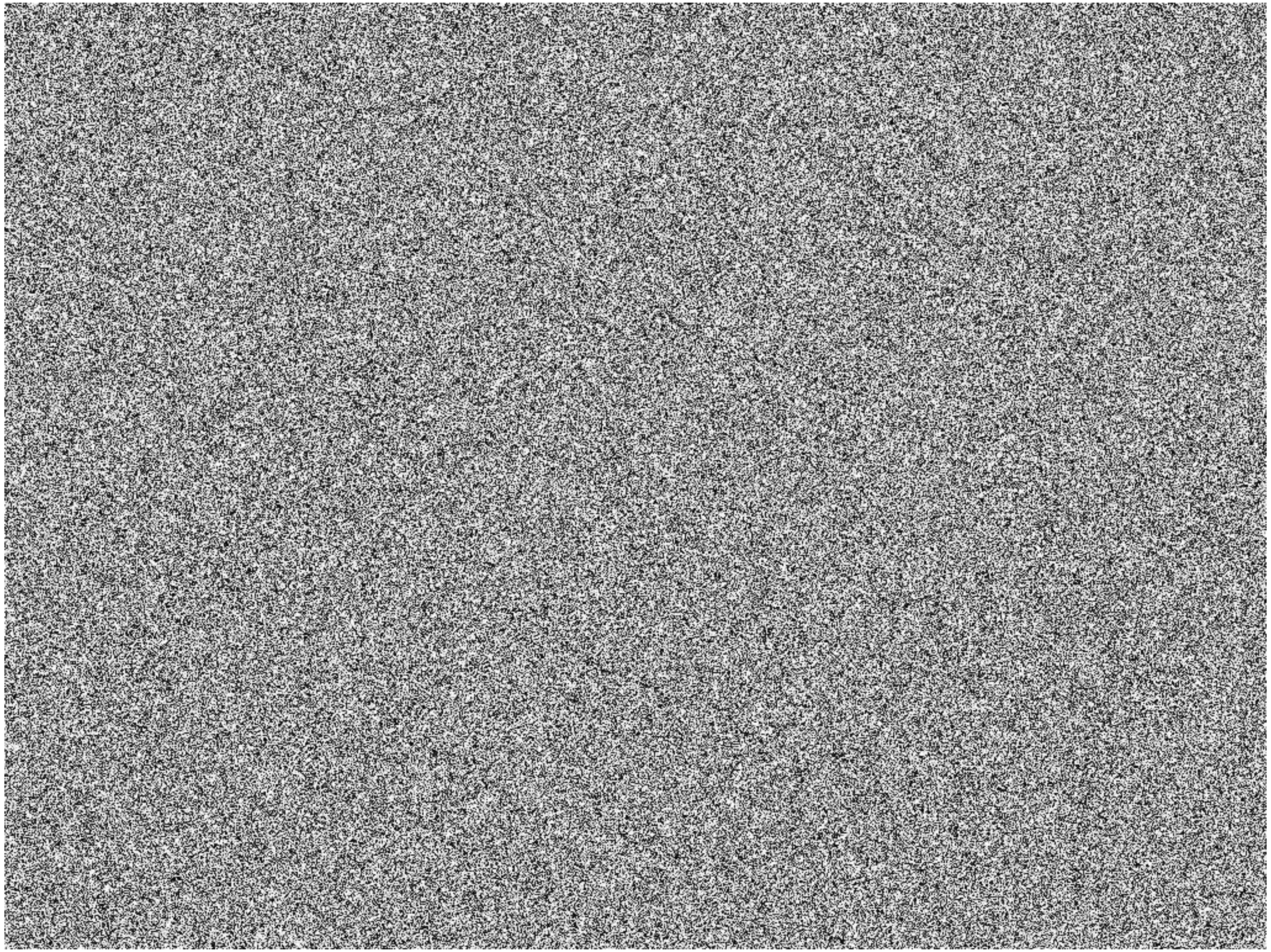
animal or not?

Example #5



animal or not?

Example #6



animal or not?

#1

#2

#3

#4

#5

#6

Human vision

Amazingly good, fast and accurate

Large amount of the brain seems to be for processing visual data

Vision is difficult!

LETTERS TO NATURE

Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

Centre de Recherche Cerveau & Cognition, UMR 5549, 31062 Toulouse, France

How long does it take for the human visual system to process a complex natural image? Subjectively, recognition of familiar objects and scenes appears to be virtually instantaneous, but measuring this processing time experimentally has proved difficult. Behavioural measures such as reaction times can be used[1], but these include not only visual processing but also the time required for response execution. However, event-related potentials (ERPs) can sometimes reveal signs of neural processing well before the motor output[2]. Here we use a go/no-go categorization task in which subjects have to decide whether a previously unseen photograph, flashed on for just 20 ms, contains an animal. ERP analysis revealed a frontal negativity specific to no-go trials that develops roughly 150 ms after stimulus onset. We conclude that the visual processing needed to perform this highly demanding task can be achieved in under 150 ms.

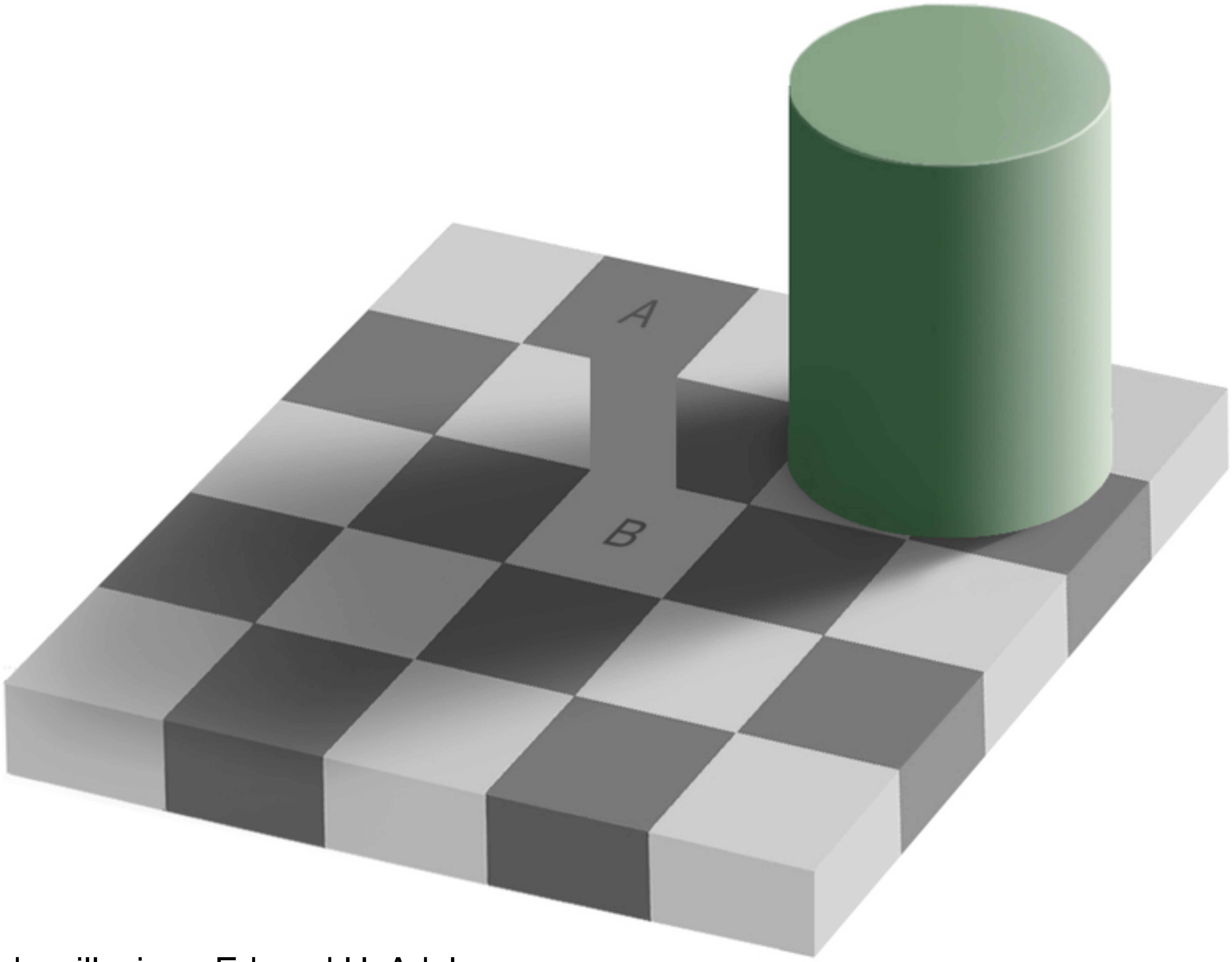
Neurophysiological measurements of the latencies of selective

such a task (the subjects had no *a priori* information about the type of animal to look for, its position or size, or even the number of animals present), performance was remarkably good. The average proportion of correct responses was 94%, with one of the fifteen subjects achieving 98% correct responses. The median reaction times on 'go' trials was 445 ms, although this value varied considerably between subjects, from a minimum of 382 ms to as much as 567 ms (Fig. 1). This remarkable level of performance was possible despite the very brief presentations, which effectively rule out the use of eye movements during image processing.

Whereas the behavioural reaction times put an upper limit on the time required for visual processing, the analysis of event-related potentials provided a much stronger constraint. By comparing average brain potentials generated on correct 'go' trials with those generated on correct 'no-go' trials, we were able to demonstrate that the two potentials diverge very sharply at ~150 ms after stimulus onset. The effect was particularly clear at frontal recording sites, and was characterized by a nearly linear increase in the voltage difference over the following 50 ms or so, the potential being more negative on no-go trials (Fig. 2). All 15 subjects showed the effect (Fig. 3), and although the onset latency varied somewhat between subjects, the differences were very minor compared with the very large differences in behavioural reaction times. Furthermore, there was no correlation whatsoever between behavioural reaction time and the onset latency for the differential response. This makes it unlikely that the differential

Source: A. Berg

We make mistakes …



Checker shadow illusion - Edward H. Adelson

Some examples of successful computer vision applications

(and cautionary tales)

Face recognition

https://hackercombat.com/free-facial-recognition-tool-to-track-people-on-social-media-sites/

https://osxdaily.com/2017/11/10/can-use-iphone-x-without-face-id/

The New York Times

San Francisco Bans Facial Recognition Technology

Attendees interacting with a facial recognition demonstration at this year's CES in Las Vegas. Joe Buglewicz for The New York Times

By Kate Conger, Richard Fausset and Serge F. Kovaleski

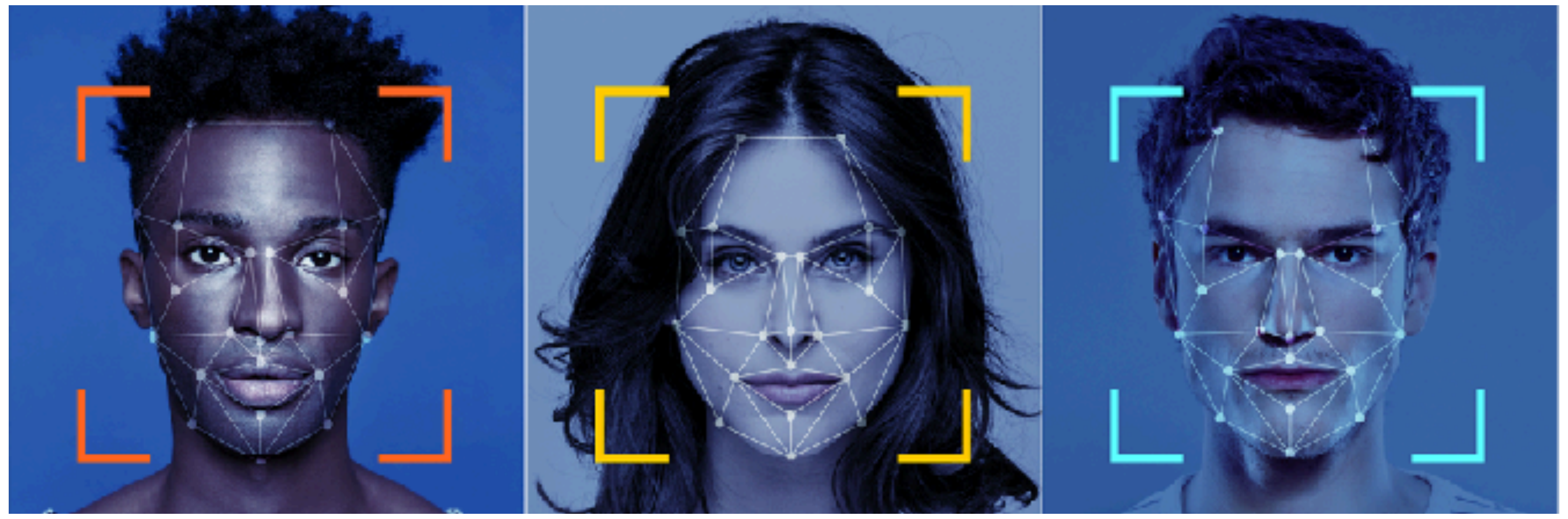
May 14, 2019

SAN FRANCISCO — San Francisco, long at the heart of the technology revolution, took a stand against potential abuse on Tuesday by banning the use of facial recognition software by the police and other agencies.

Boston Bans Use Of Facial Recognition Technology. It's The 2nd-Largest City To Do So

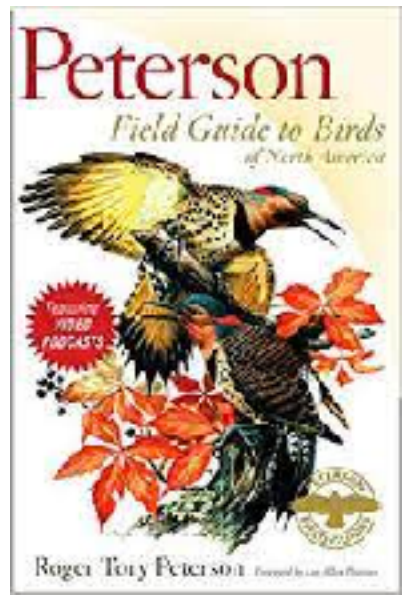
Updated June 24, 2020 By Ally Jarmanning

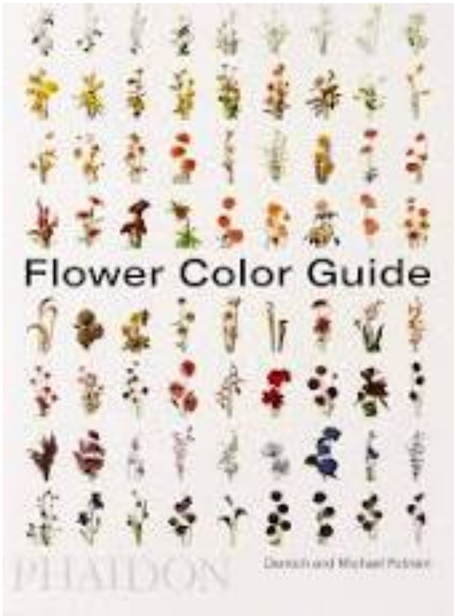
A close-up of a police facial recognition camera in use at the Cardiff City Stadium in Cardiff, Wales. (Matthew Horwood/Getty Images)

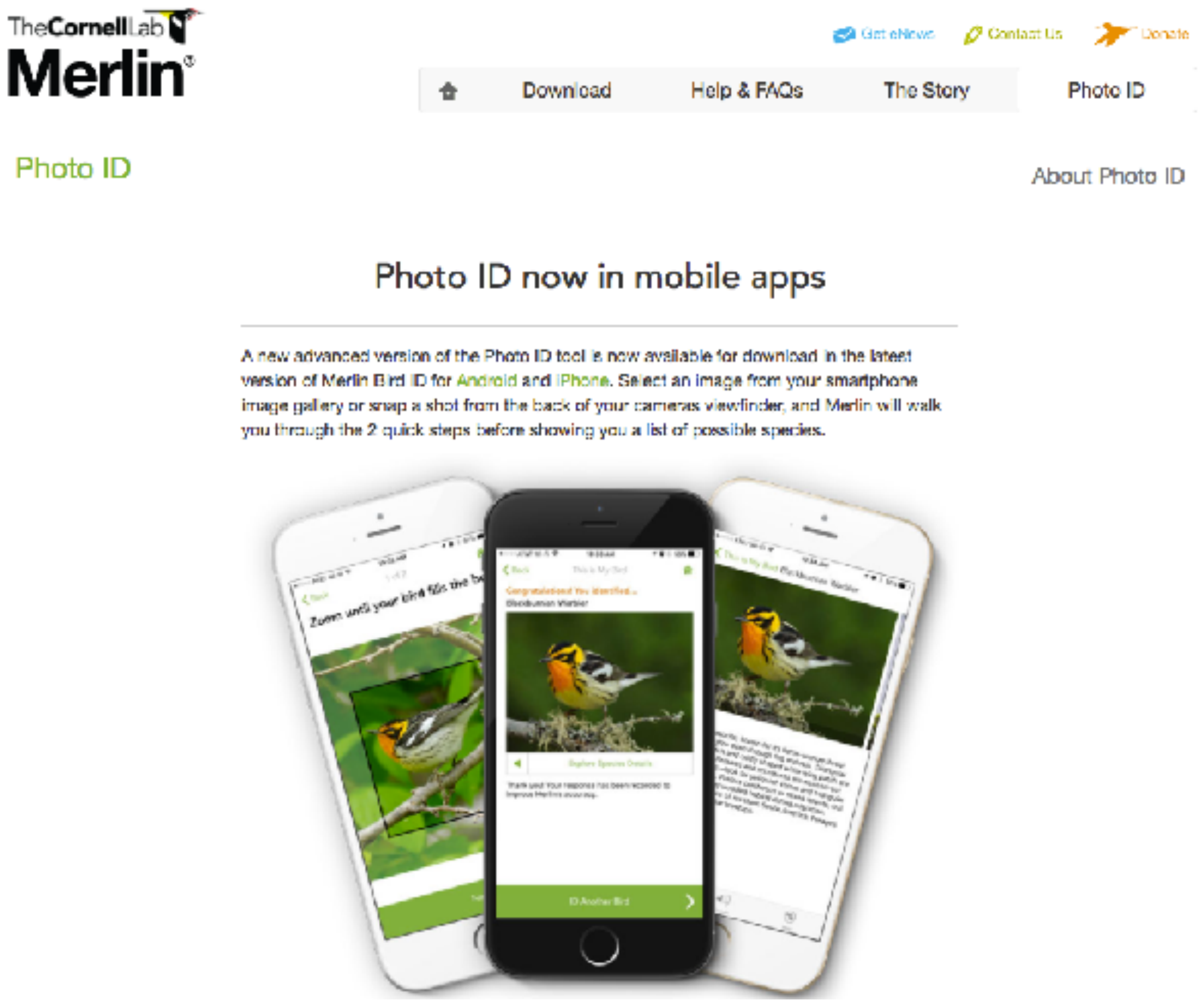


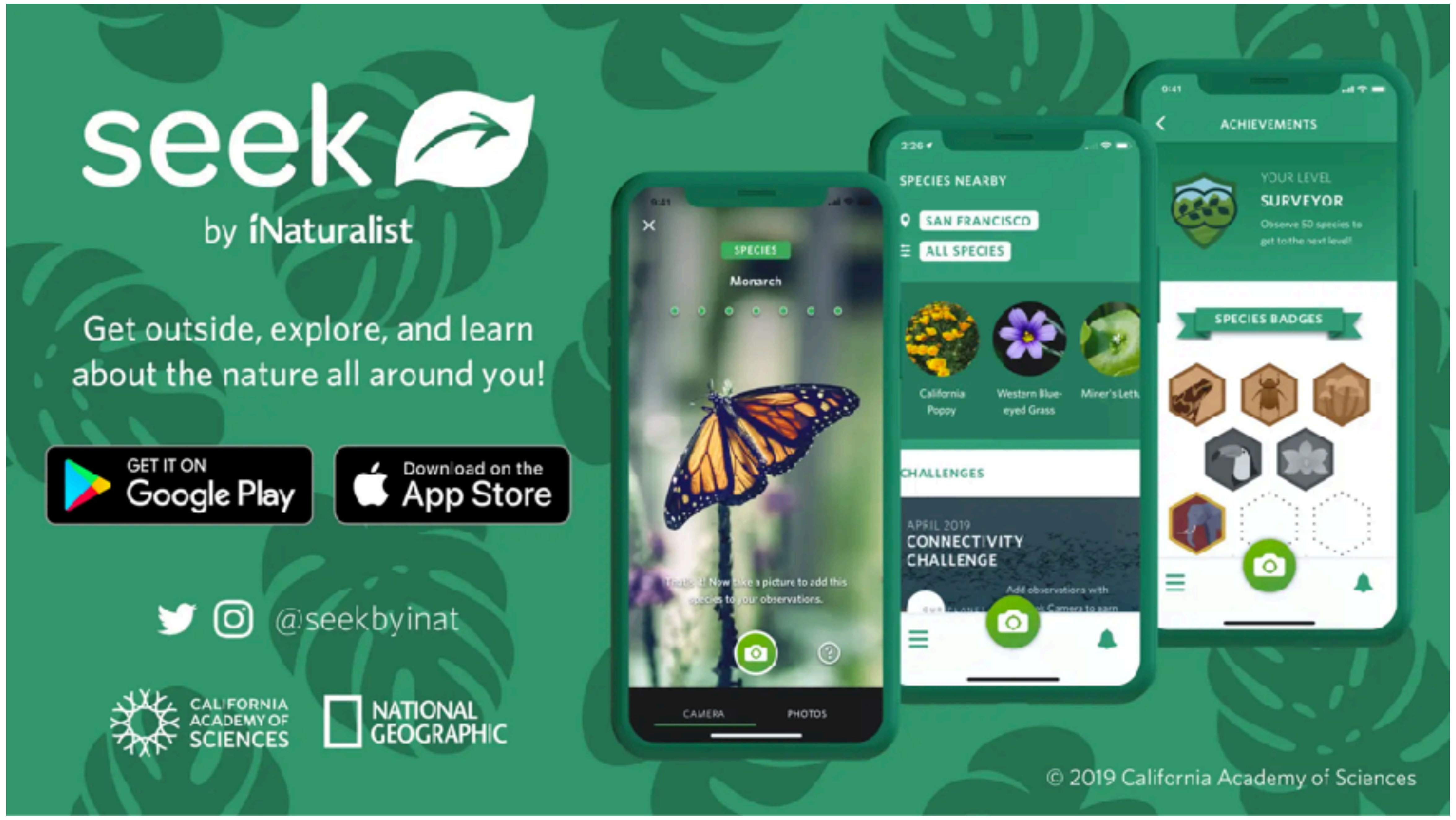
https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/

Electronic field guides









Electronic field guides

FGVC7 Home Competitions Submission Program Organizers

FGVC7 Competitions

iWildCam2020

Identify different species of animals in camera trap images.

https://www.kaggle.com/c/iwildcam-2020-fgvc7

Plant Pathology Challenge

Classify images of diseased plants.

https://www.kaggle.com/c/plant-pathology-2020-fgvc7

Semi-Supervised Fine-Grained Recognition Challenge

Semi-supervised classification of birds images.

https://www.kaggle.com/c/semi-inat-2020

Herbarium 2020 Challenge

Identify plant species from a large, long-tailed, collection of herbarium specimens.

https://www.kaggle.com/c/herbarium-2020-fgvc7

iMat Fashion2020

Apparel instance segmentation and fine-grained attribute classification.

https://www.kaggle.com/c/imaterialist-fashion-2020-fgvc7

iMet2020

Fine-grained attributes classification of works of art.

https://www.kaggle.com/c/imet-2020-fgvc7

https://sites.google.com/view/fgvc7/

Google AI Blog

The latest news from Google AI

Announcing the 7th Fine-Grained Visual Categorization Workshop

Wednesday, May 20, 2020

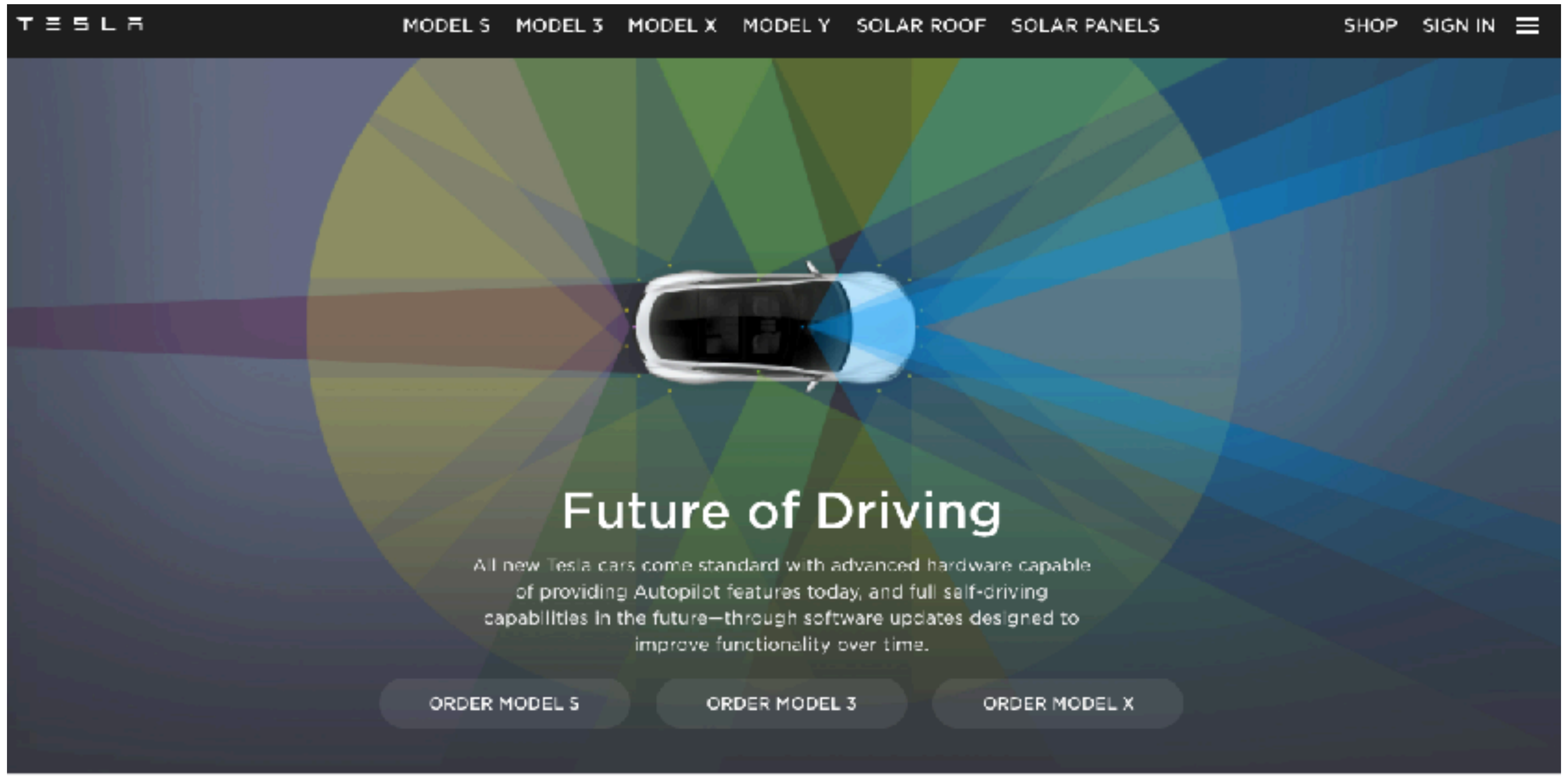
Posted by Christine Kaeser-Chen, Software Engineer and Serge Belongie, Visiting Faculty, Google Research

Fine-grained visual categorization refers to the problem of distinguishing between images of closely related entities, e.g., a monarch butterfly (*Danaus plexippus*) from a viceroy (*Limenitis archippus*). At the time of the first FGVC workshop in 2011, very few fine-grained datasets existed, and the ones that were available (e.g., the CUB dataset of 200 bird species, launched at that workshop) presented a formidable challenge to the leading classification algorithms of the time. Fast forward to 2020, and the computer vision landscape has undergone breathtaking changes. Deep learning based methods helped CUB-200-2011 accuracy rocket from 17% to 90% and fine-grained datasets have proliferated, with data arriving from a diverse array of institutions, such as art museums, apparel retailers, and cassava farms.

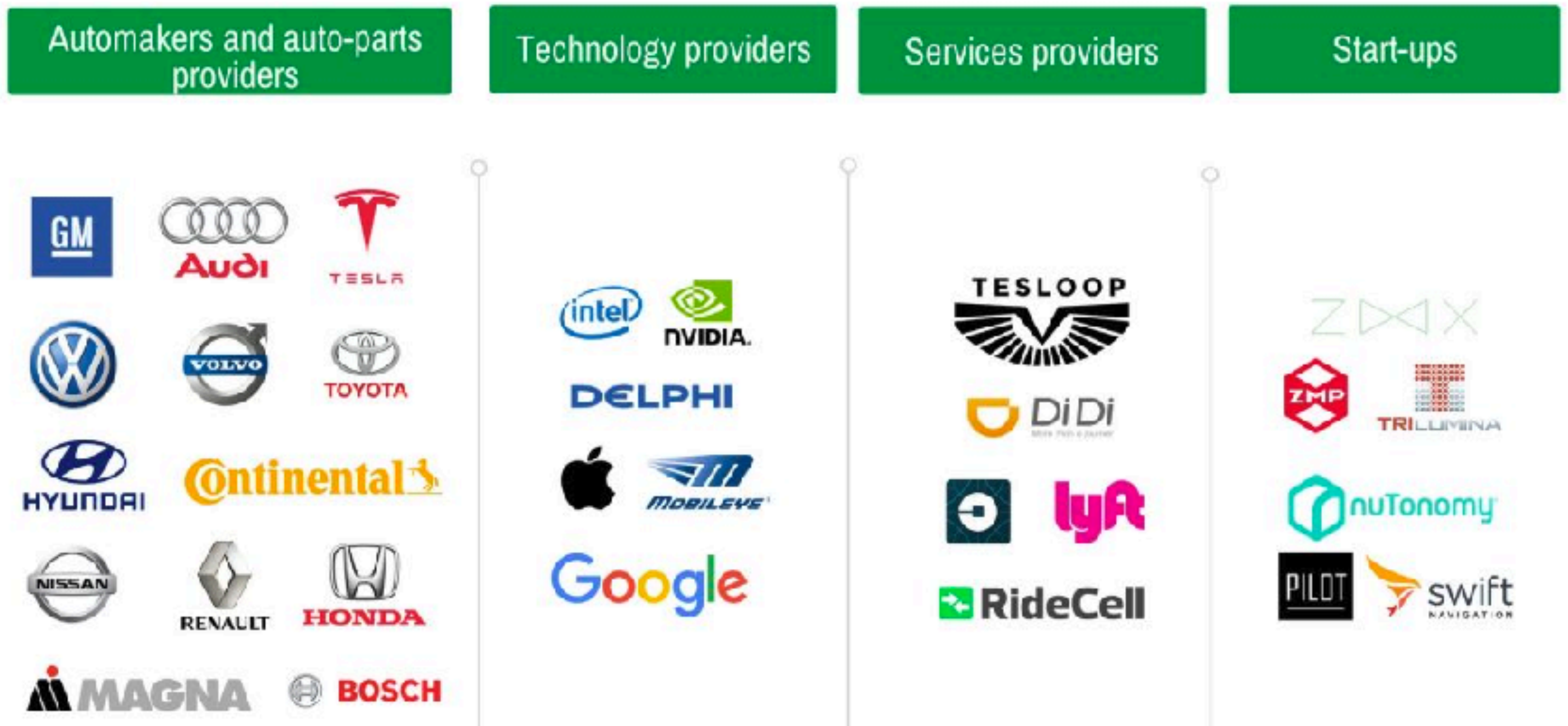
In order to help support even further progress in this field, we are excited to sponsor and co-organize the 7th Workshop on Fine-Grained Visual Categorization (FGVC7), which will take place as a virtual gathering on June 19, 2020, in conjunction with the IEEE conference on Computer Vision and Pattern Recognition (CVPR). We're excited to highlight this year's world-class lineup of fine-grained challenges, ranging from fruit tree disease prediction to fashion attributes, and we invite computer vision researchers from across the world to participate in the workshop.

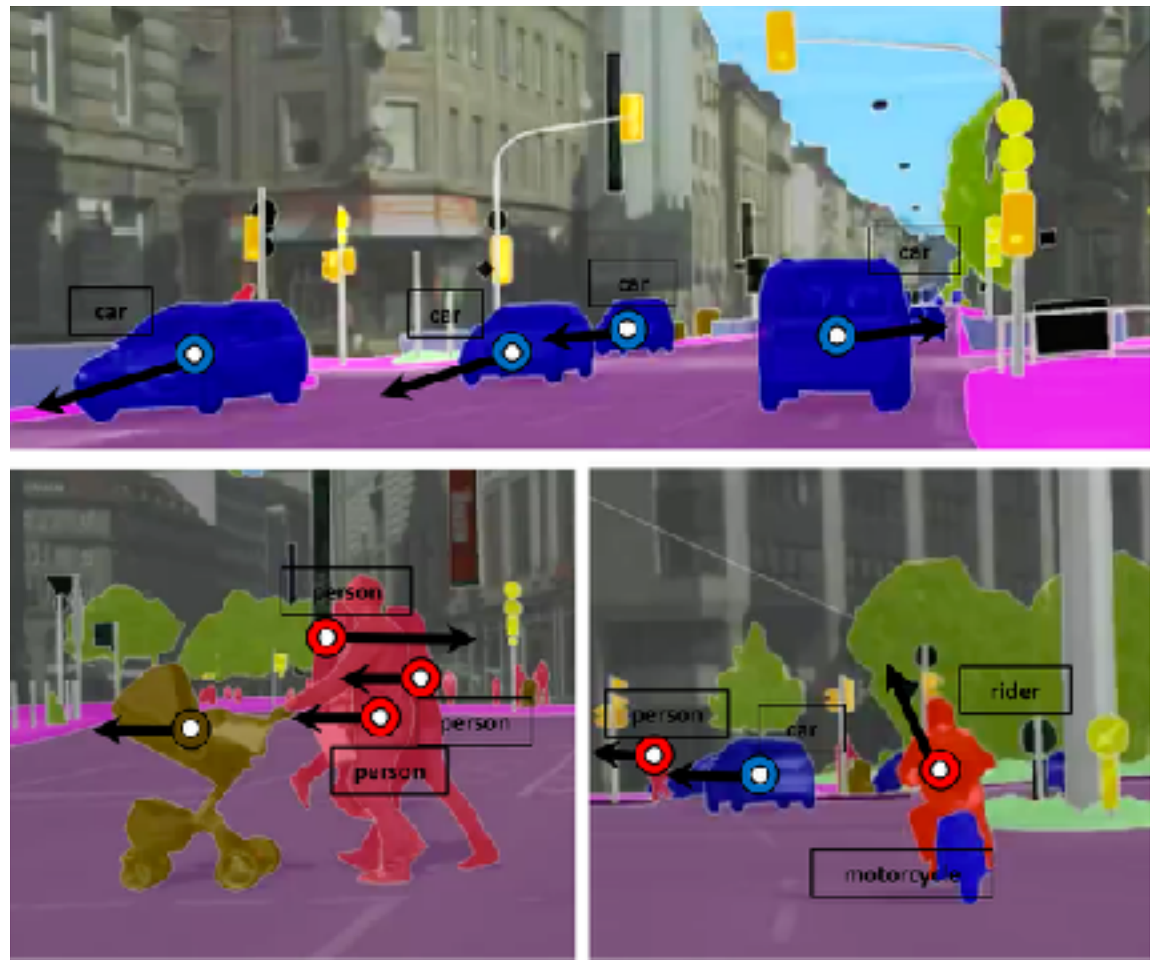
https://ai.googleblog.com/2020/05/announcing-7th-fine-grained-visual.html

Autonomous driving



https://www.tesla.com/autopilot





https://www.trafficsafetystore.com/blog/could-ford-lead-the-future-of-autonomous-cars/

https://hal.archives-ouvertes.fr/hal-01494296

Many others …

Industrial inspection
Sports analytics
Advertisement
Assistive technology
Product recognition and recommendation
Activity recognition
Emotion analysis
Scene text detection and recognition
Document analysis
Medical imaging and screening
…

source: http://cvpr2020.thecvf.com/sponsors

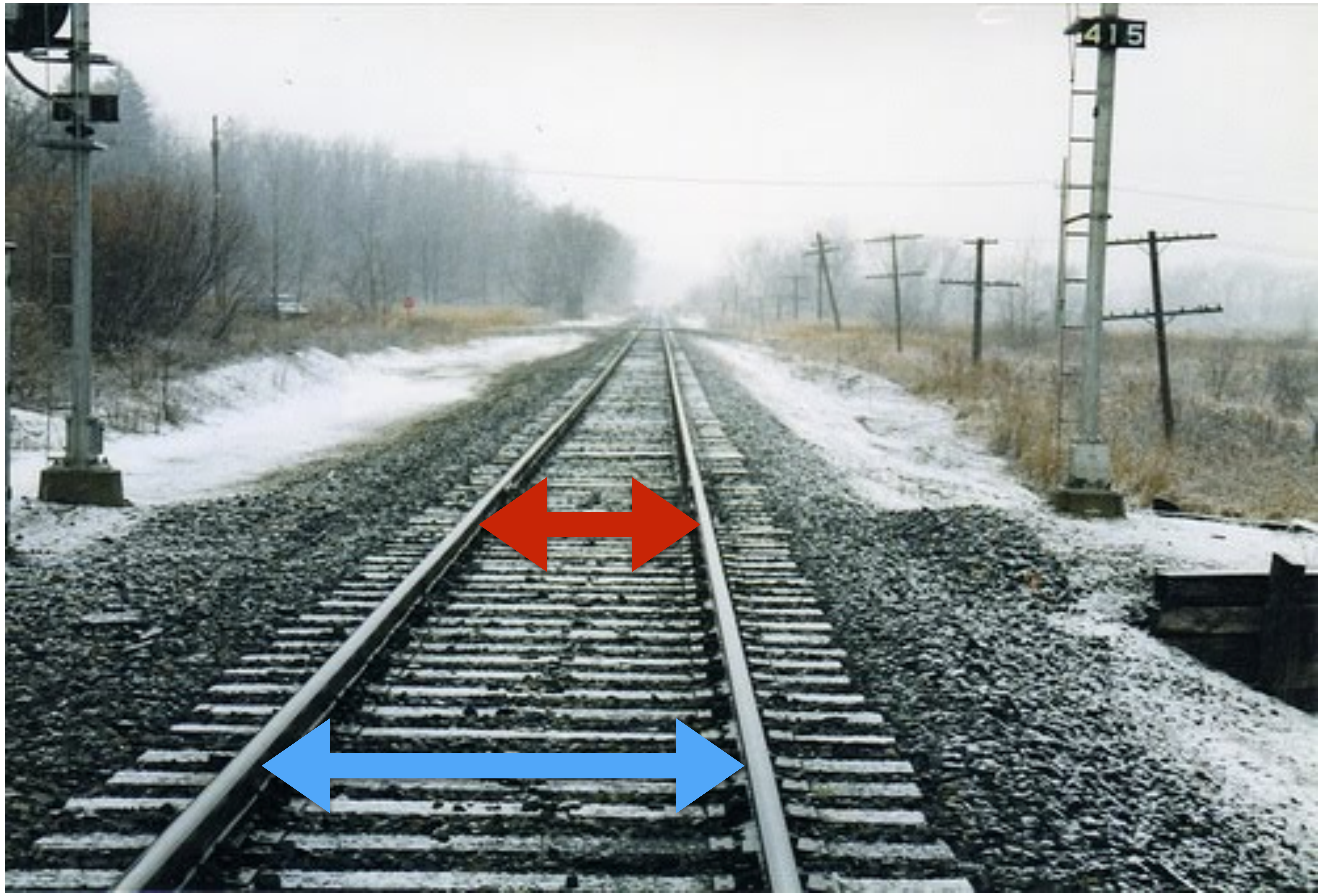
What tools do we have?

Many possibilities — how do we solve this ambiguity?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Slide credit: J. Koenderink

Tool 1: Physics and Geometry



http://kalisdigitalphotos.blogspot.com

Parallel lines merge at the horizon

Analyzing parallel lines to estimate space

Tool 1: Physics and Geometry

Scattering of skylight by particles in the air adds to the luminosity

Photo by **Éole Wind**

As the distance of the object from the viewer *increases*, the contrast between the object and its background *decreases*.

Tool 1: Physics and Geometry

Occlusions

Chicago loop, image source: **wikipedia**

Tool 1: Physics and Geometry

Light and shading

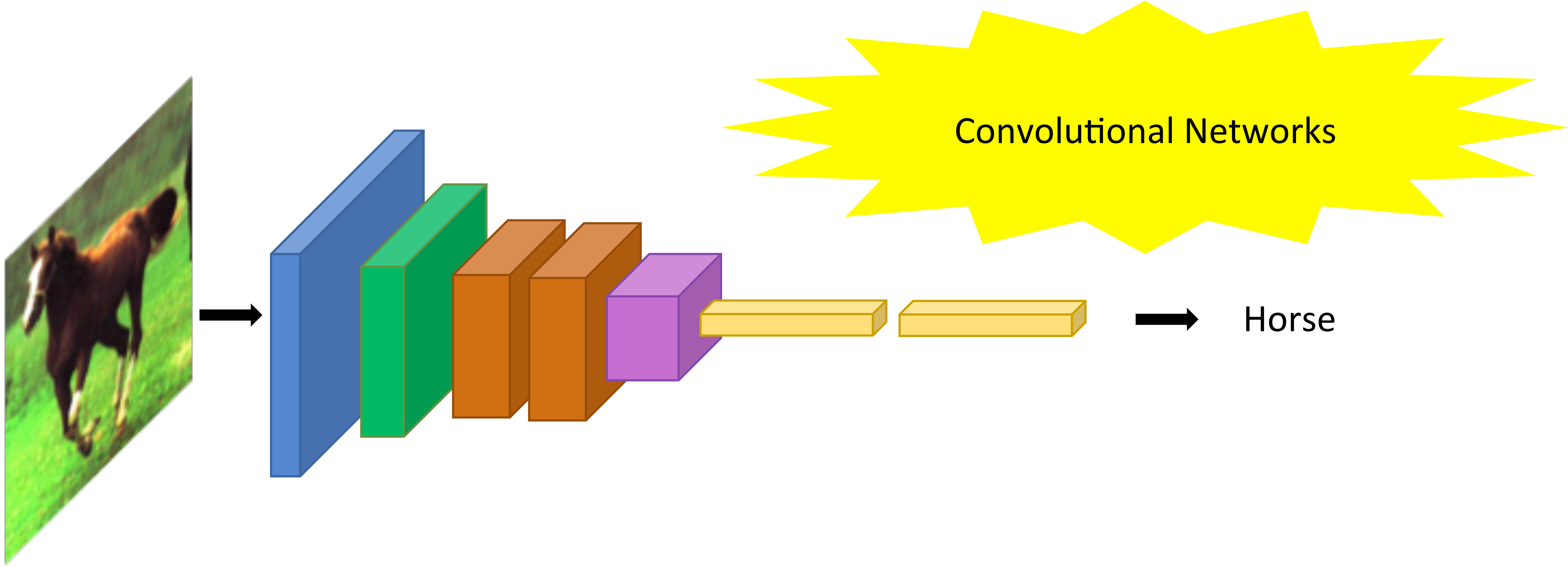
“The four seasons” sculpture set

The Rathas of Mahabalipuram, India

Vision is hard

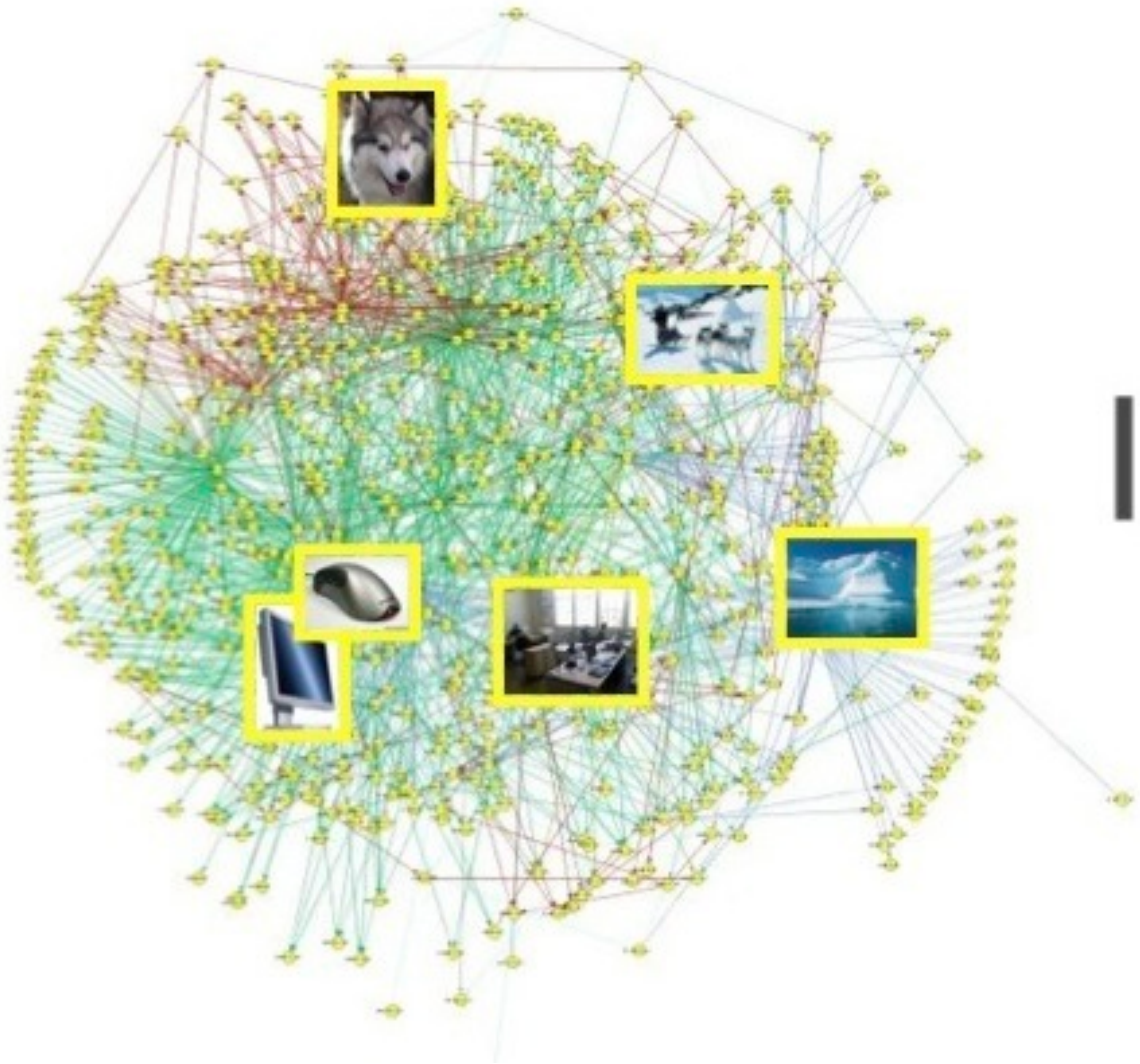
Tools from geometry and physics are often not sufficient

Tool 2: Data and machine learning



1. Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86.11 (1998): 2278-2324.
2. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS* 2012.

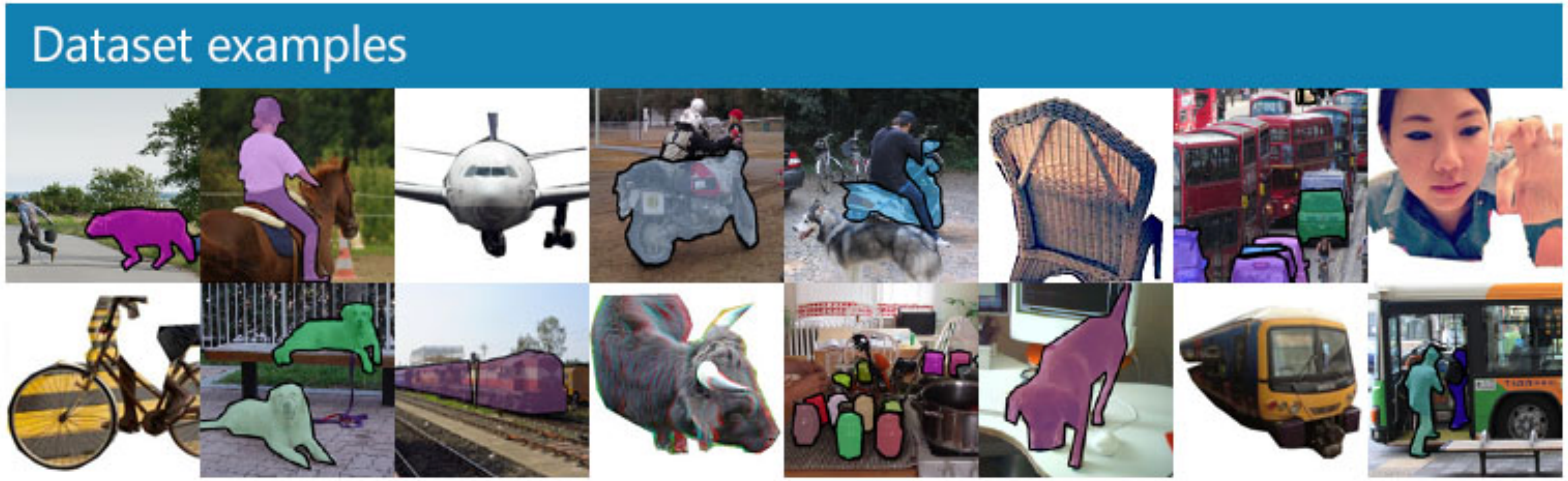
Tool 2: Data and machine learning



[Deng et al. CVPR 2009]

- 14+ million labeled images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk
- The challenge: 1.2 million training images, 1000 classes

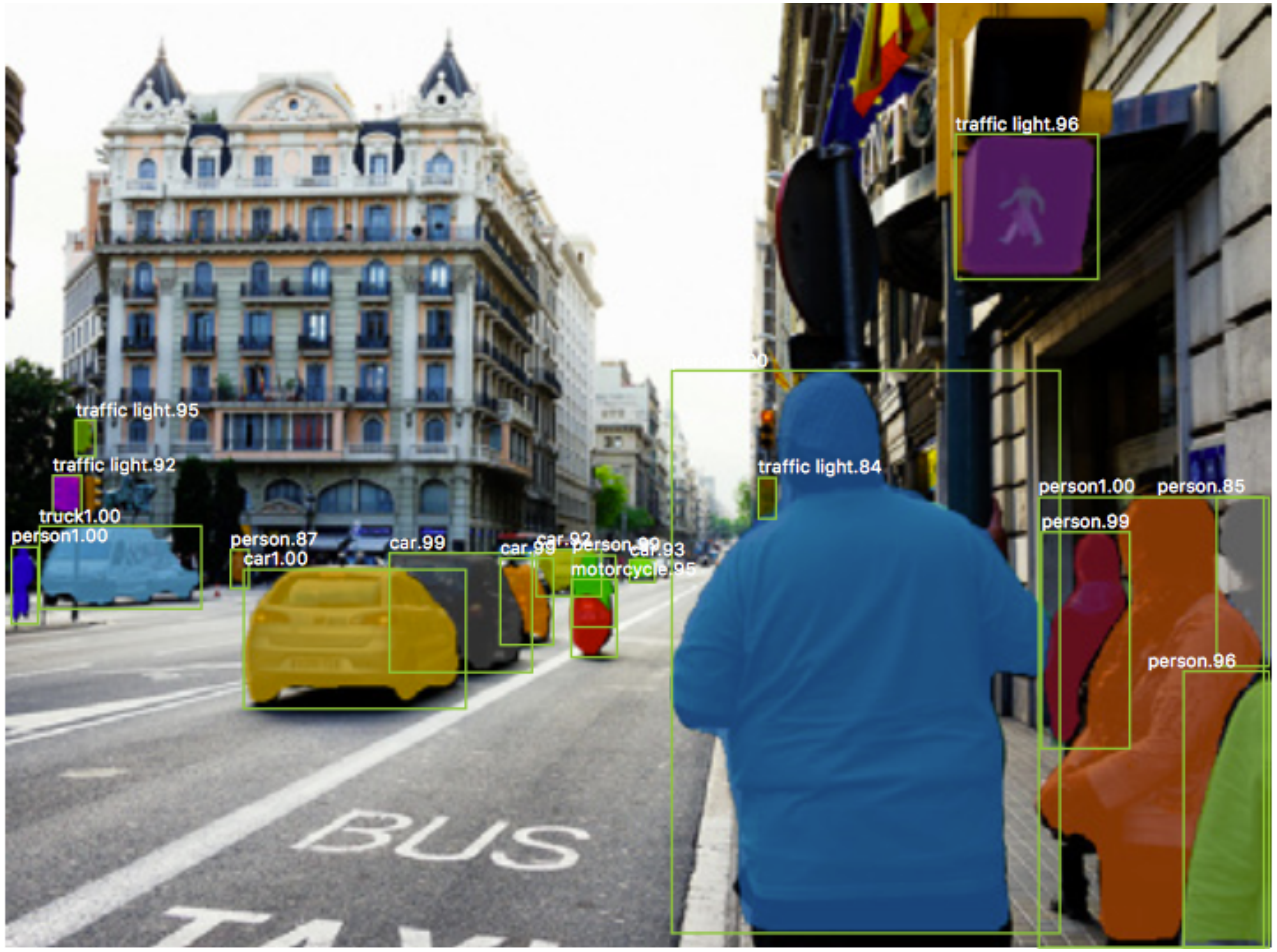
Tool 2: Data and machine learning



mscoco

Microsoft COCO: Common Objects in Context
Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár
ECCV, 2014

Tool 2: Data and machine learning



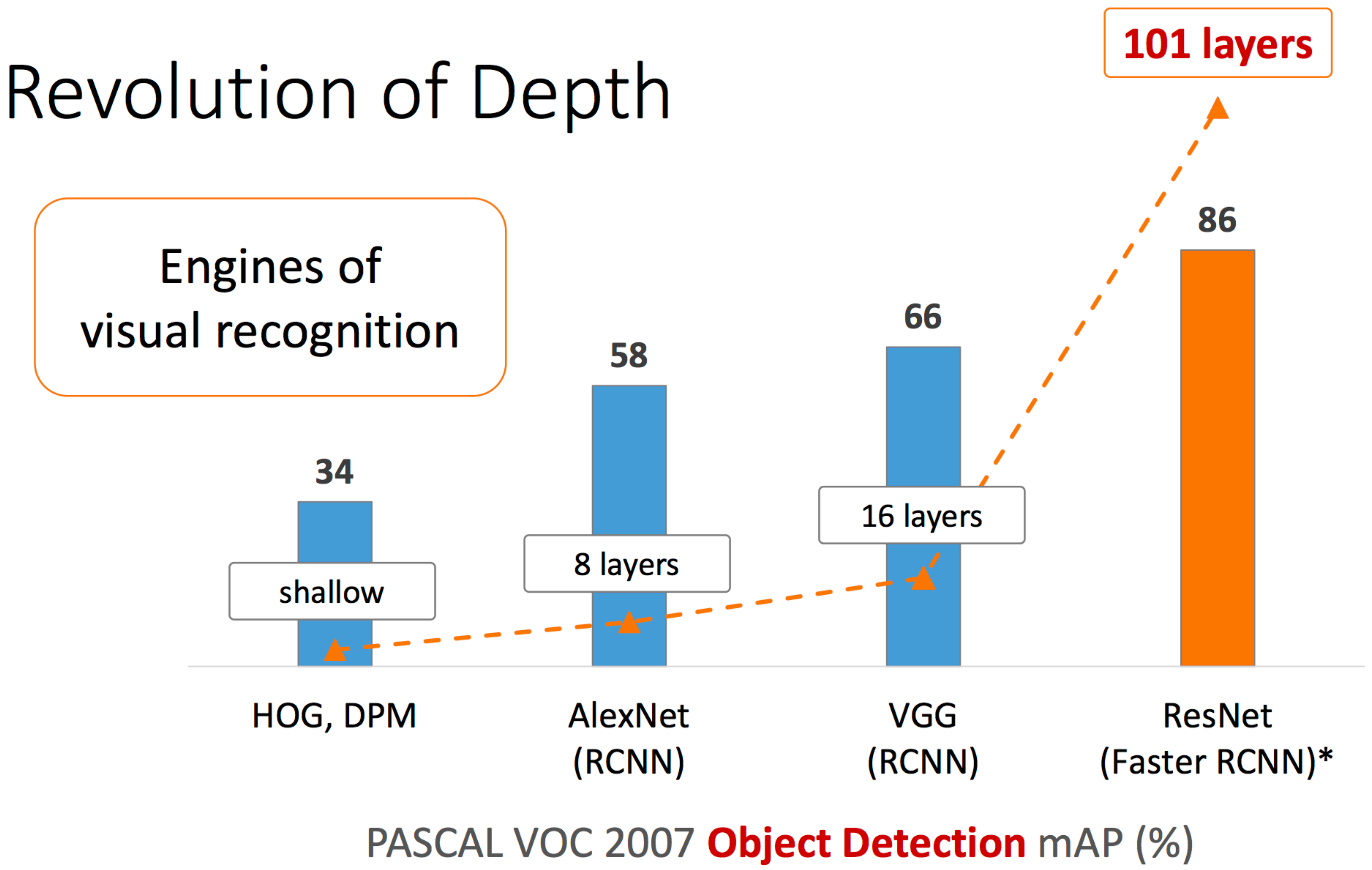
Mask R-CNN

Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick

(Submitted on 20 Mar 2017 (v1), last revised 5 Apr 2017 (this version, v2))

What next? — Improving recognition

Architectures for recognition



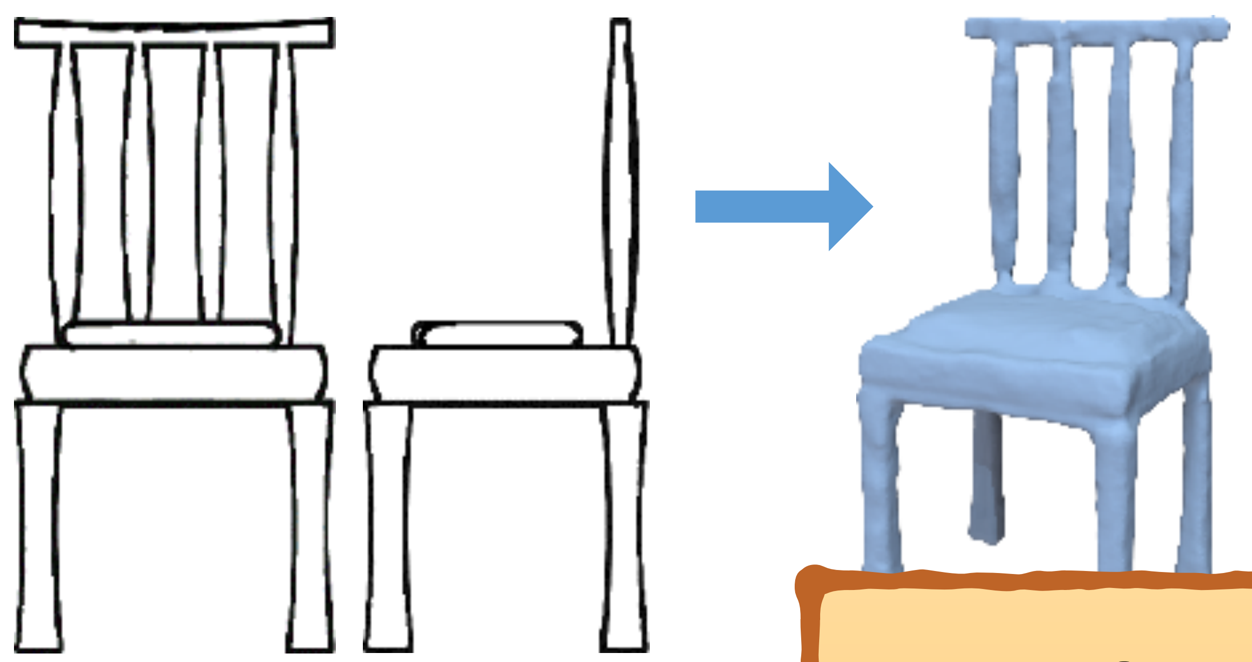
Deep Residual Learning for Image Recognition
Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. CVPR 2016.

Learning with less supervision

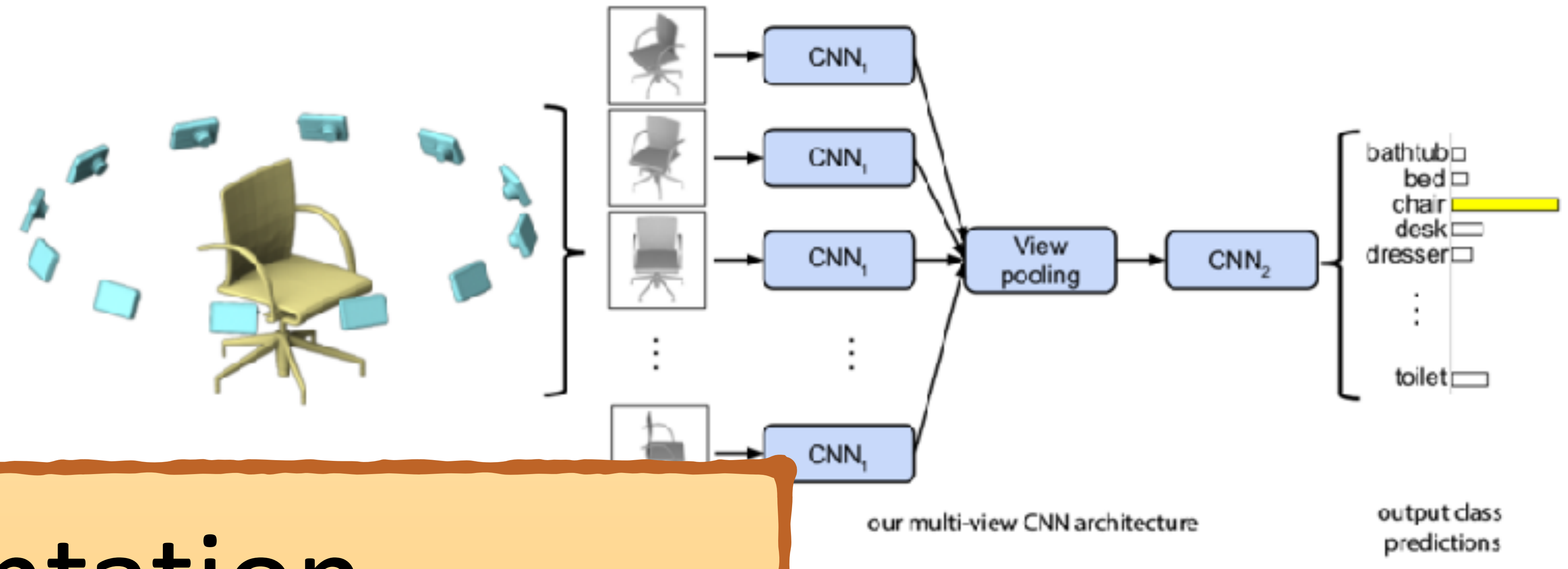
- Multi-tasking
- Transfer learning
- Domain adaptation
- Multi-modal data
- Theory

What next? — 3D shape understanding

Inferring 3D shape from a single image



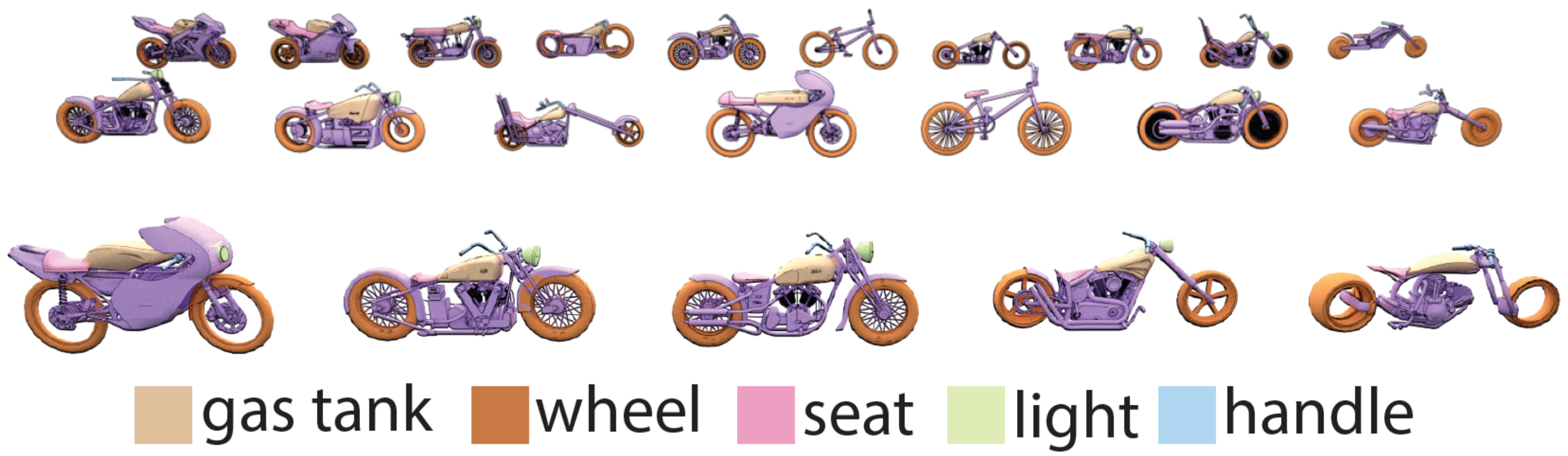
Categorizing or retrieving 3D shapes



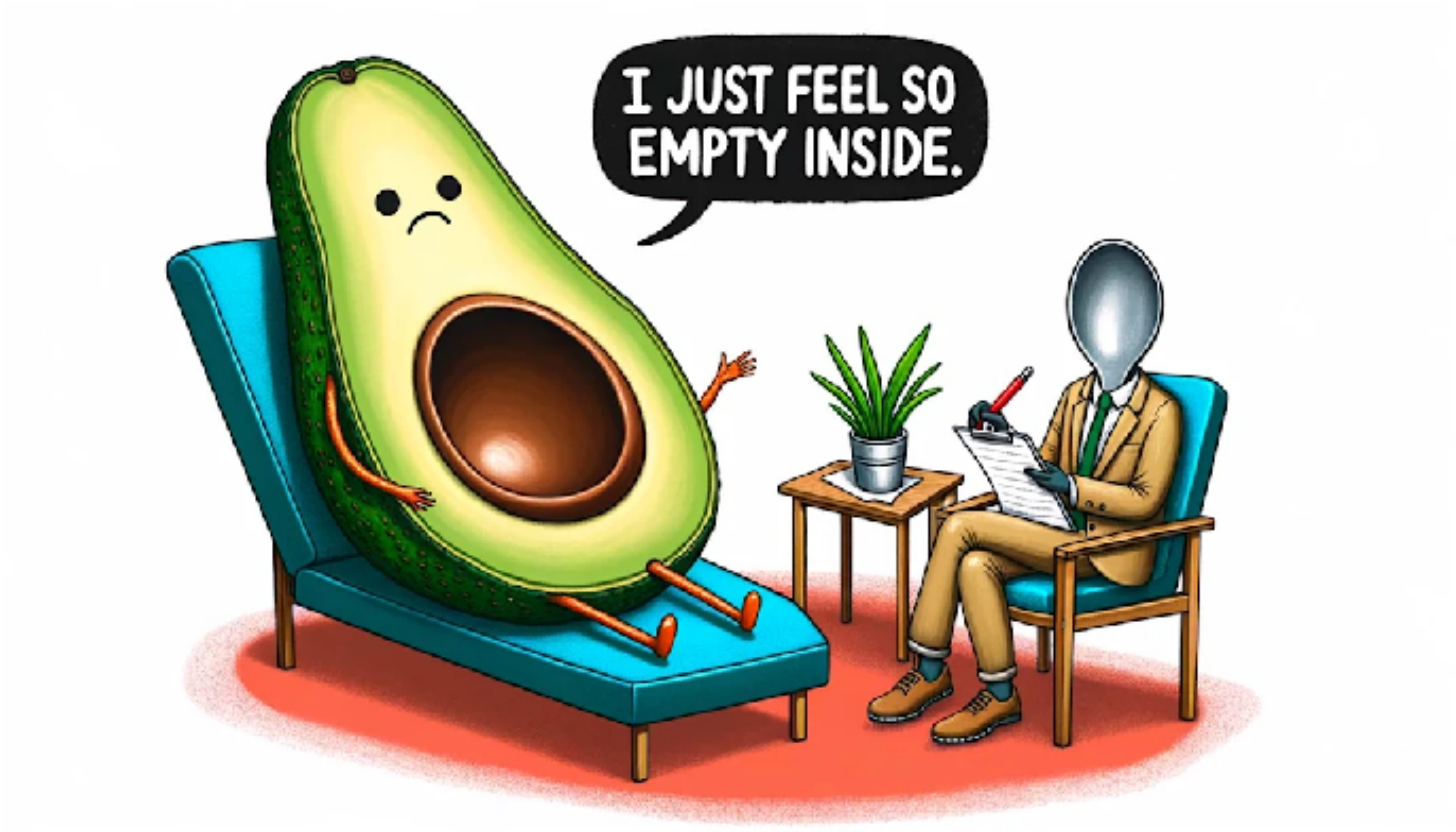
1. Choice of representation
2. Incorporate physics + geometry

Estimating 3D shape from image collections

Segmenting 3D shapes into parts



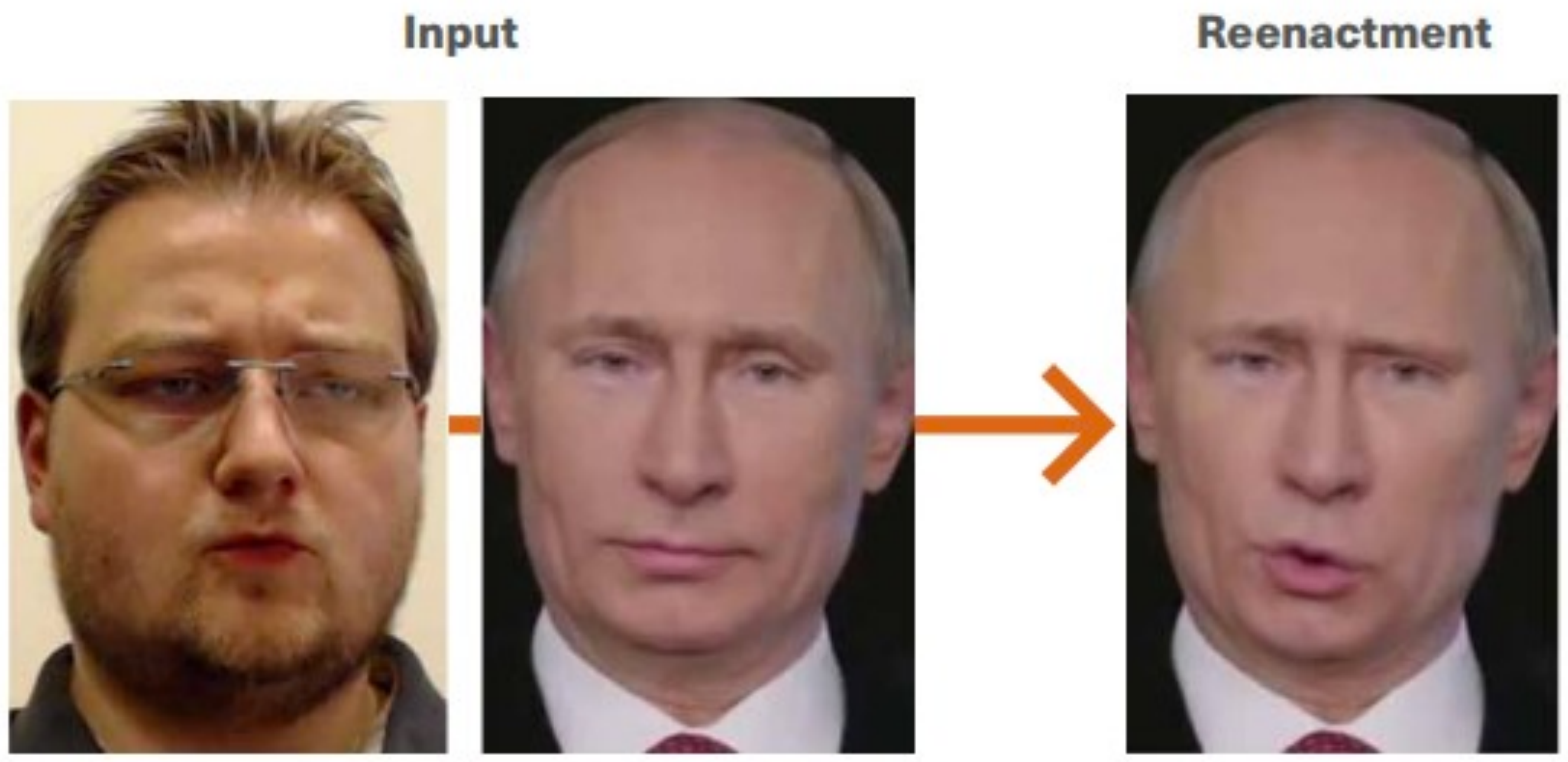
What next? — Better generative models



An illustration of an avocado sitting in a therapist's chair, saying 'I just feel so empty inside' with a pit-sized hole in its center. The therapist, a spoon, scribbles notes.

https://openai.com/dall-e-3/

But deepfakes …



Face2Face, Thies et al., 2016


The New York Times

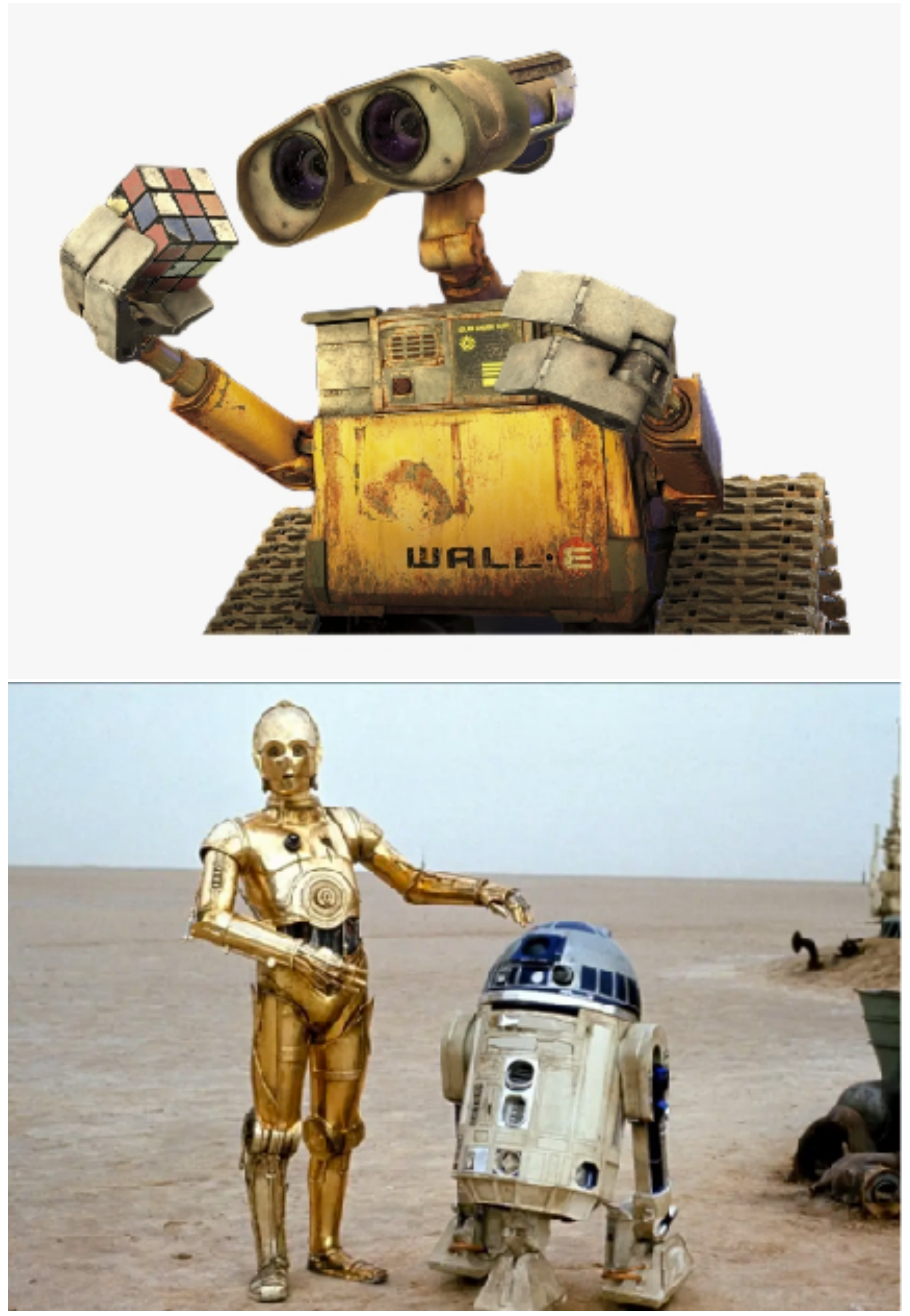
The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

What next? — better understanding of the world

Multi-modal understanding

Language, vision, sound



LucasFilm

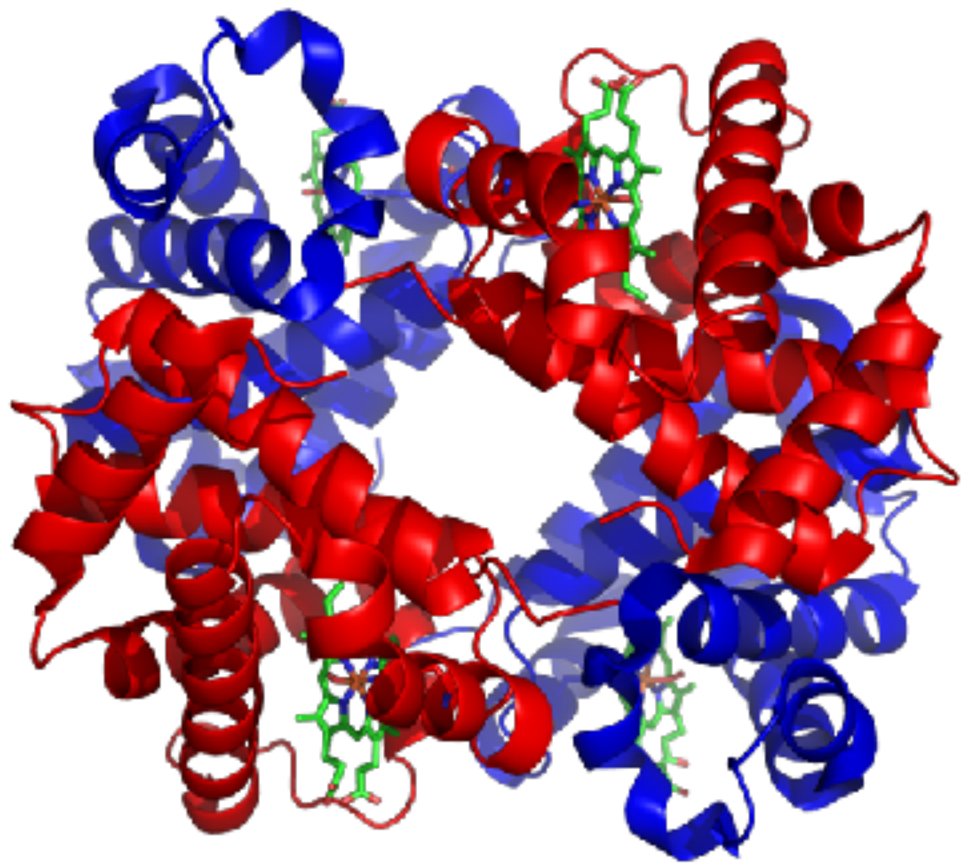
Embodied cognition

Solving planet-scale problems

Advance science

Amino acid sequence

AlphaFold



https://en.wikipedia.org/wiki/Hemoglobin

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › news › article

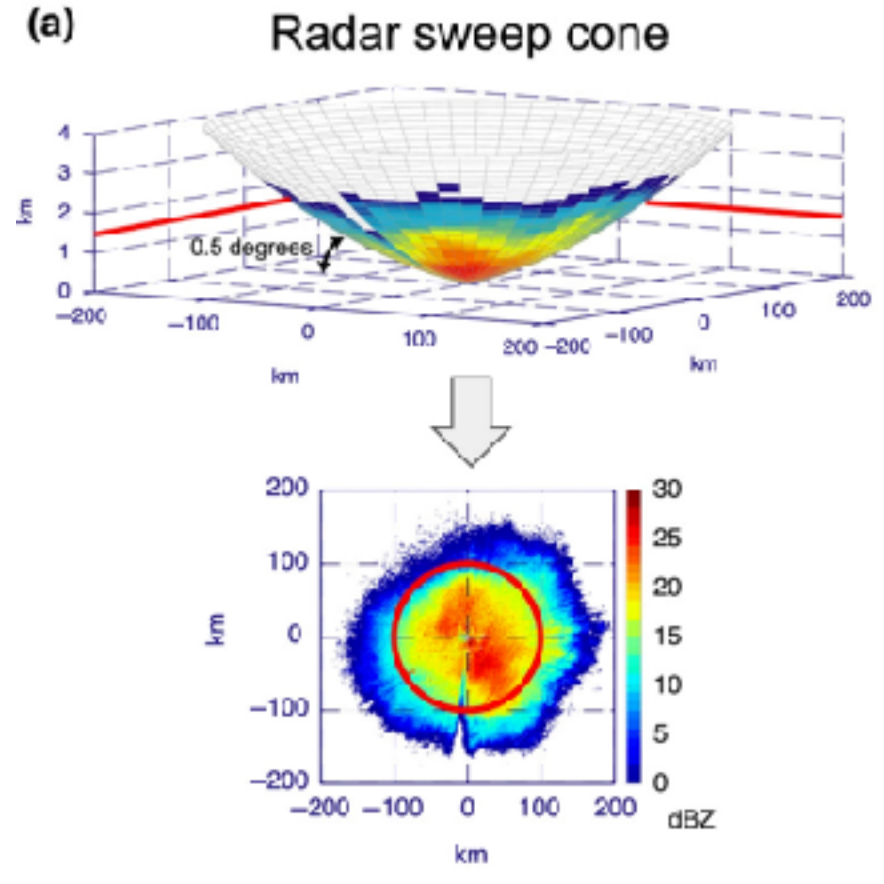
NEWS | 09 October 2024

Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures

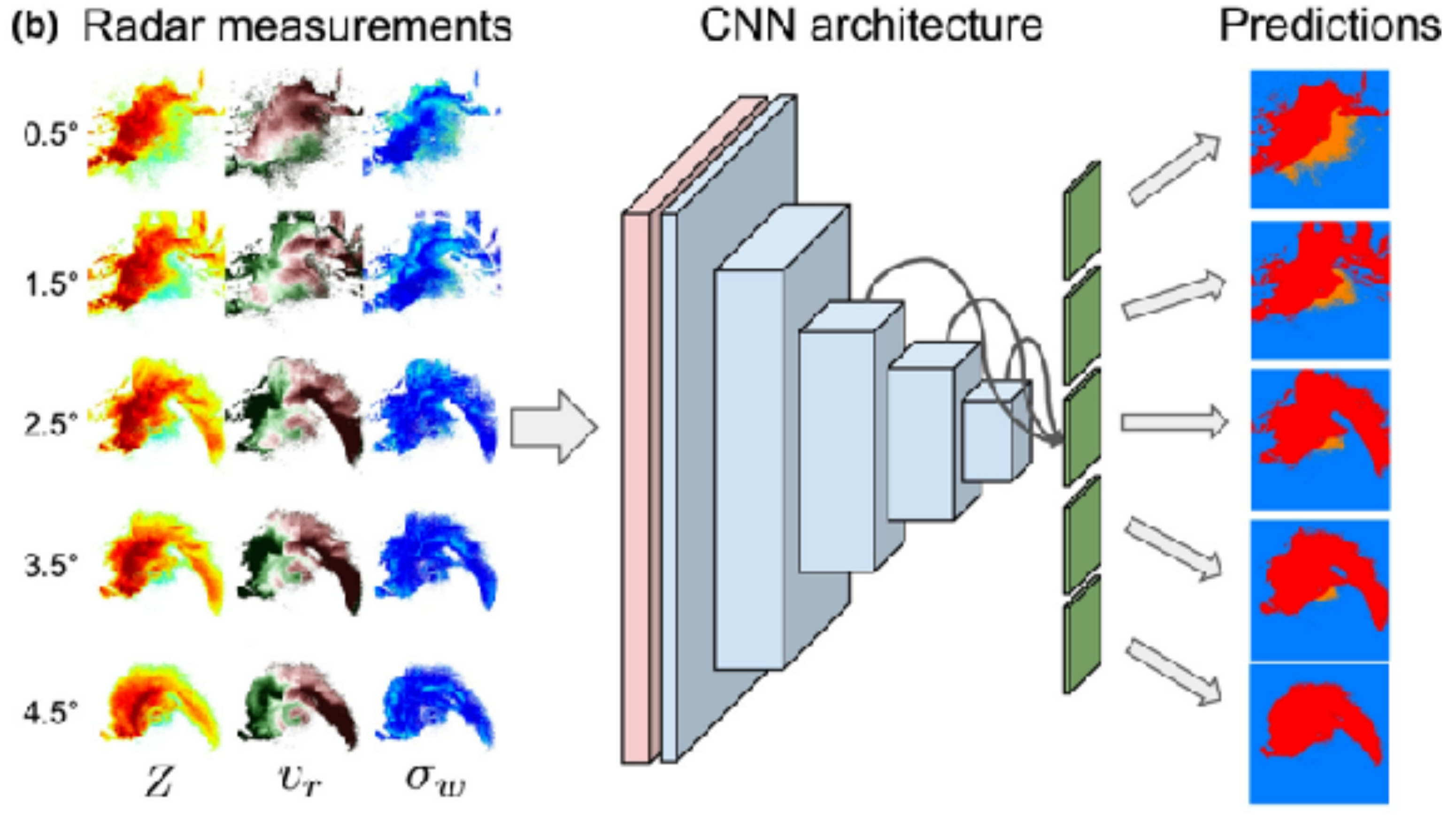
This year's prize celebrates computational tools that have transformed biology and have the potential to revolutionize drug discovery.

Advance science …

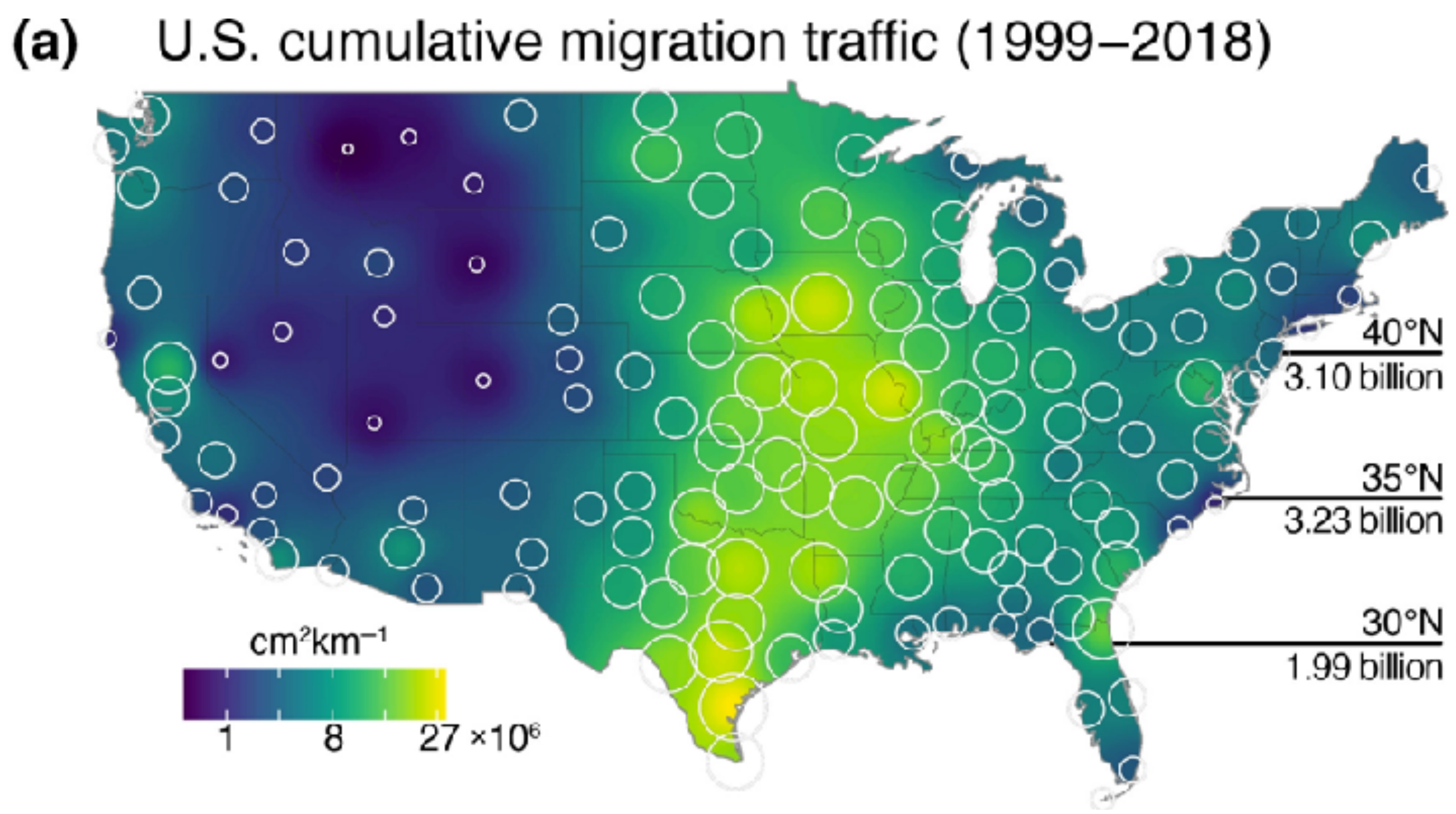
Measuring bird migration from RADAR data @ dark ecology project



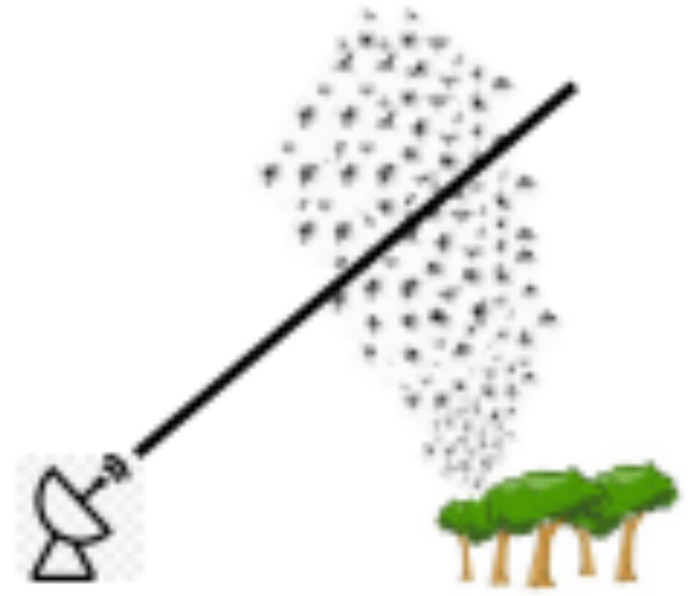
Radar geometry



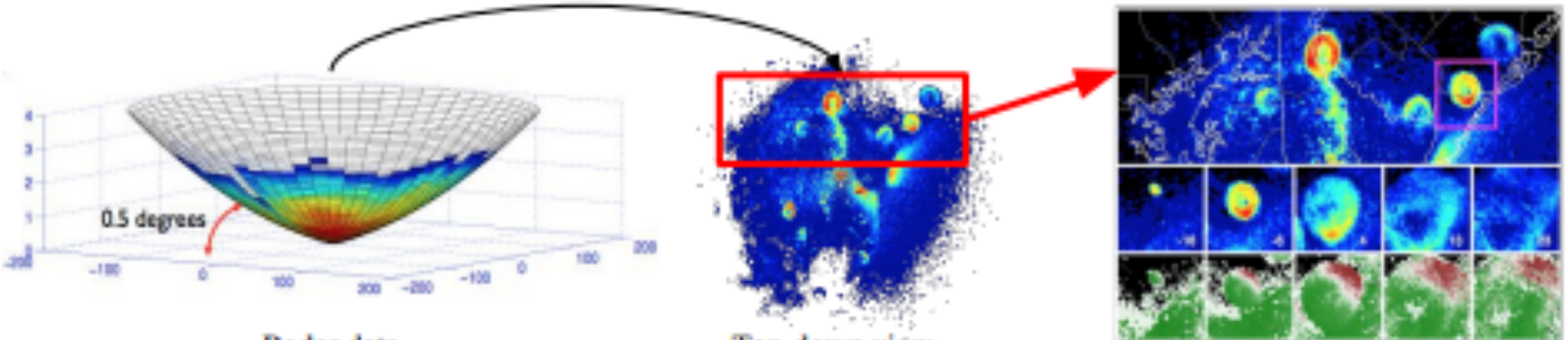
Overview of the processing pipeline



Spatial mapping



(a) Communal roost exodus



(b) Radar geometry and roost appearance

Welcome!