Introduction to Crowdsourcing for Computer Vision

Danna Gurari

The University of Texas at Austin

Fall 2019

https://www.ischool.utexas.edu/~dannag/Courses/CrowdsourcingForCV/CourseContent.html

Today’s Topics

- Class logistics
- Computer vision: past, present, & future
- Computer vision: what makes it hard?
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

Today’s Topics

- Class logistics
- Computer vision: past, present, & future
- Computer vision: what makes it hard?
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

Introductions

Instructor: Danna Gurari

Danna: pronounced like “Donna”

Gurari: rhymes with Ferrari

Pronouns: she/her/hers

Interdisciplinary class:

- introduce yourself
- share about your career aspirations

Q&A: “Do I have the appropriate pre-requisites/background?”

- Yes. There are no pre-requisites.

- You will be expected to further develop skills we cover in class **on your own**; e.g.,
  - Programming; e.g., html, css, javascript, command line tools

Q&A: ““What are required textbooks?”

None.

We will read research papers and online tutorials.

Class Logistics & Overview

- Grading:

	% of Final Class Grade
Class Participation	5%
Reading Assignments	25%
Lab Assignments	30%
Final Project	40%

- Website
  - https://www.ischool.utexas.edu/~dannag/Courses/CrowdsourcingForCV/
- Objectives, schedule, assignments, and policies
  - https://www.ischool.utexas.edu/~dannag/Courses/CrowdsourcingForCV/Syllabus/Syllabus.pdf

Class Format

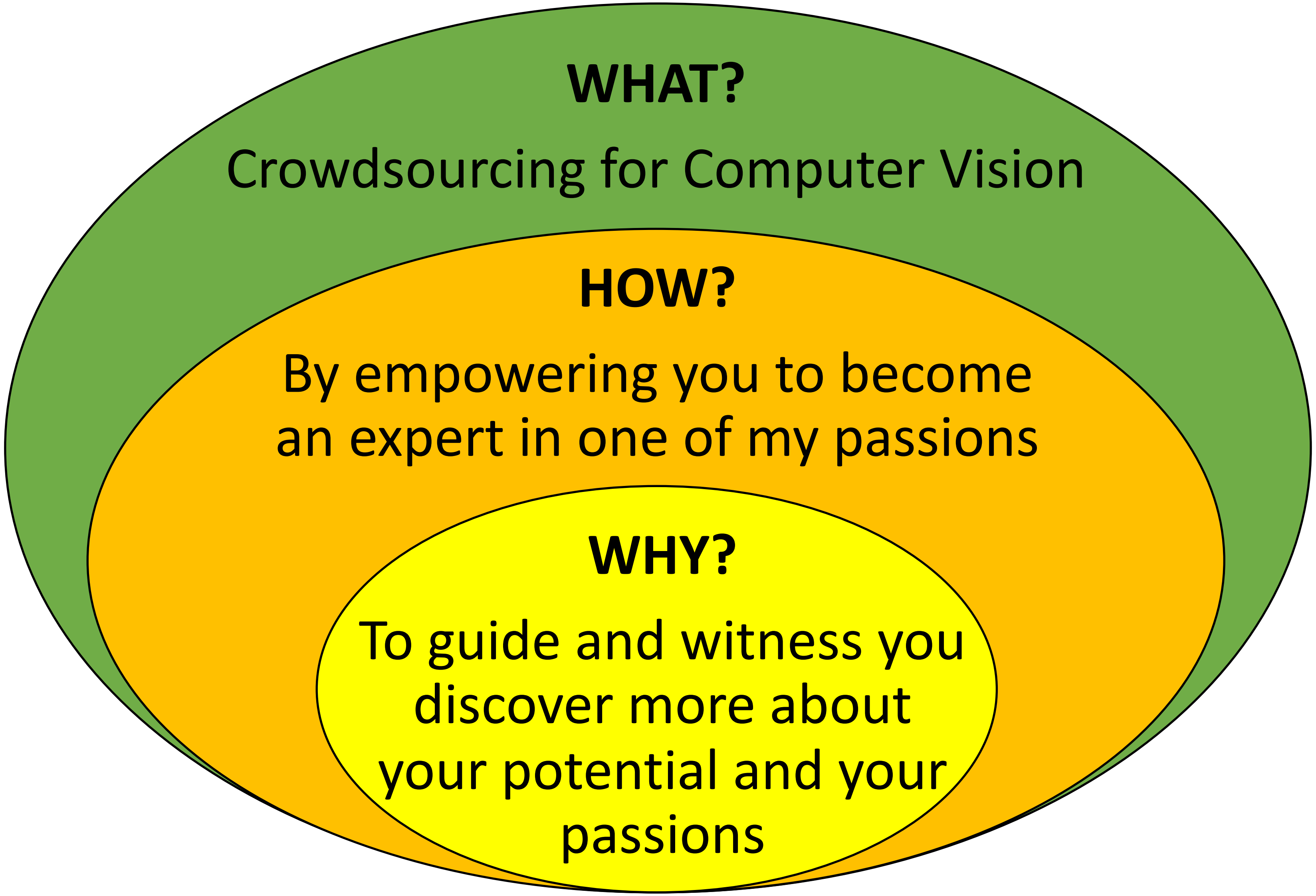
- First half = lecture & group discussions
- Break
- Second half = hands-on lab session

Congratulations!

- By taking this class, you receive a gift of:

- Thanks to: Microsoft Azure

What is My “Why” for Teaching You…



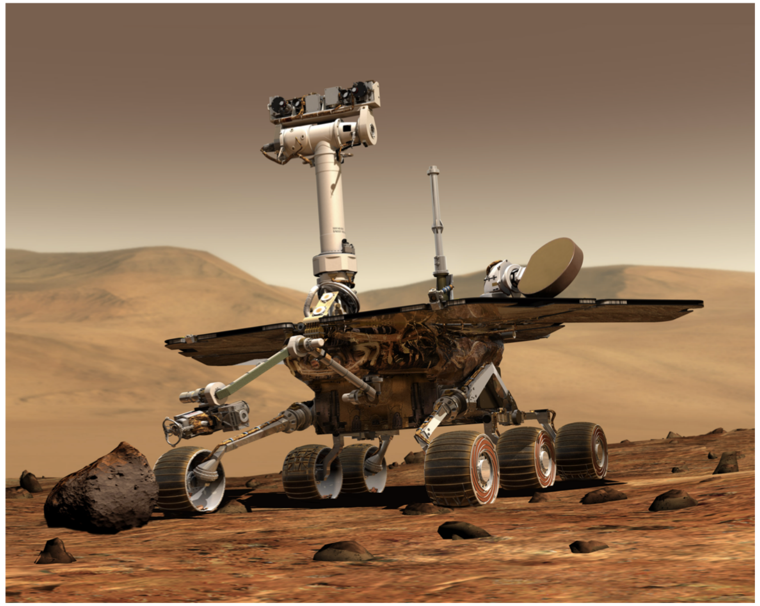
Today’s Topics

- Class logistics
- Computer vision: past, present, & future
- Computer vision: what makes it hard?
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

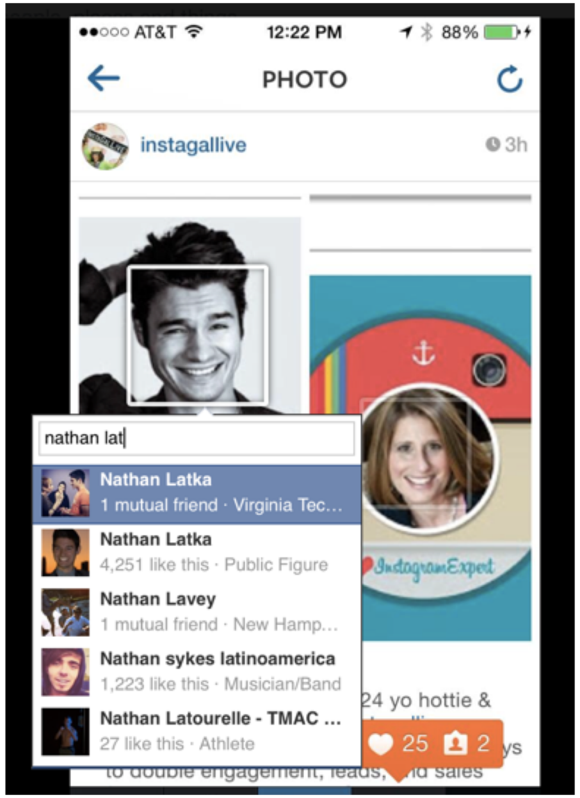
What is Computer Vision?

Algorithms that allow computers to “see”

Modern Examples of Computer Vision



e.g., self-driving vehicle on Mars



e.g., recognizing people

e.g., shopping without a cashier

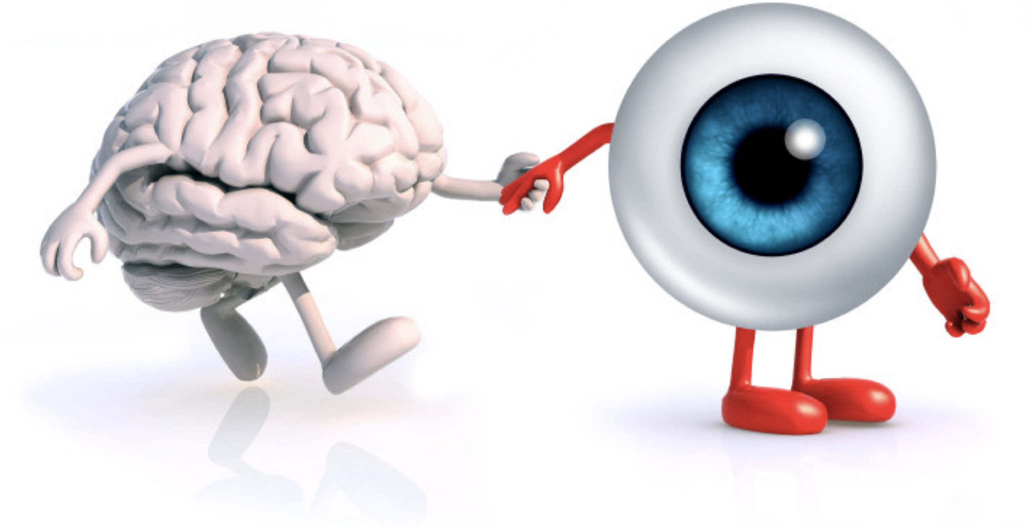
Modern Examples of Computer Vision

With > 85% of internet data in the form of images and videos, there are many opportunities for computer vision applications!

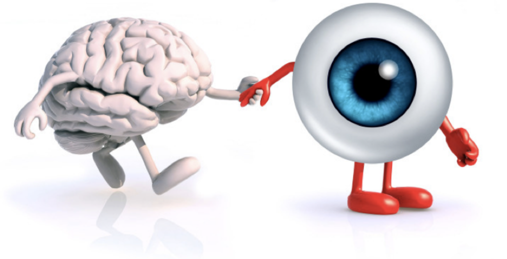
https://sevenshinestudios.wordpress.com/computer-vision-and-deep-learning/

Origins of Computer Vision

- Emulate ingredients that power human sight: brain, eyes, & conductor



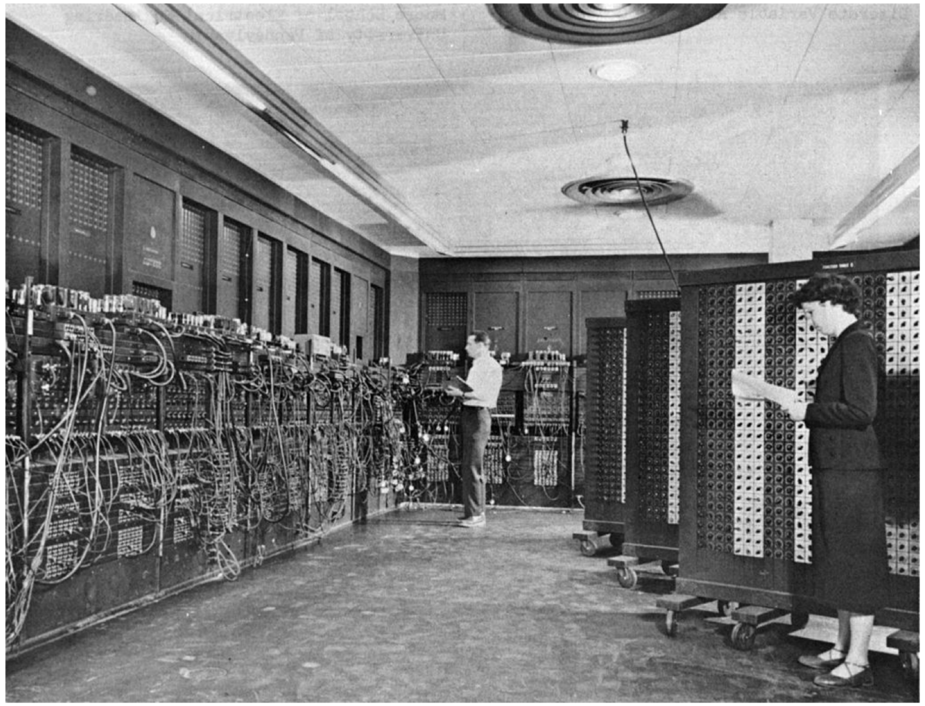
Origins of Computer Vision:



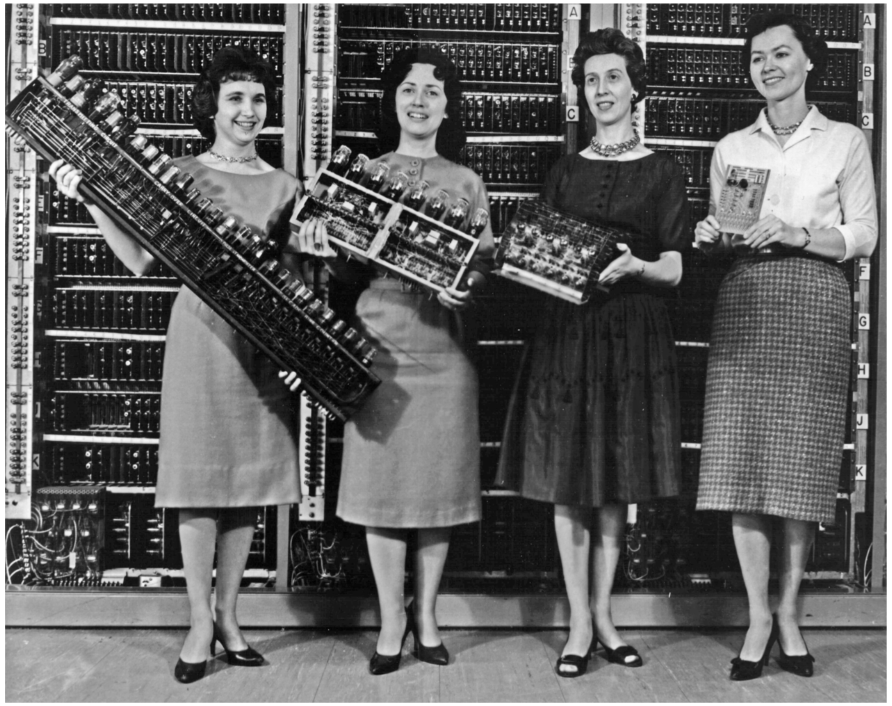
1945

Origins of Computer Vision:

1945



ENIAC (Electronic Numerical Integrator and Computer) created during World War II



First programmers

Origins of Computer Vision:

1945

Origins of Computer Vision:

Origins of Computer Vision:

1945

1957

First digital image (176 x 176 pixels)

Origins of Computer Vision:

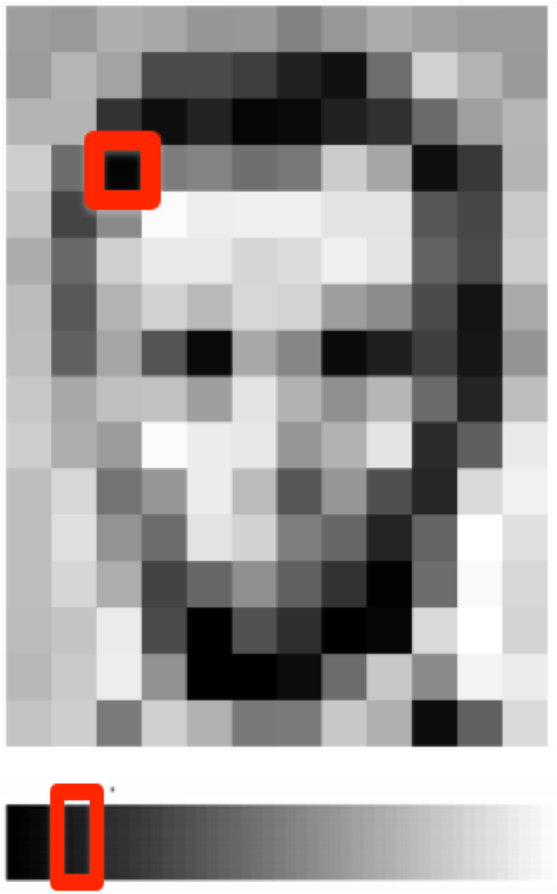
1945

1957

Digital images

What a Computer Sees:

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	[illegible]	131	111	120	204	166	15	56	180
194	68	[illegible]	[illegible]	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218



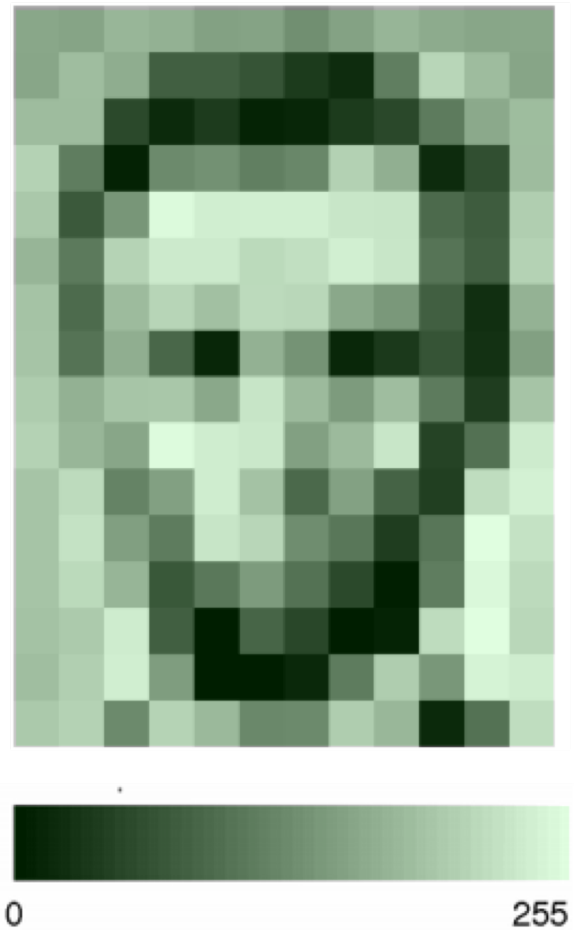
Origins of Computer Vision:

1945 — 1957

Digital images

What a Computer Sees:

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218



Origins of Computer Vision:

1945

1957

Digital images

What a Computer Sees:

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

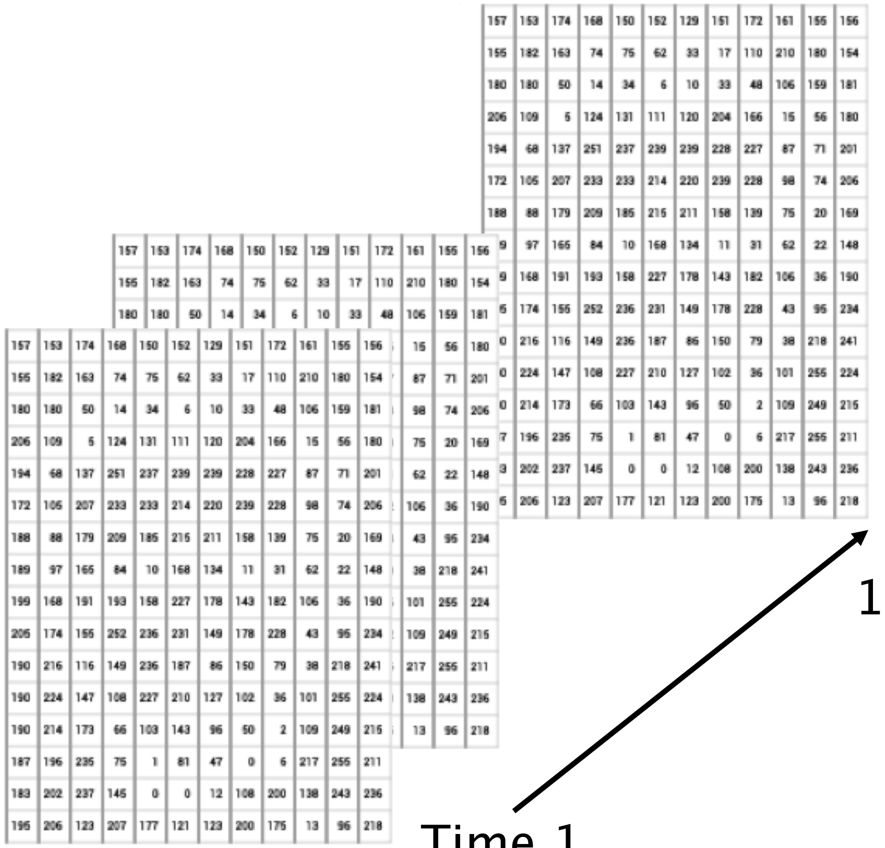
Origins of Computer Vision:

1945

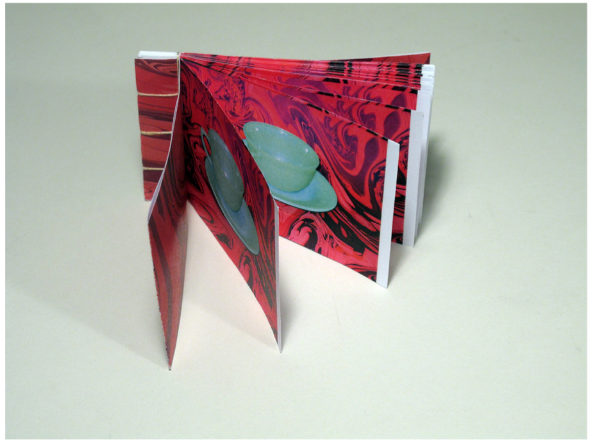
1957

Digital images

What a Computer Sees:



Analogous to (for video):



Origins of Computer Vision

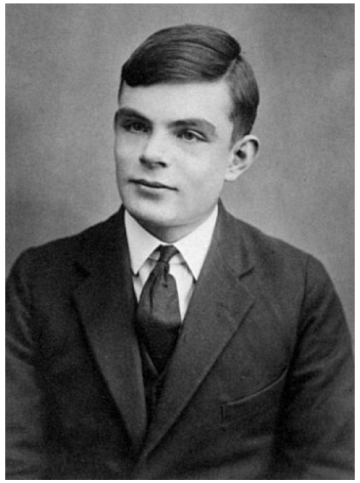
1945

1957

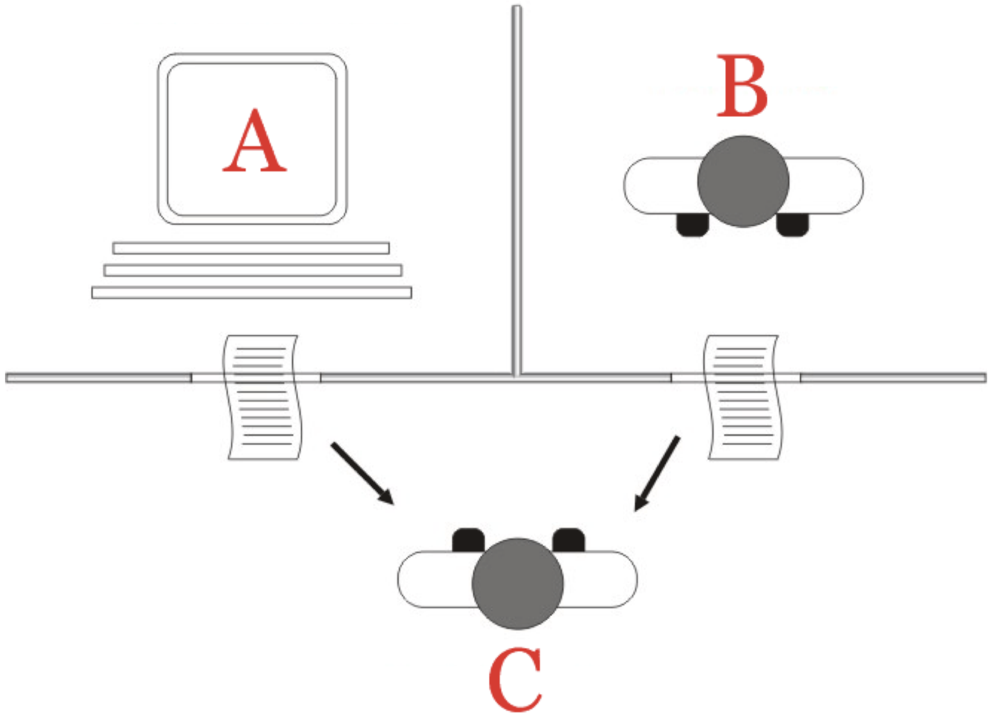
Origins of Computer Vision

1945 1950 1957

Turing Test



Alan Turing
(1912-1954)

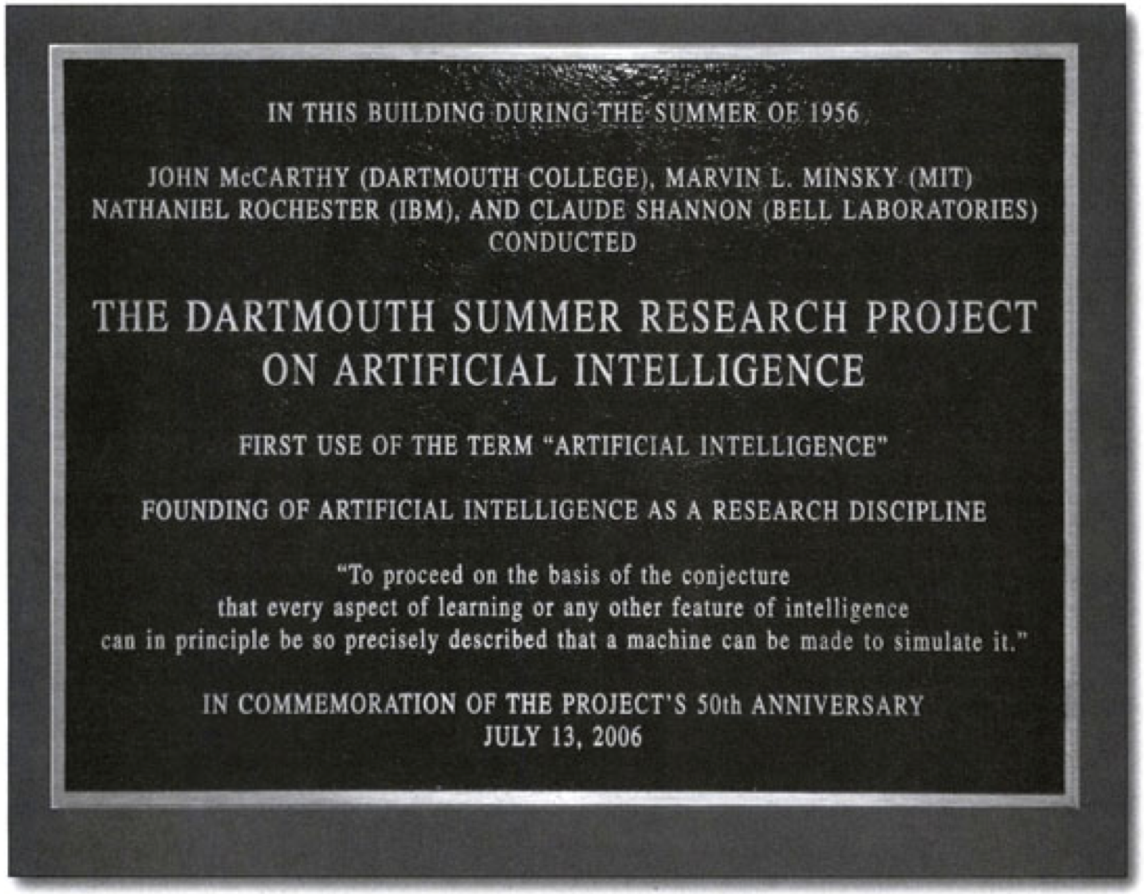


Turing Test: can ”C” decide whether text responses come from a machine or human

Origins of Computer Vision

1945 1950 1956

Turing Test AI Birth



Origins of Computer Vision

1945 1950 1956

Turing Test AI Birth

Workshop Proposal: "… We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer…"

Origins of Computer Vision

1945 | 1950 | 1956

Turing Test | AI Birth

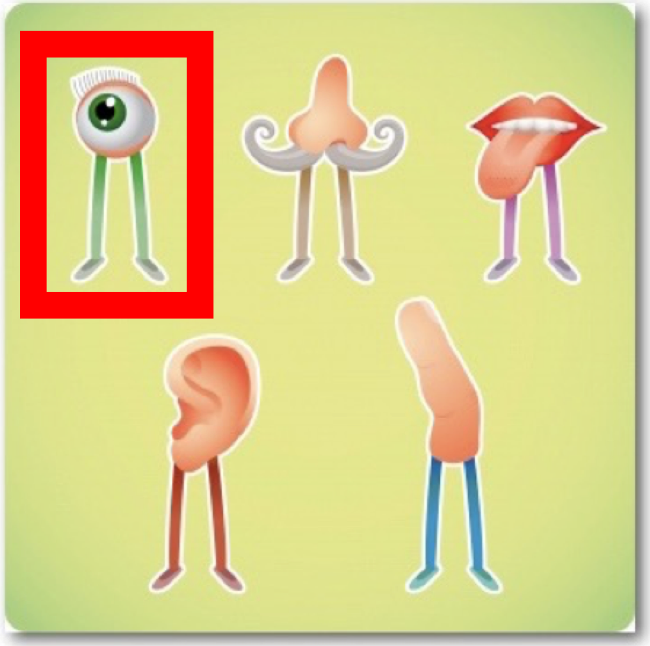
Workshop Proposal: "… We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer…"

Origins of Computer Vision

1945 1950 1956

Turing Test AI Birth

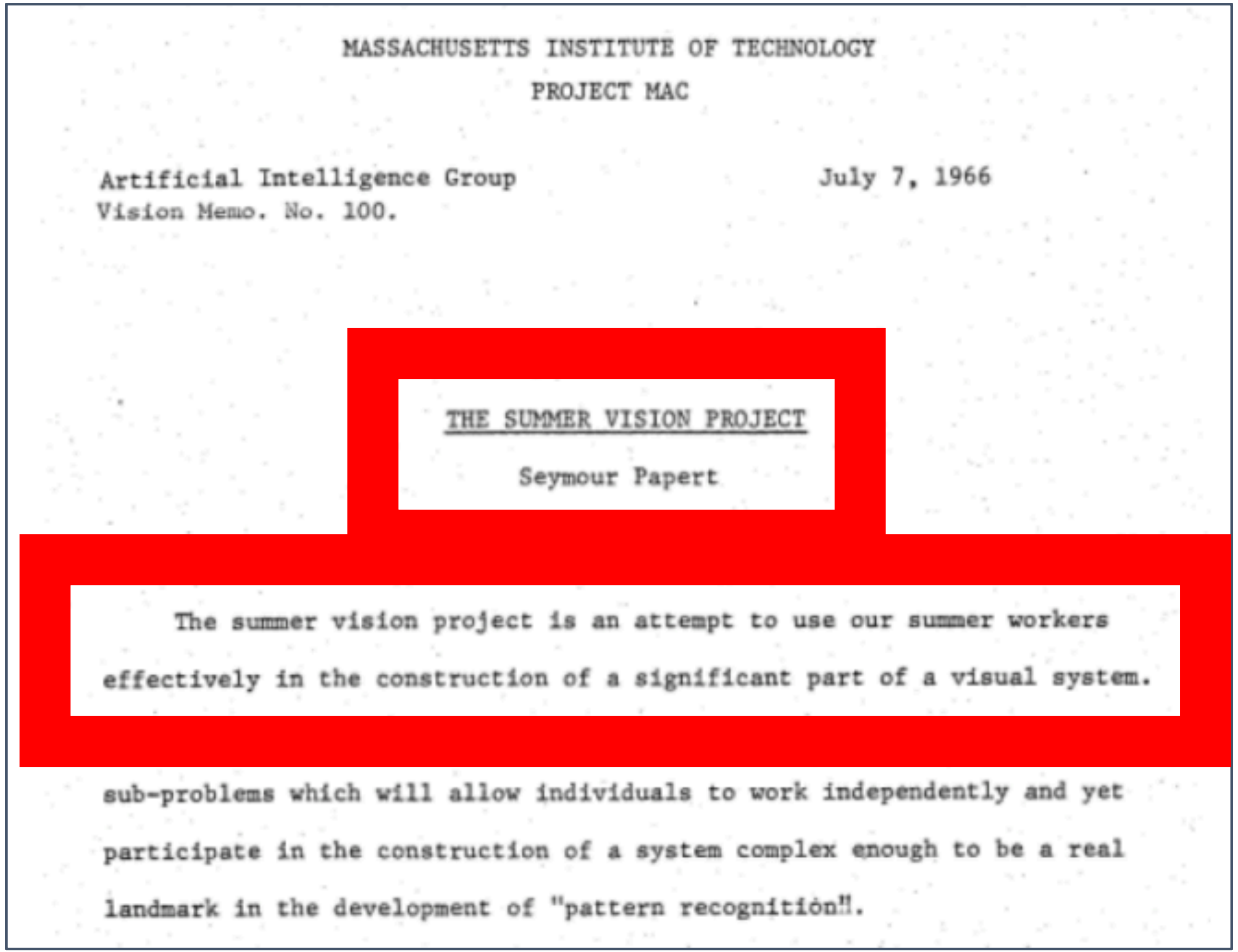
What human intelligence might machines imitate?



Origins of Computer Vision

1945 1950 1956 1966

Turing Test AI Birth Birth of Computer Vision



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

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

THE SUMMER VISION PROJECT

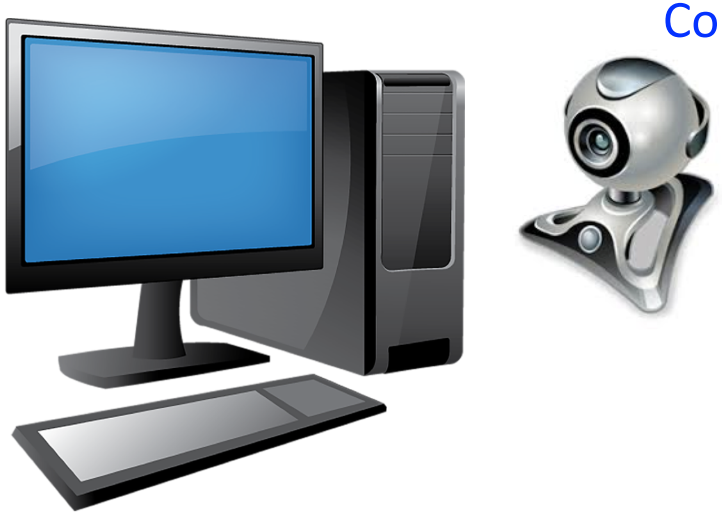
Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. 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".

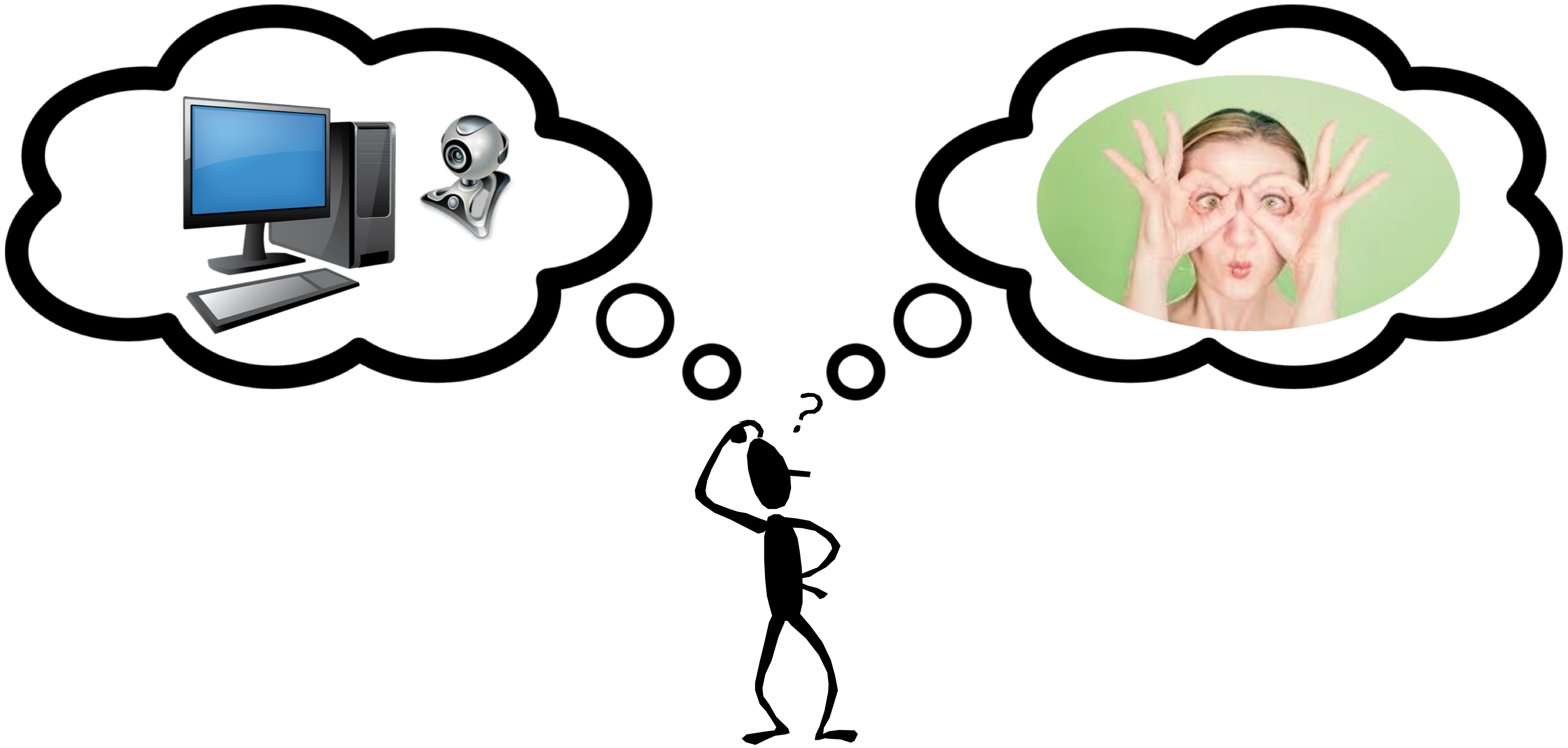
Origins of Computer Vision

1945 1950 1956 1966

Turing Test AI Birth Birth of Computer Vision



Turing Test: design "computer vision" that is indistinguishable from "human vision"



Origins of Computer Vision

1945 1950 1956 1966

Turing Test AI Birth Birth of Computer Vision

e.g.,

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

What is this?
- A picture of a person

Could you describe this person?
- Long face
- Angular jaw
- Has a beard

Who is this person?
- Abraham Lincoln

Is this person happy?
- I am not sure.

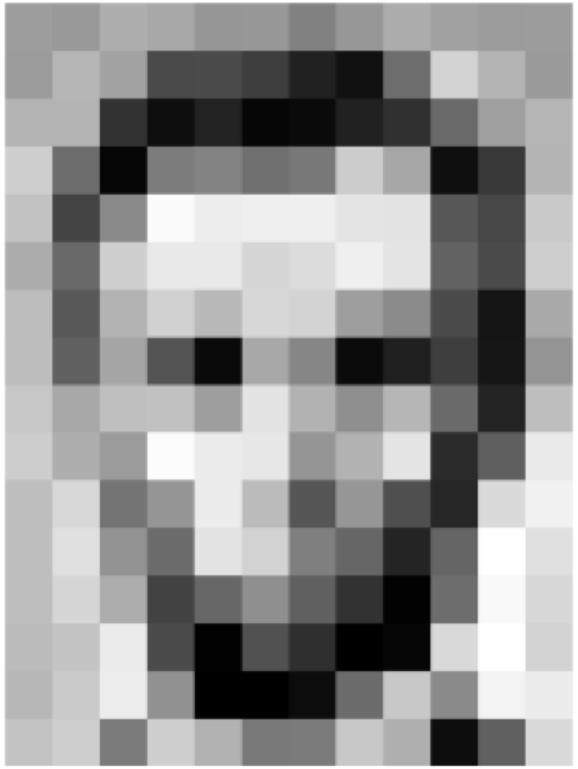
Is this person attractive?
- ~70% of people would say “yes”

Origins of Computer Vision

1945 — 1950 — 1956 — 1966

Turing Test — AI Birth — Birth of Computer Vision

e.g.,



What is this?
- A picture of a person

Could you describe this person?
- Long face
- Angular jaw
- Has a beard

Who is this person?
- Abraham Lincoln

Is this person happy?
- I am not sure.

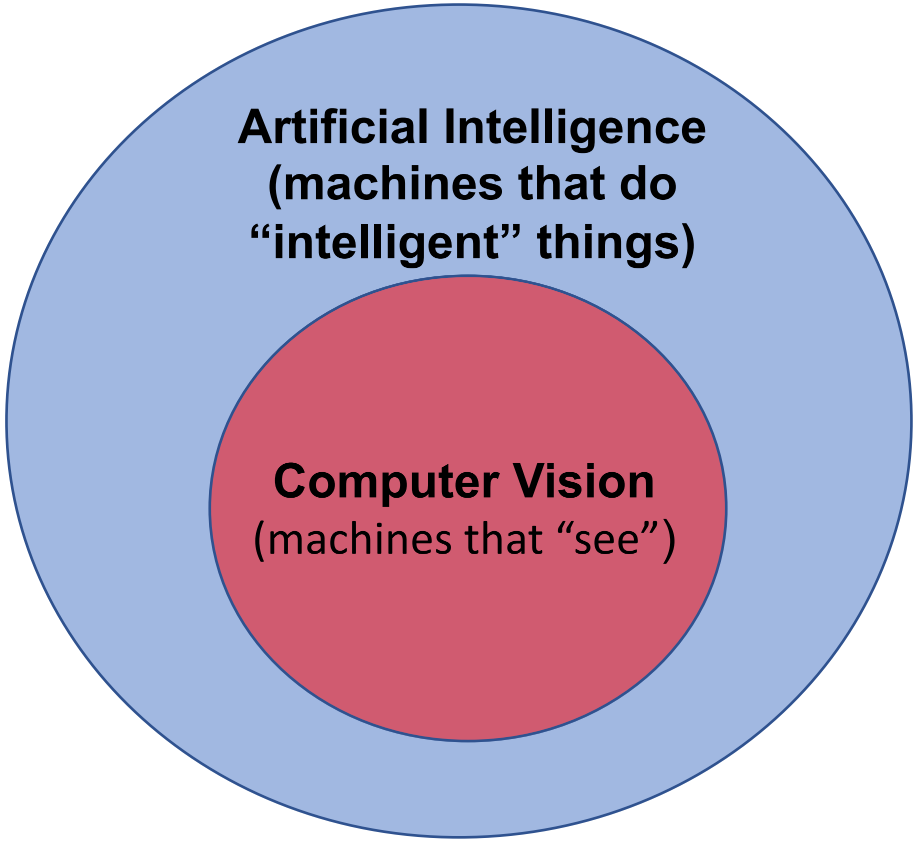
Is this person attractive?
- ~70% of people would say “yes”

Origins of Computer Vision

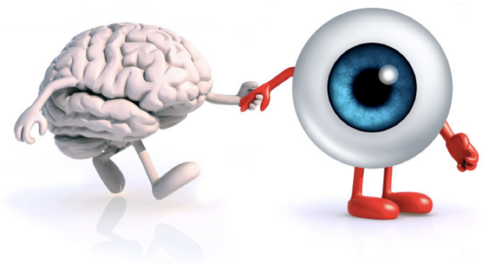
1945 1950 1956 1966

Turing Test AI Birth

Birth of Computer Vision



Origins of Computer Vision



1945 1957 1966

Brain, eyes, & conductor needed to emulate human sight were born over ~20 years!

Computer Vision in **Research**

1945 1957 1966 1983 1987 1990

- 1983: 1rst Computer Vision and Pattern Recognition (CVPR)
- 1987: 1rst International Conference on Computer Vision (ICCV)
- 1990: 1rst European Conference on Computer Vision (ECCV)

And more including:

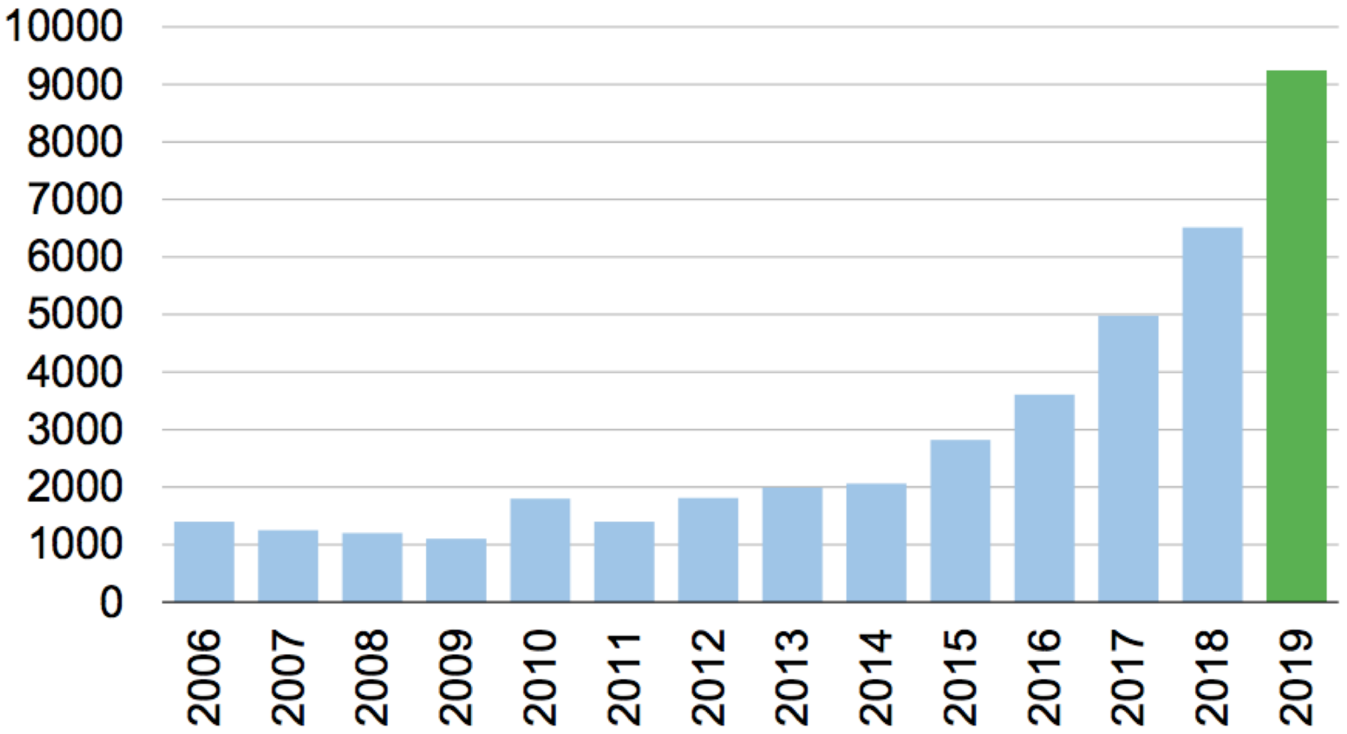
- Asian Conference on Computer Vision (ACCV)
- British Machine Vision Conference (BMVC)
- Winter conference on Applications in Computer Vision (WACV)
- Medical Image Computing and Computer-Assisted Intervention (MICCAI)
- Conference on Automatic Face and Gesture Recognition (IEEE FG)

Computer Vision in Research

1945 — 1957 — 1966 — 1983 — 1987 — 1990

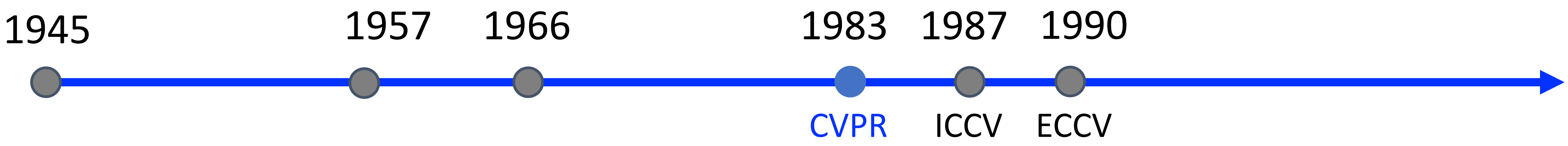
CVPR (1983), ICCV (1987), ECCV (1990)

Number of attendees:

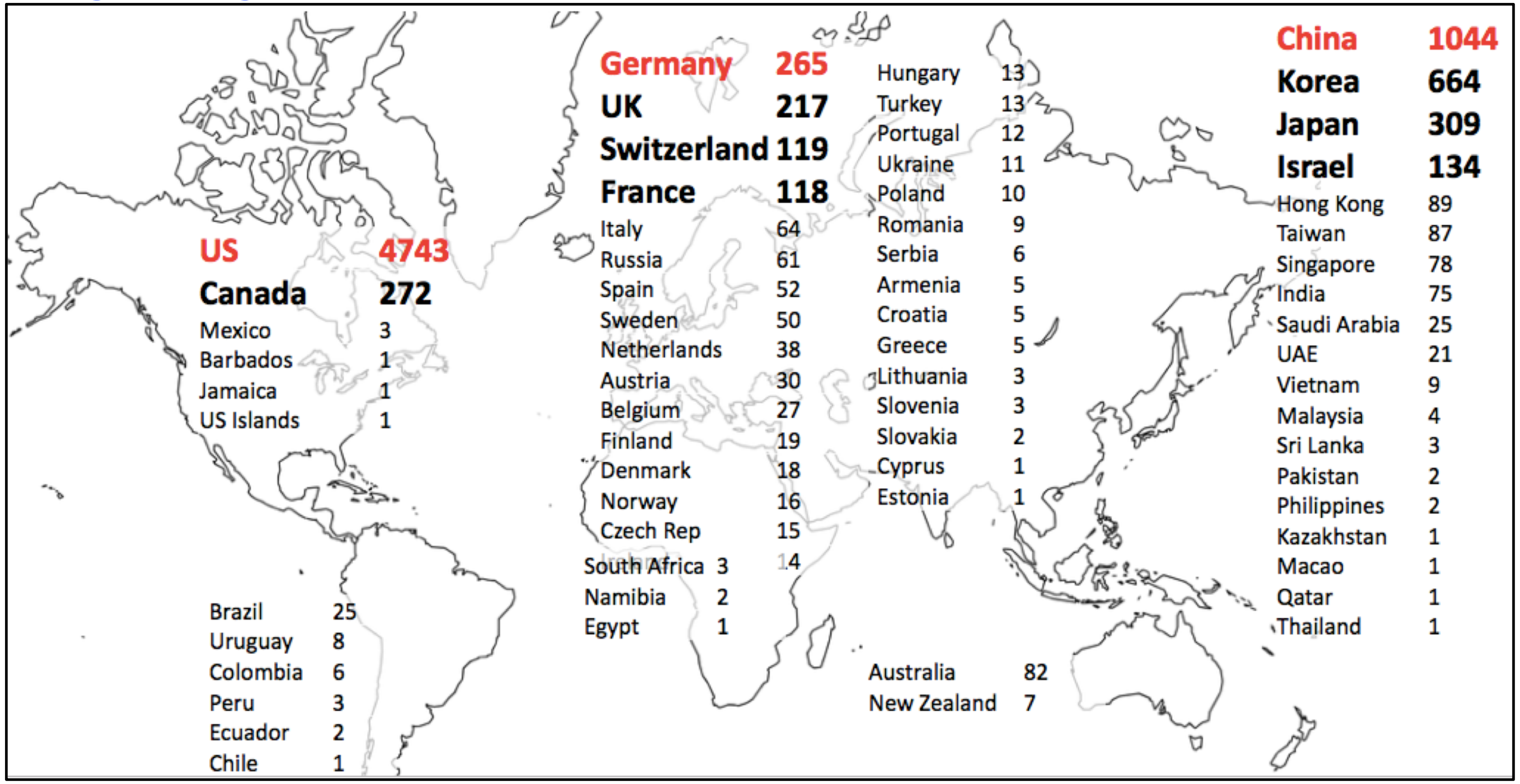


http://cvpr2019.thecvf.com/

Computer Vision in Research



Originating countries of attendees in 2019:

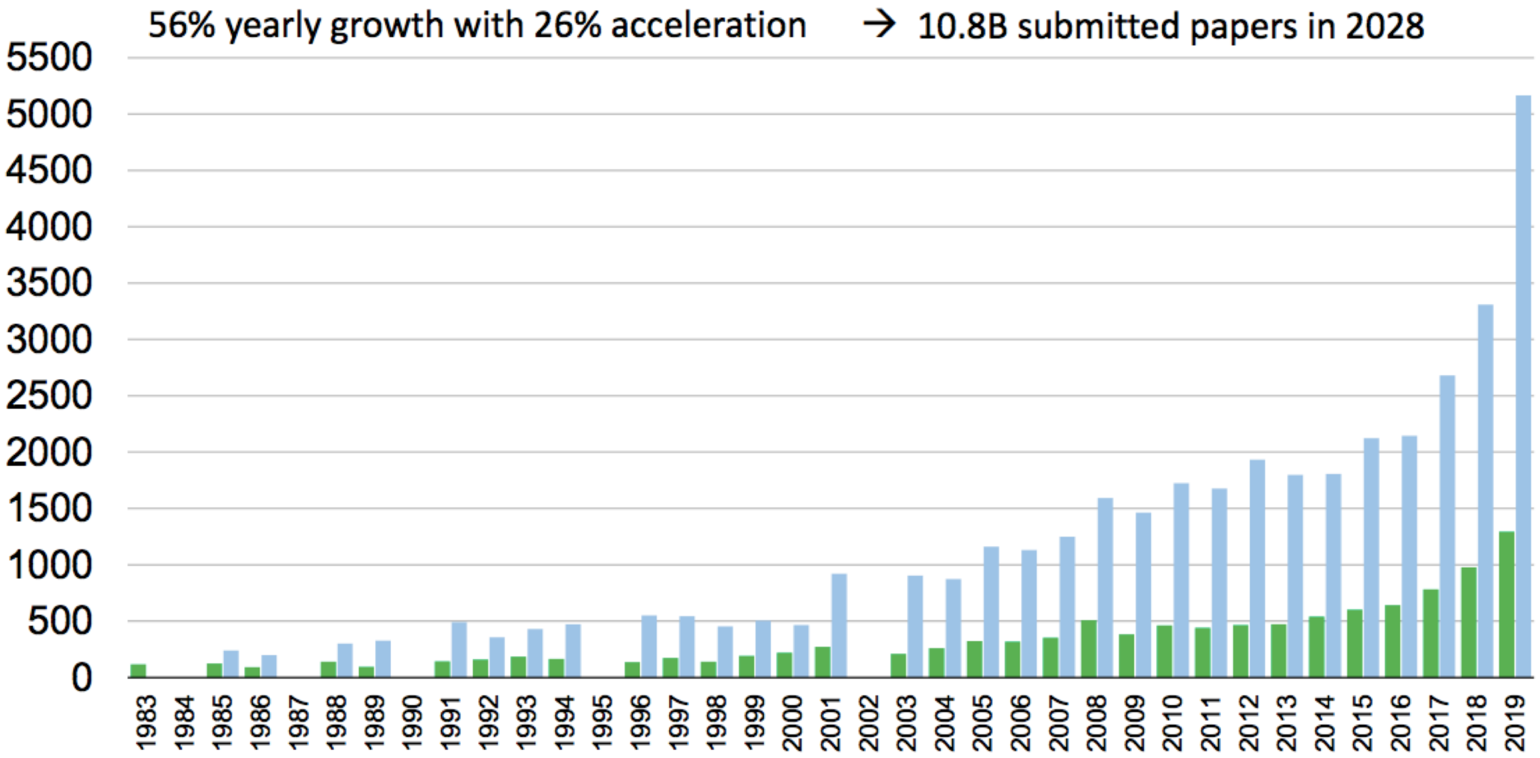


http://cvpr2019.thecvf.com/

Computer Vision in Research

1945 — 1957 — 1966 — 1983 (CVPR) — 1987 (ICCV) — 1990 (ECCV)

Number of submitted and accepted papers:



http://cvpr2019.thecvf.com/

Computer Vision in Industry

1945 — 1957 — 1966 — 1983 — 1987 — 1990 — 2015

1983: CVPR

1987: ICCV

1990: ECCV

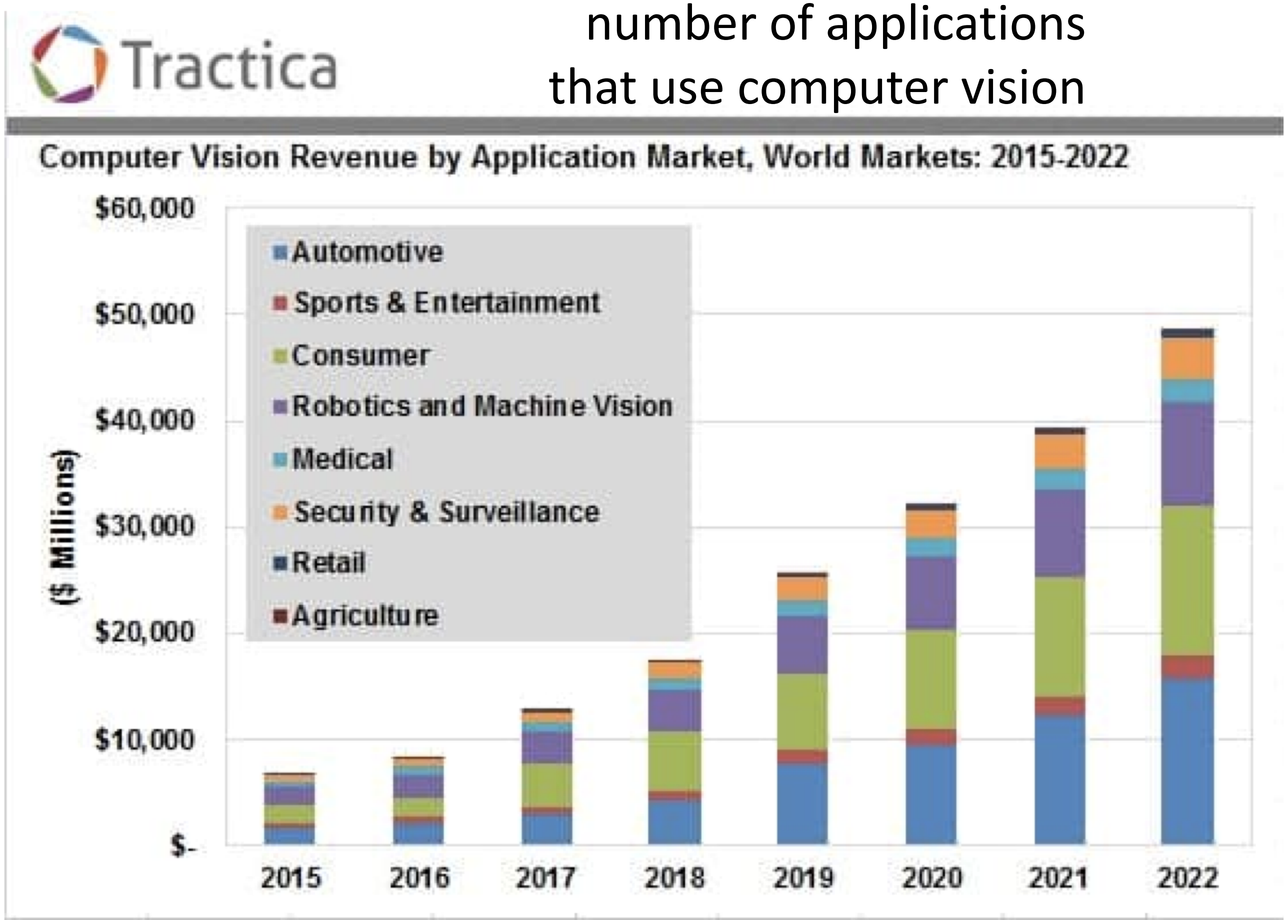
2015: Rapid growth in the number of applications that use computer vision

Computer Vision in Industry

1945 1957 1966 1983 1987 1990 2015

CVPR ICCV ECCV

Rapid growth in the number of applications that use computer vision

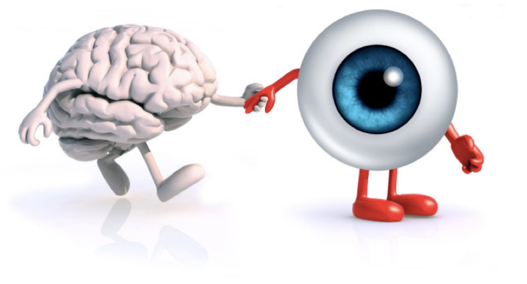


https://sevenshinestudios.wordpress.com/computer-vision-and-deep-learning/

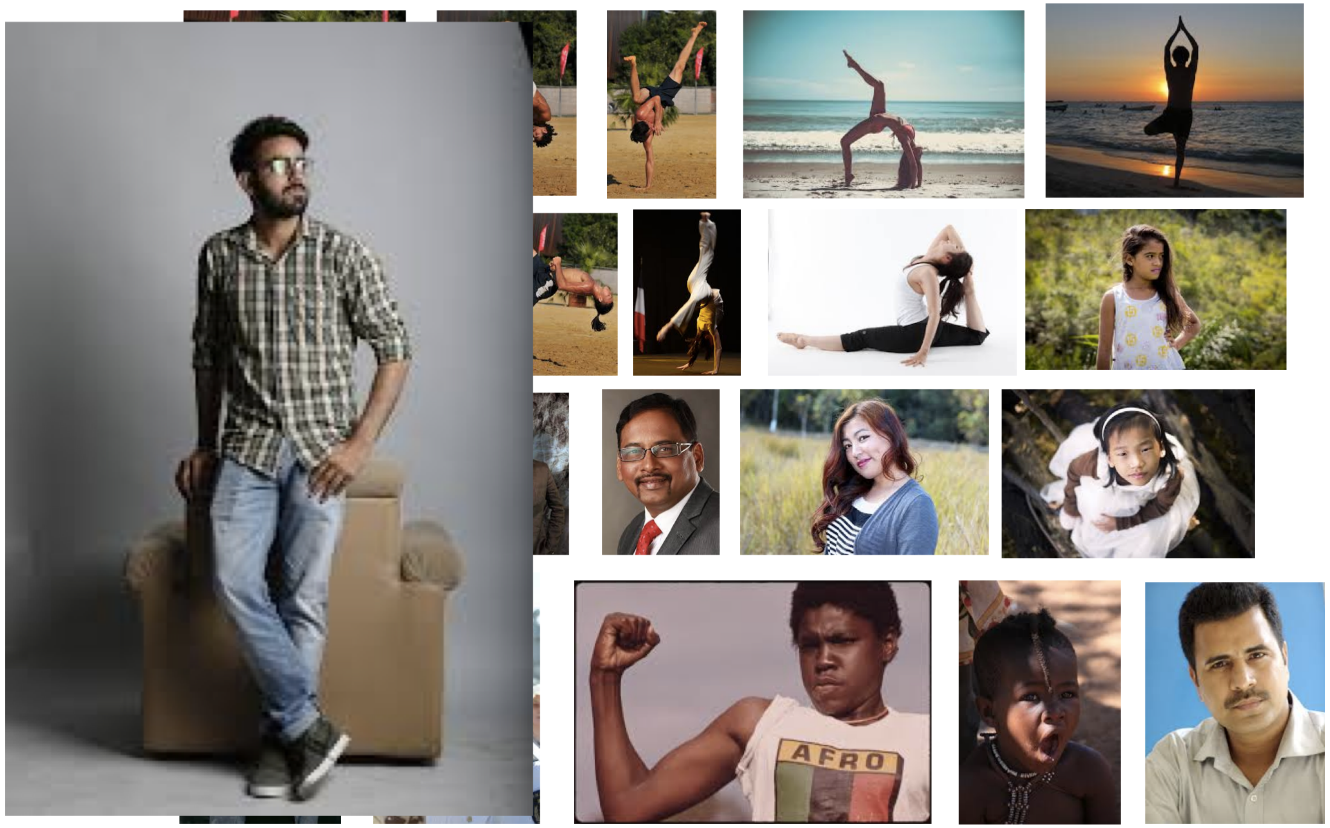
Today’s Topics

- Class logistics
- Computer vision: past, present, & future
- Computer vision: what makes it hard?
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

Group Discussion



How would you instruct a computer to answer: “Is a person in the image?”



So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

So Much Variation for So Many Tasks! (List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

e.g., take a picture of an object and find where to buy it

So Much Variation for So Many Tasks! (List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

Kitchen

Store

So Much Variation for So Many Tasks! (List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

e.g., describe a bird to learn what type it is
Demo: https://www.youtube.com/watch?v=UPcz9Y17iCc

So Much Variation for So Many Tasks! (List Restricted to Topics We Will Cover in Class)

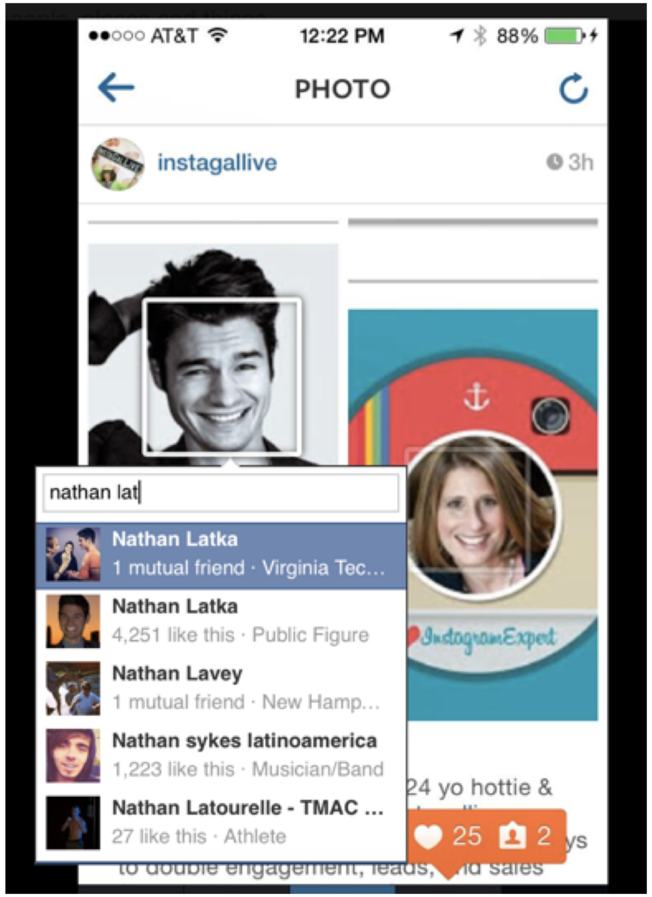
- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

e.g., rotoscoping (more examples on Wiki)

https://www.starnow.co.uk/ahmedmohammed1/photos/4650871/before-and-after-rotoscopinggreen-screening

So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

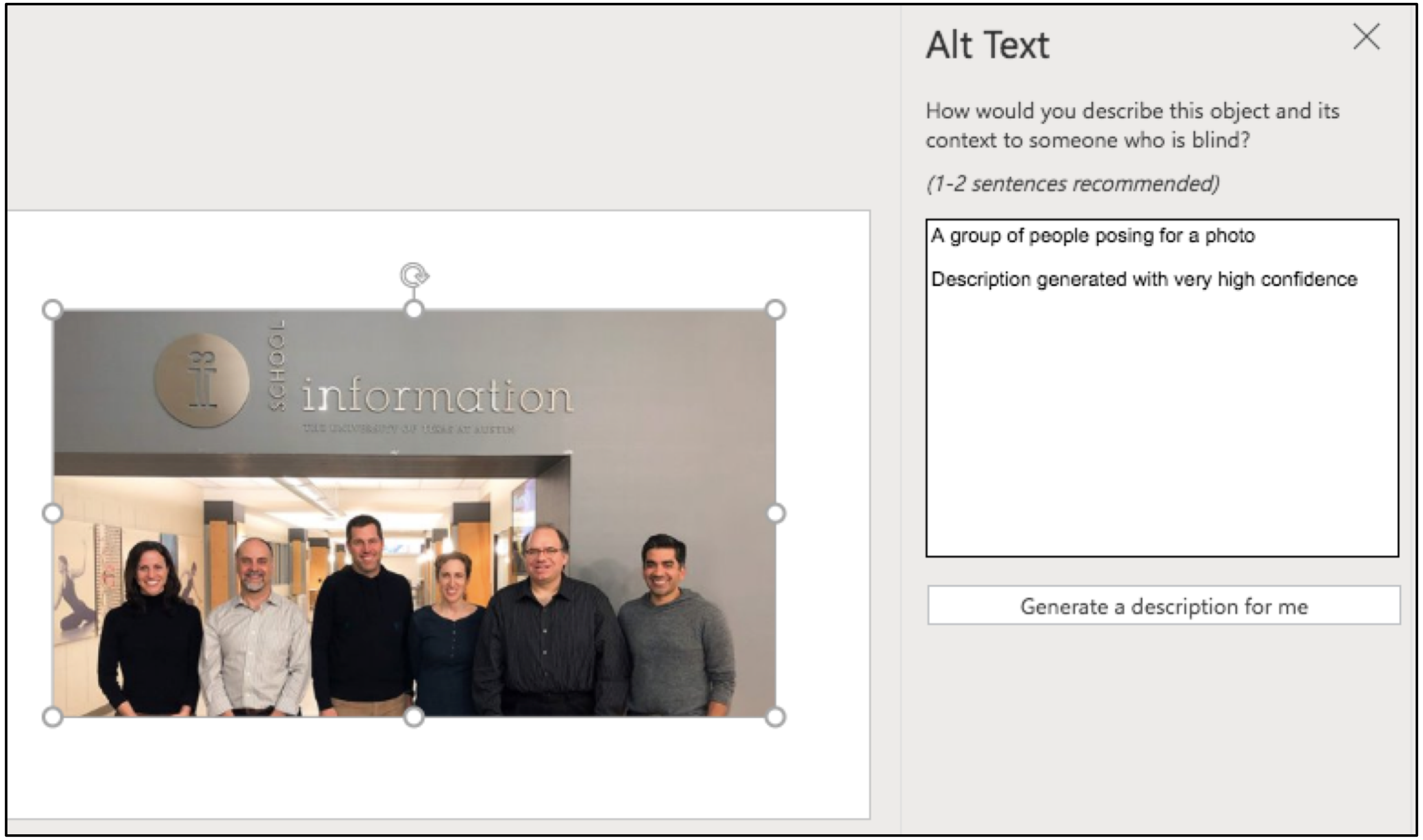
- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…



e.g., detect faces to tag

So Much Variation for So Many Tasks! (List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- **Image Captioning**
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…



e.g., Microsoft Power Point (Office 365 demo)

So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- **Activity/Event Recognition**
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

e.g., shopping without a cashier

So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

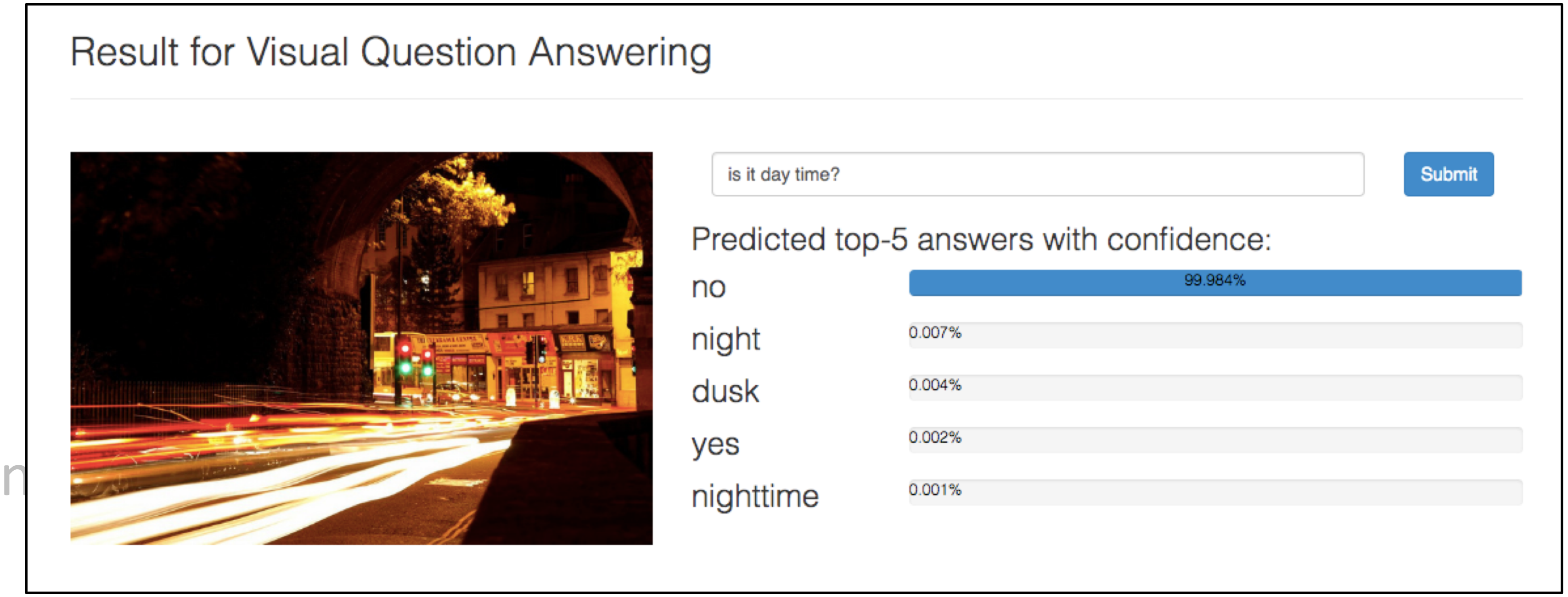
- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- Subjective Problems
- And more…

e.g., track bowling ball path

e.g., calculate bat speed

So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

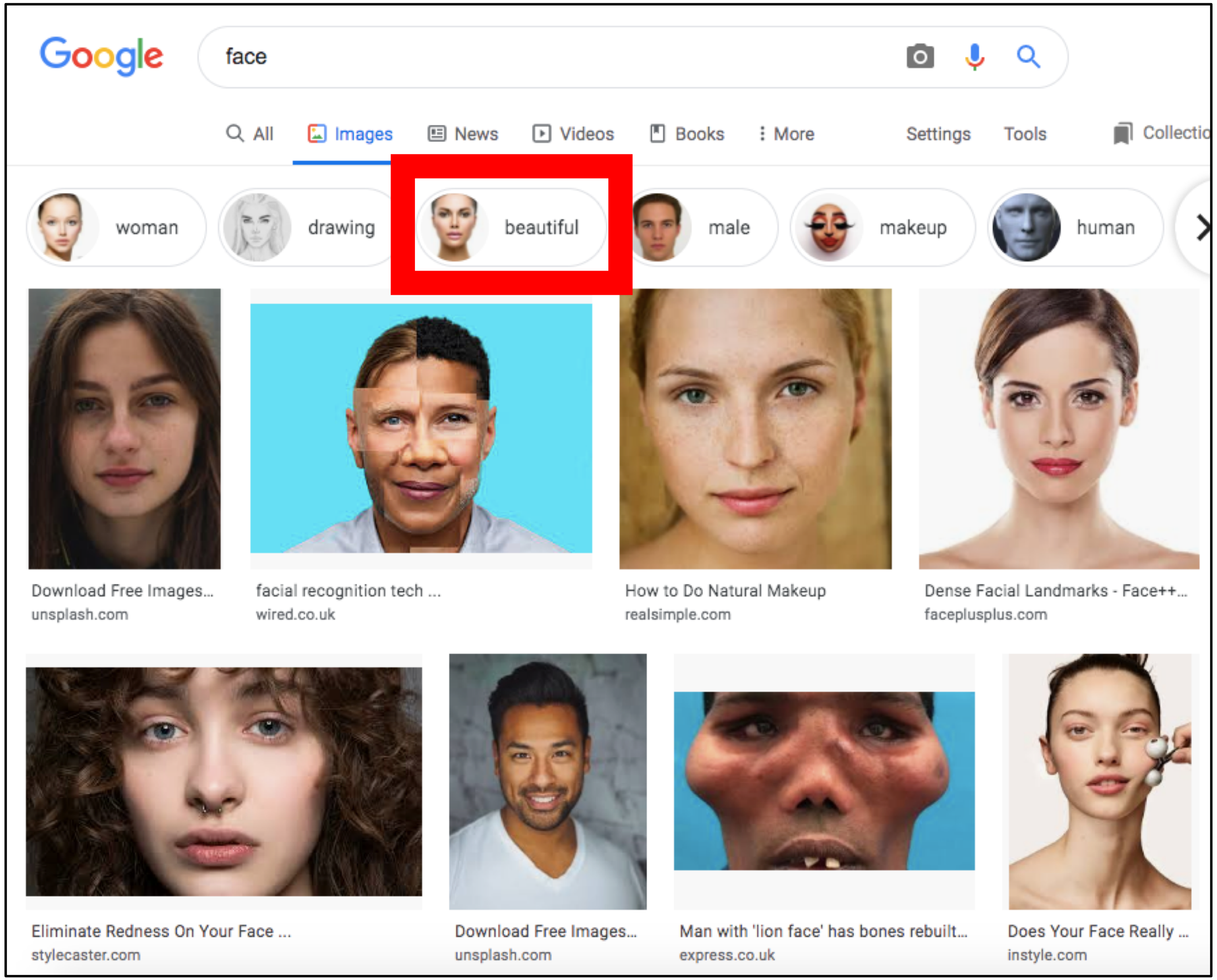
- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recogn
- Object Tracking
- **Visual Question Answering**
- Subjective Problems
- And more…



Demo: http://vqa.cloudcv.org/

So Much Variation for So Many Tasks!
(List Restricted to Topics We Will Cover in Class)

- Object recognition
- Scene classification
- Attribute labeling
- Segmentation
- Object detection
- Image Captioning
- Activity/Event Recognition
- Object Tracking
- Visual Question Answering
- **Subjective Problems**
- And more…



Key Challenge: Replicate Human Vision for **So Much Variation** for **So Many Tasks**!

Design computer vision that is indistinguishable from **human** vision

Little Training Highly Trained

Key Challenge: Replicate Human Vision for **So Much Variation** for **So Many Tasks**!

images on hard drive: 10^5
(500 GB/2 MB = 250,000)

images seen during my first 10 years: 10^9
(24 images/sec * 60 sec * 60 min * 16 hr * 365 days * 10 yrs = 5,045,760,000)

images seen by all humanity: 10^{20}
(7.5 billion humans[1] * 24 images/sec * 60 * 60 * 16 * 365 * 60 yrs = $2.23 * 10^{20}$)
[1] http://www.worldometers.info/world-population/

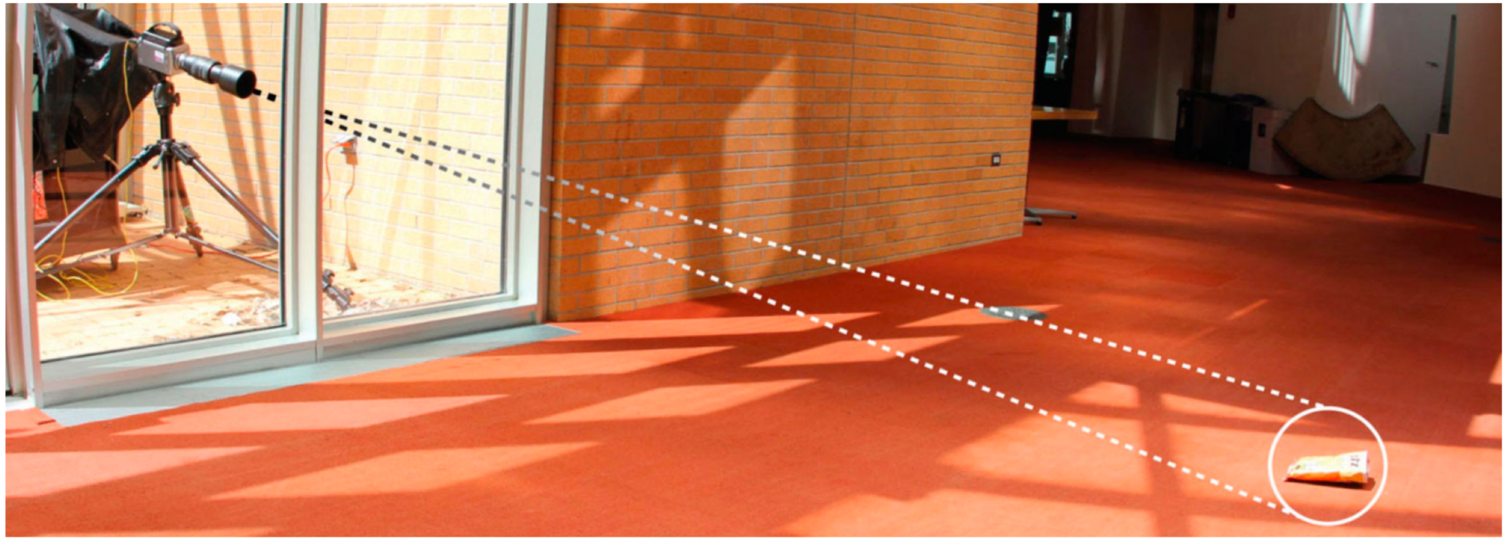
Adapted from slides by Antonio Torralba

Computer Vision Beyond Human Vision

e.g., face re-enactment
Demo: https://www.youtube.com/watch?v=Cx54WPwsG2w
Demo: https://www.youtube.com/watch?v=ttGUiwfTYvg

Computer Vision Beyond Human Vision

Detecting Heart Rate
Demo: https://www.youtube.com/watch?v=9JNkSZJuDJ8

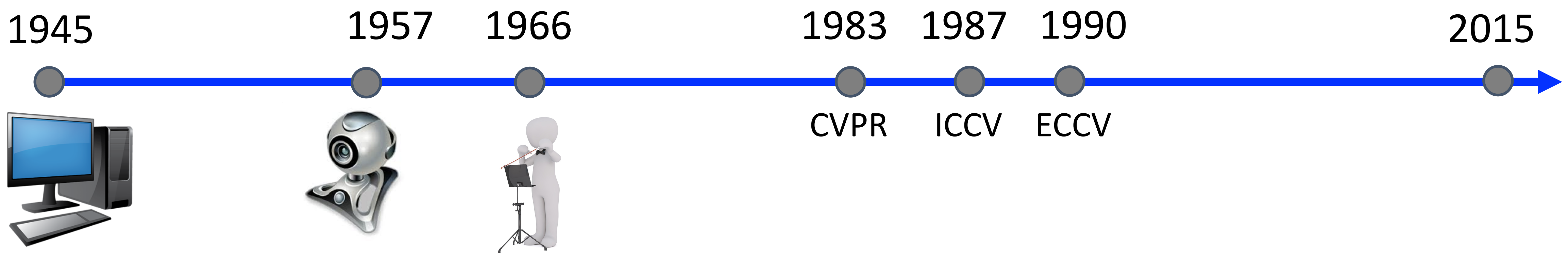


Recovering Sound
Demo: https://www.youtube.com/watch?v=npNYP2vzaPo

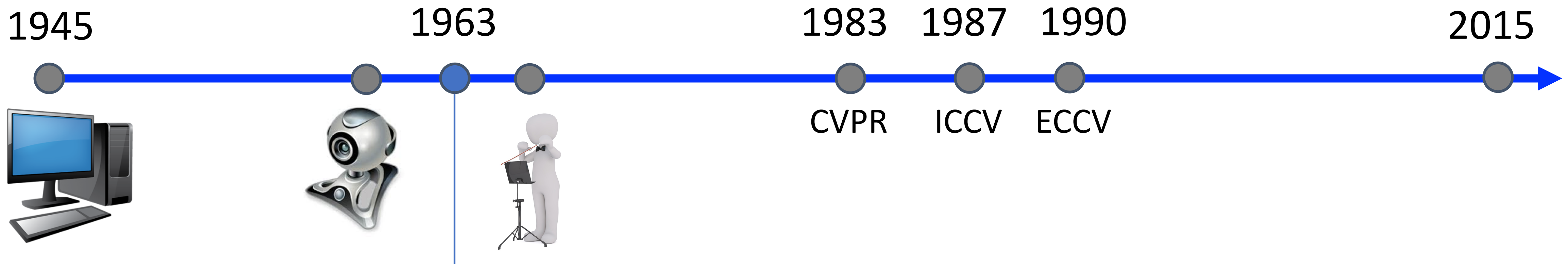
Today’s Topics

- Class logistics
- Computer vision: past, present, & future
- Computer vision: what makes it hard?
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

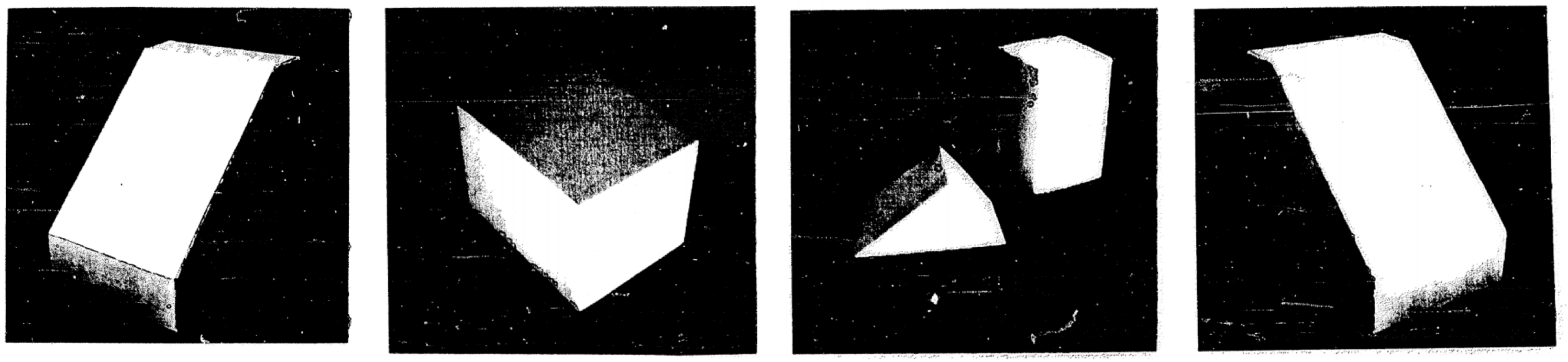
Progress Charted By Emulating Human Perception



Progress Charted By Emulating Human Perception



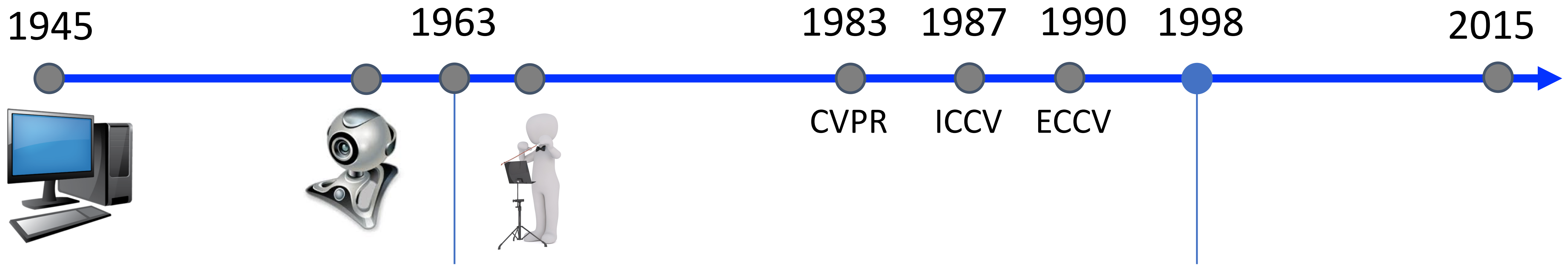
3D Objects



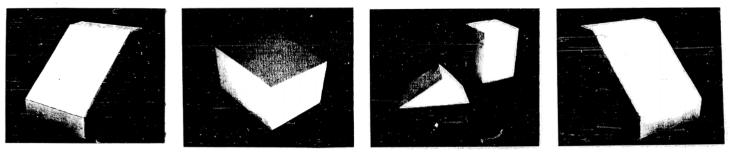
- 4 256x256 images could be stored in the computer memory
- Scanning each image to the computer took ~3 minutes

LG Roberts. Machine perception of three-dimensional solids. 1963

Progress Charted By Emulating Human Perception

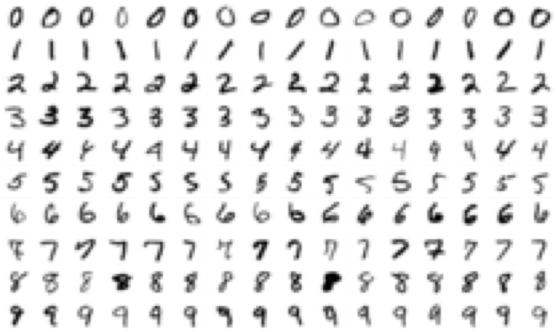


3D Objects



- 4 256x256 images

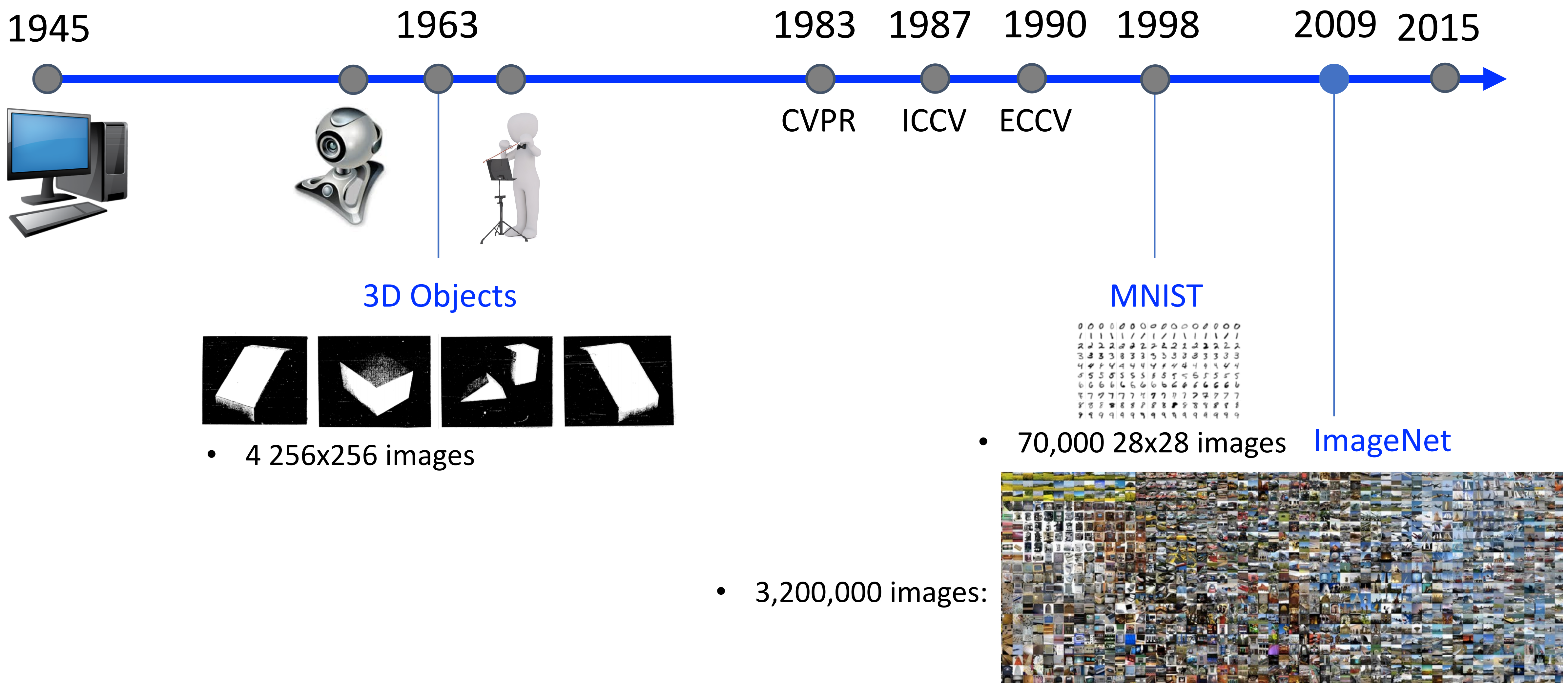
MNIST



- 70,000 28x28 images

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. 1998.

Progress Charted By Emulating Human Perception



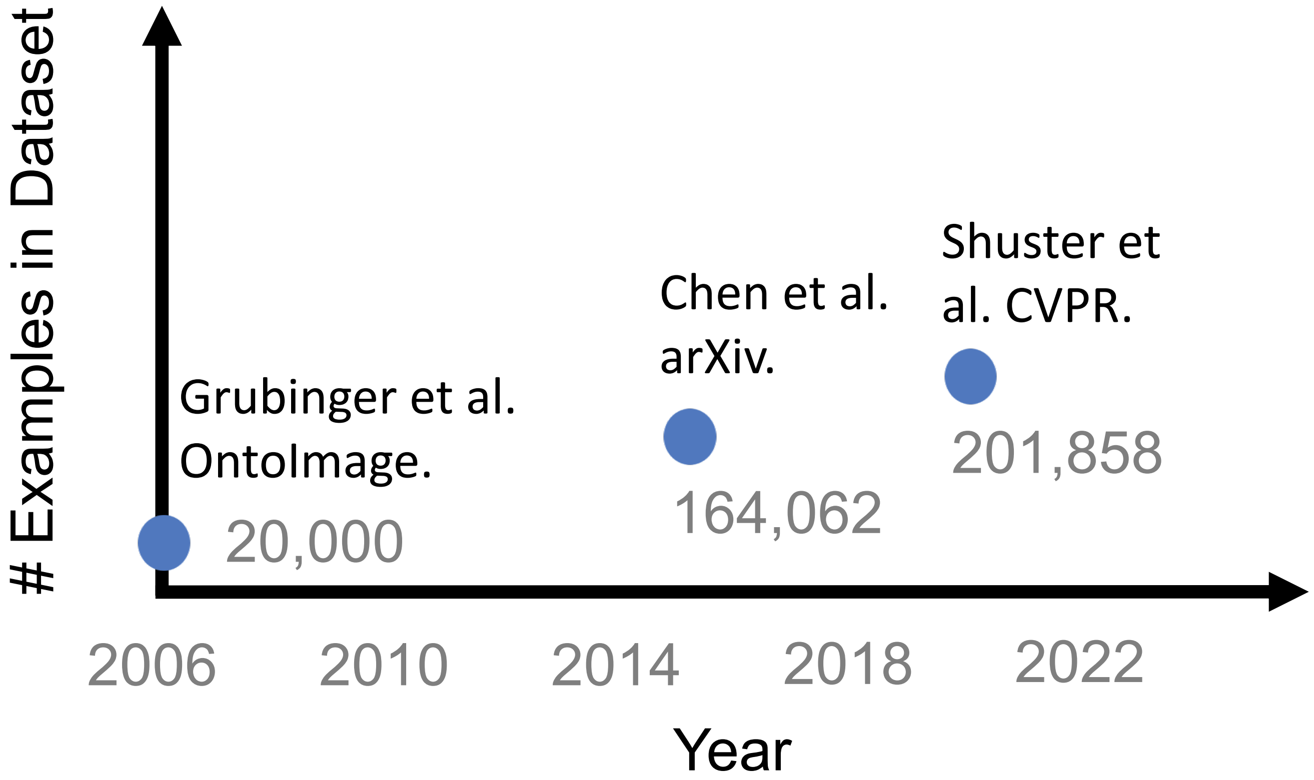
J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. 2009.

Progress Charted By Emulating Human Perception

Trend is to build bigger human-labelled datasets for many vision tasks!

Trend is to Build Bigger Labelled Datasets; e.g.,

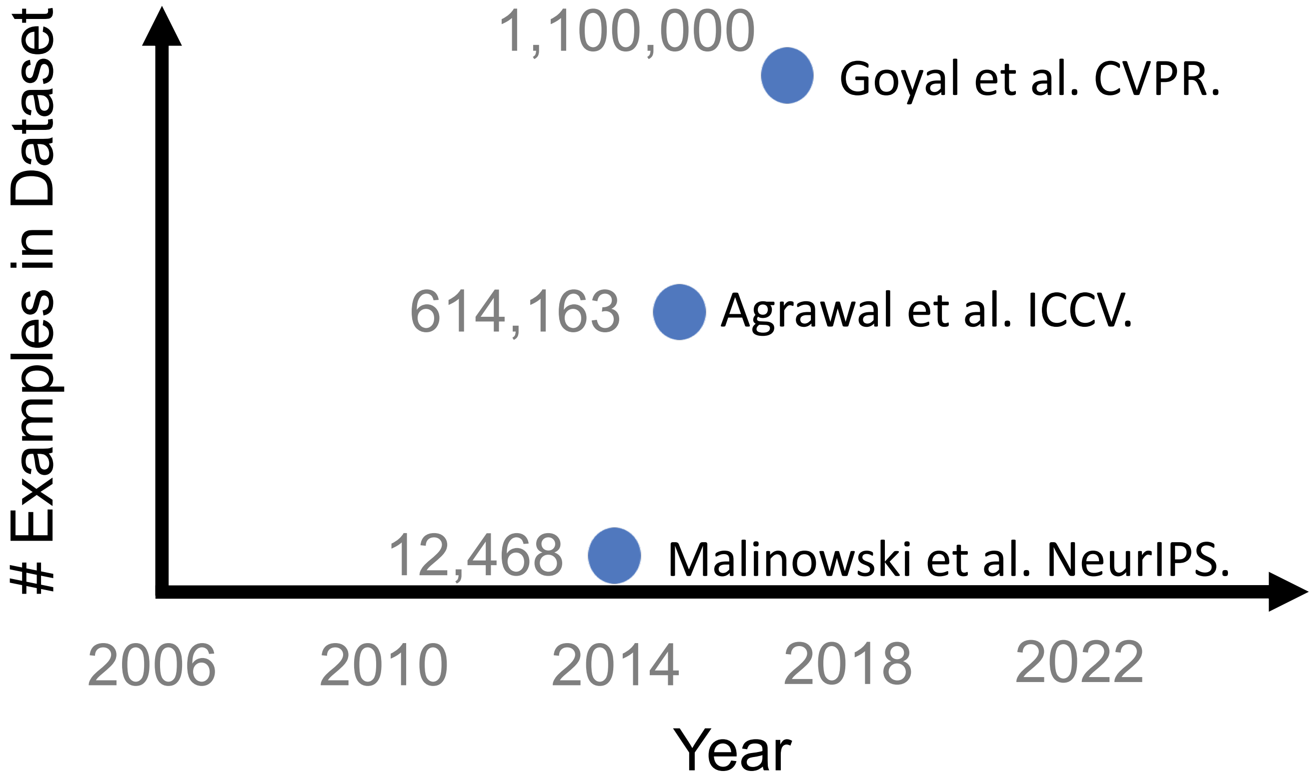
Caption:
Two tee shirts, one white the other black, hanging in a closet.



Trend is to Build Bigger Labelled Datasets; e.g.,

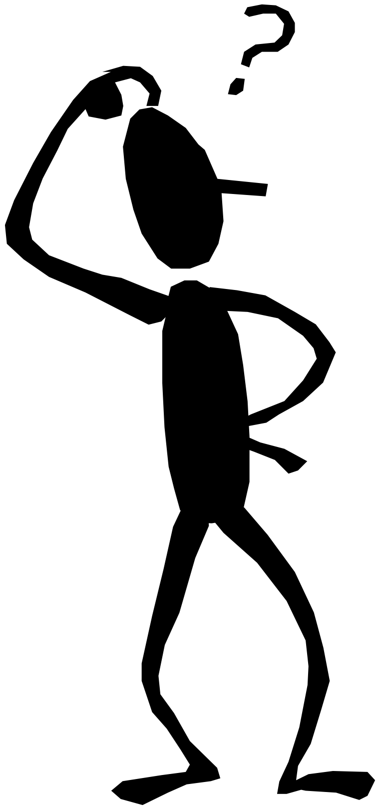
Visual QA
Q: Which shirt is black?

A: Right

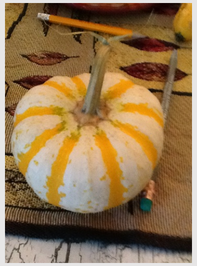


Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?



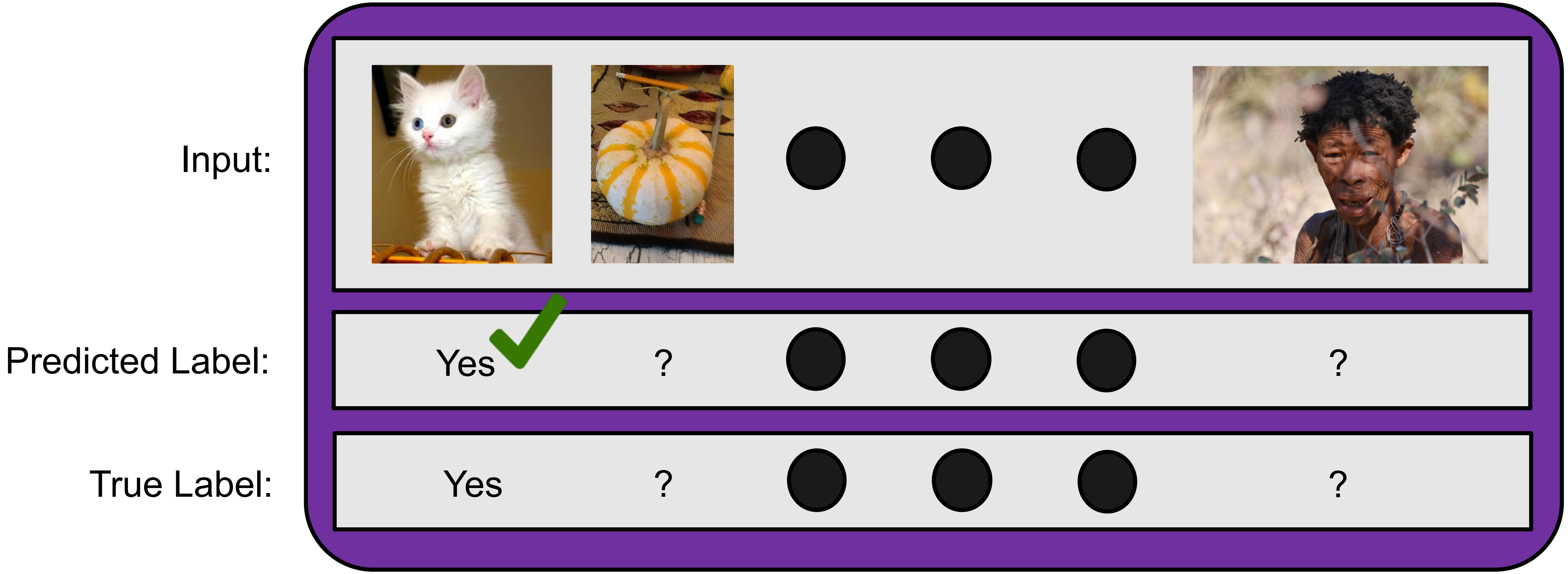
e.g., Is a cat in the image?

+ Input:			●	●	●	
Predicted Label:	Yes	?	●	●	●	?
True Label:	Yes	?	●	●	●	?

Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?

e.g., Is a cat in the image?



Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?

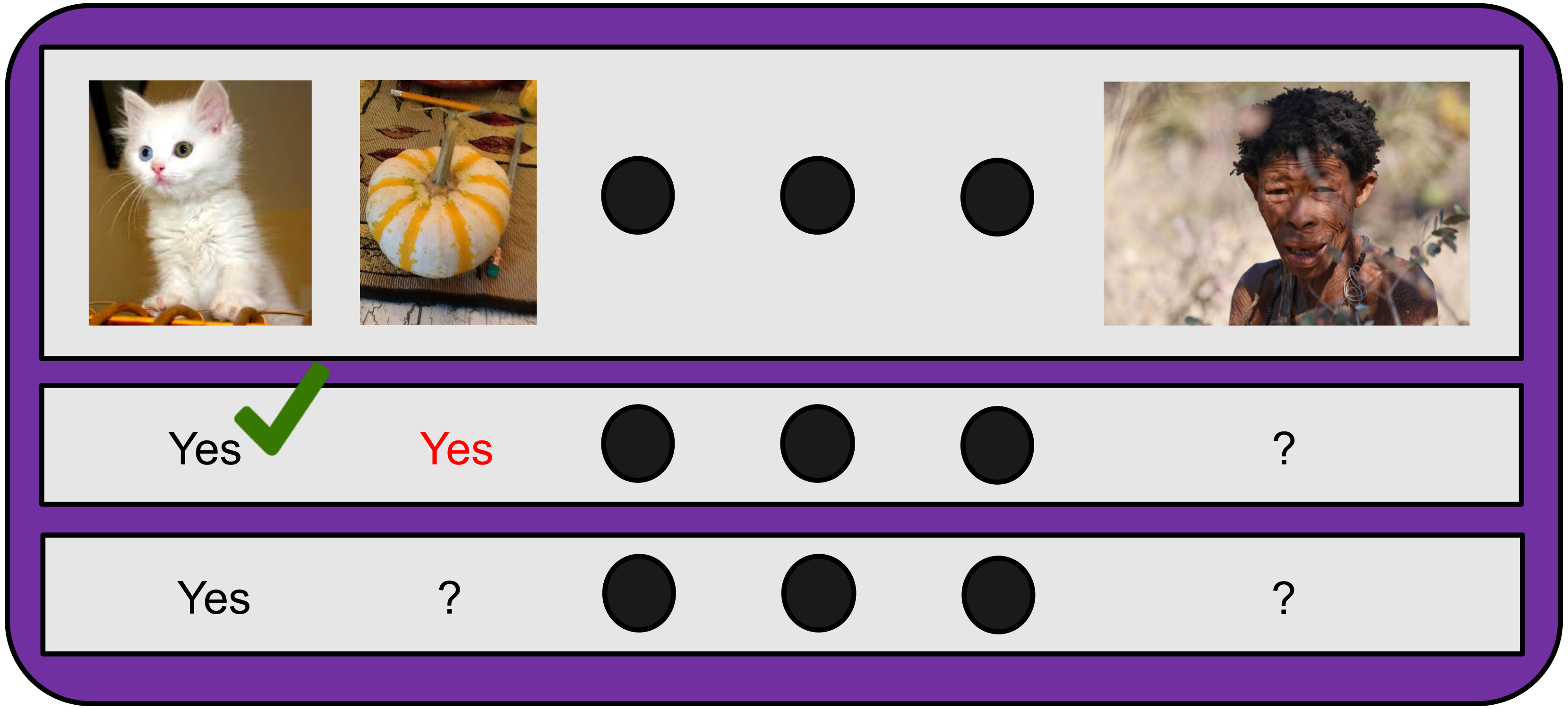
Prediction Model

e.g., Is a cat in the image?

Input:

Predicted Label:

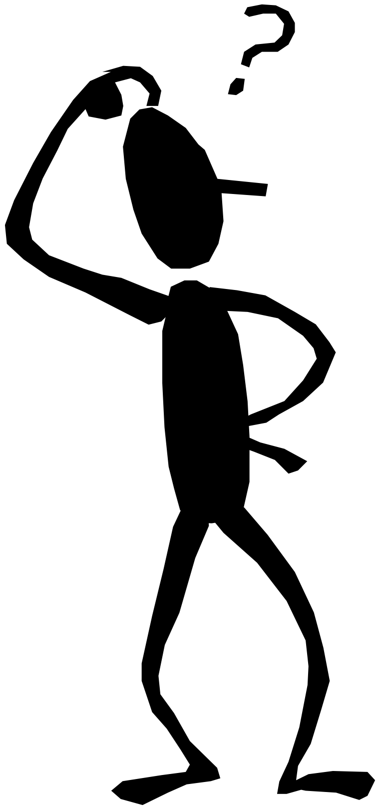
True Label:

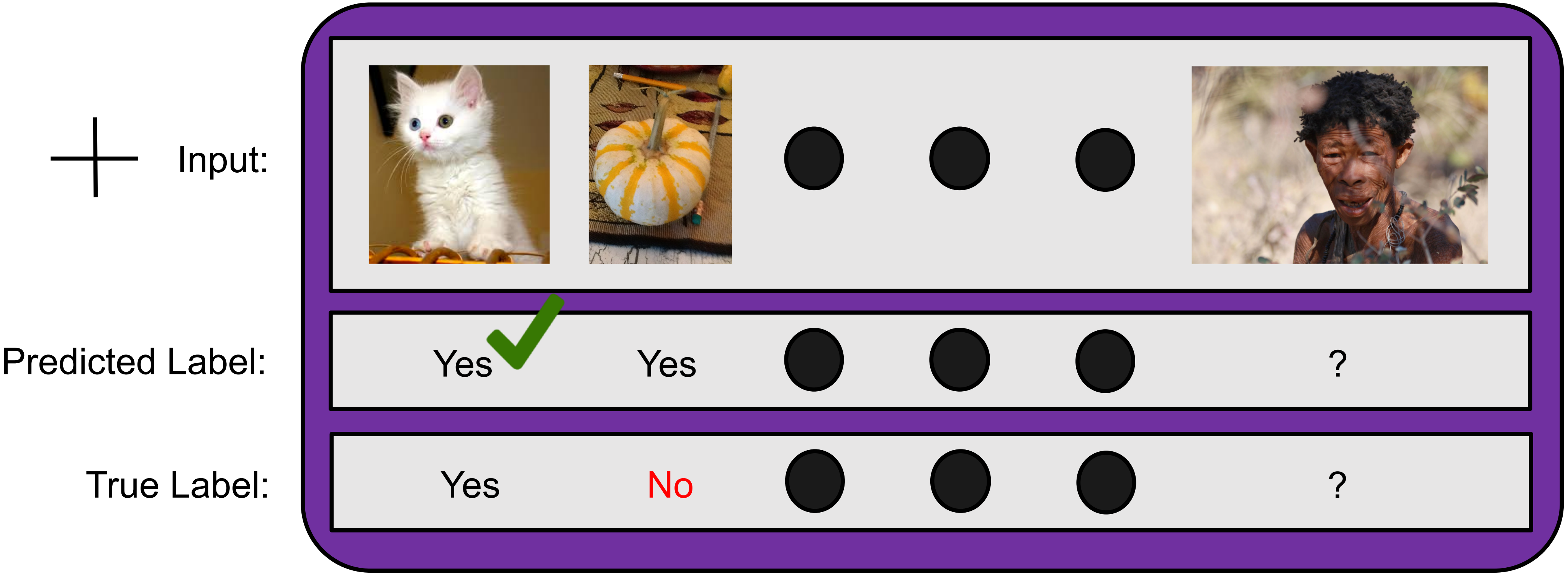


Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?

e.g., Is a cat in the image?

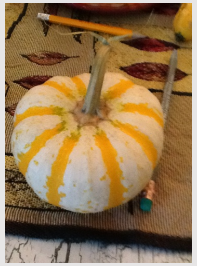




Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?

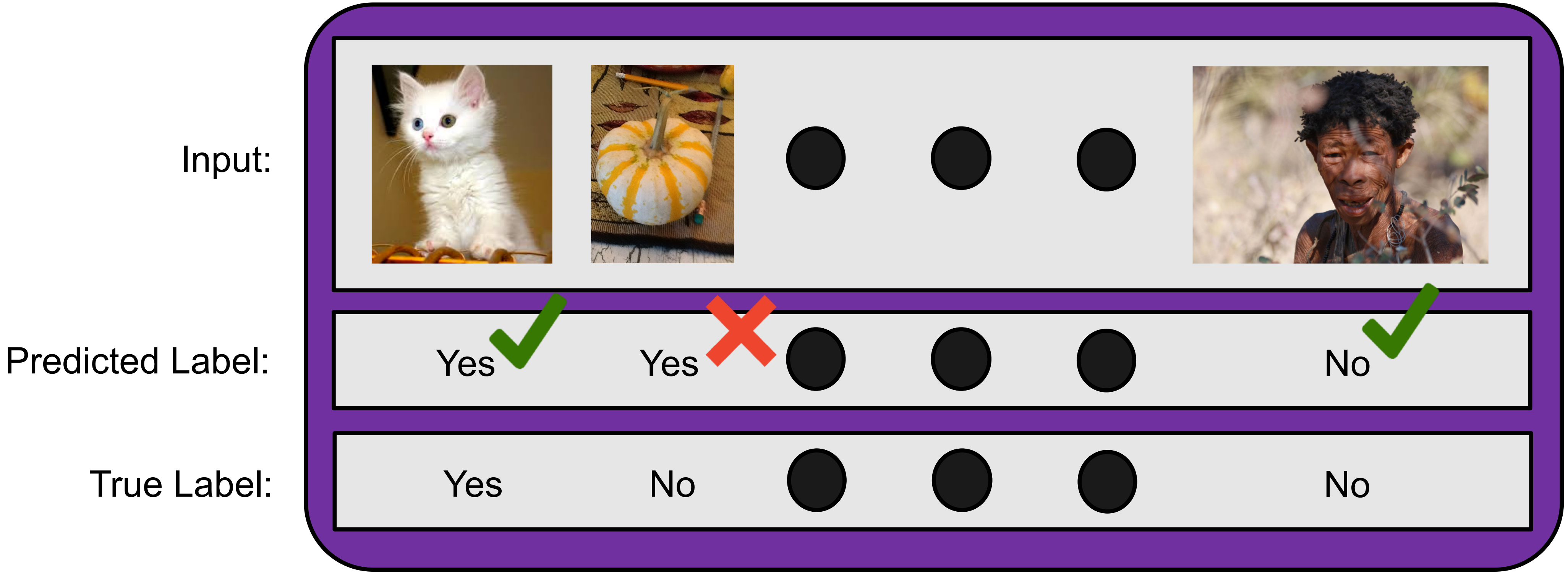
e.g., Is a cat in the image?

Input:			●	●	●	
Predicted Label:	Yes	Yes	●	●	●	?
True Label:	Yes	No	●	●	●	?

Why Bigger Datasets? To Evaluate CV Algorithms

How good is an algorithm?

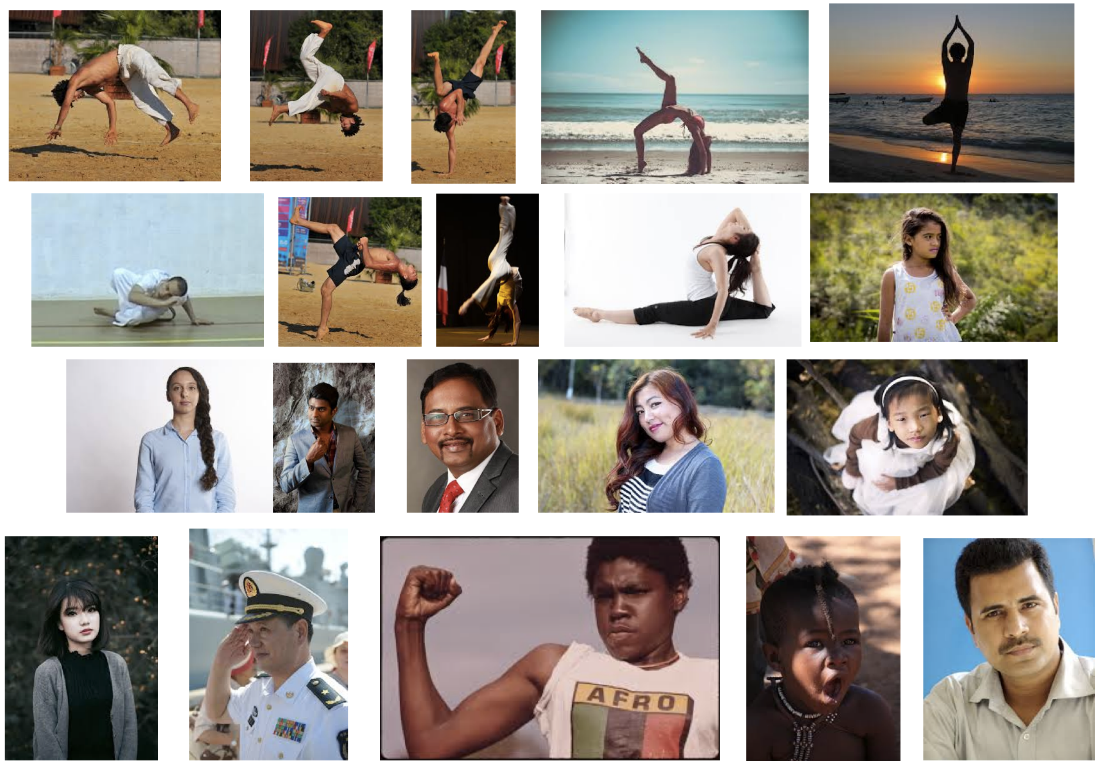
e.g., Is a cat in the image?



Why Bigger Datasets? To Teach CV Algorithms

Typical approach: train using as many labelled examples as possible

e.g., recognize a person



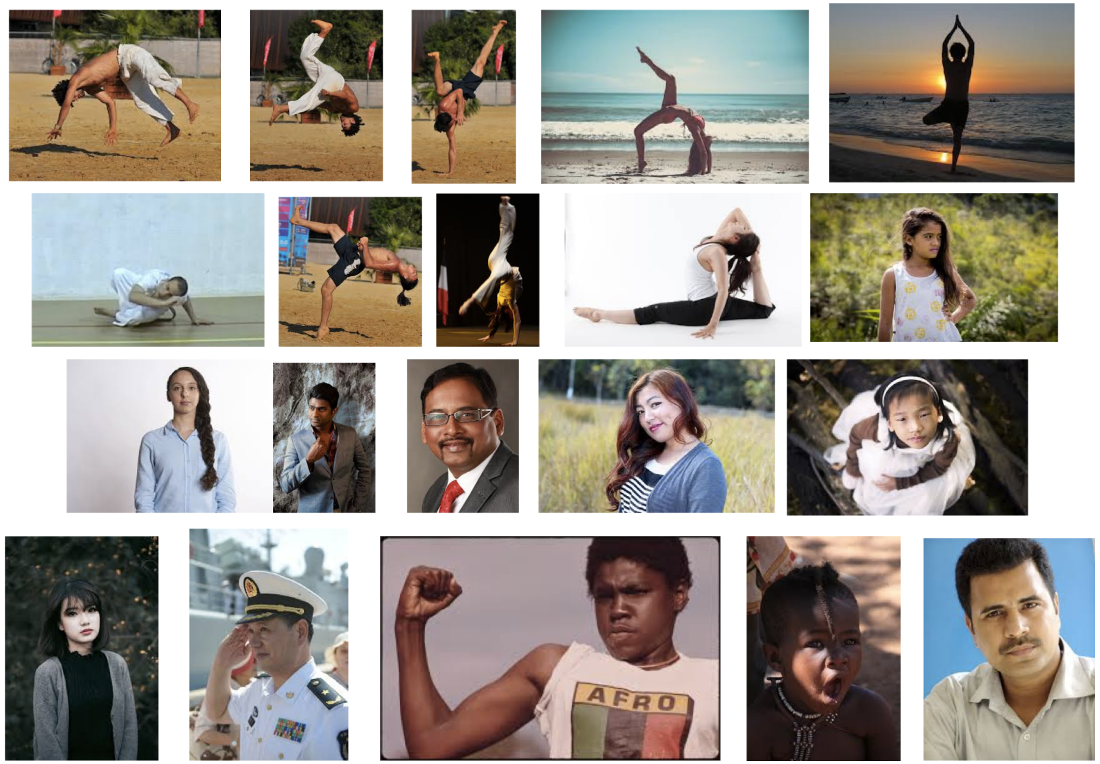
(Intuition: how we teach a person)

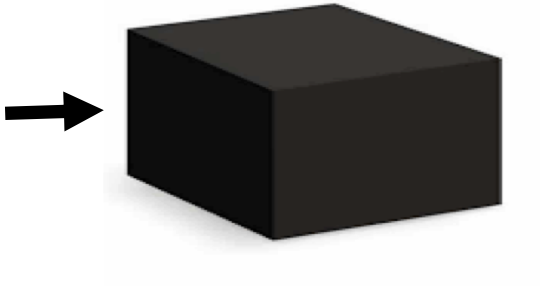
Often people perform better when trained on more examples!

Why Bigger Datasets? To Teach CV Algorithms

Typical approach: train using as many labelled examples as possible

e.g., recognize a person

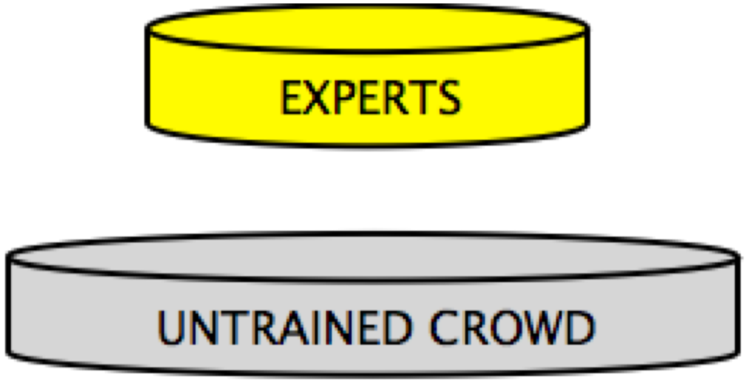




Algorithms also perform better when trained on more examples!

Why Crowdsourcing?

- Crowds offer a scalable, less expensive alternative to the gold standard of "experts" to generate large human-labelled datasets



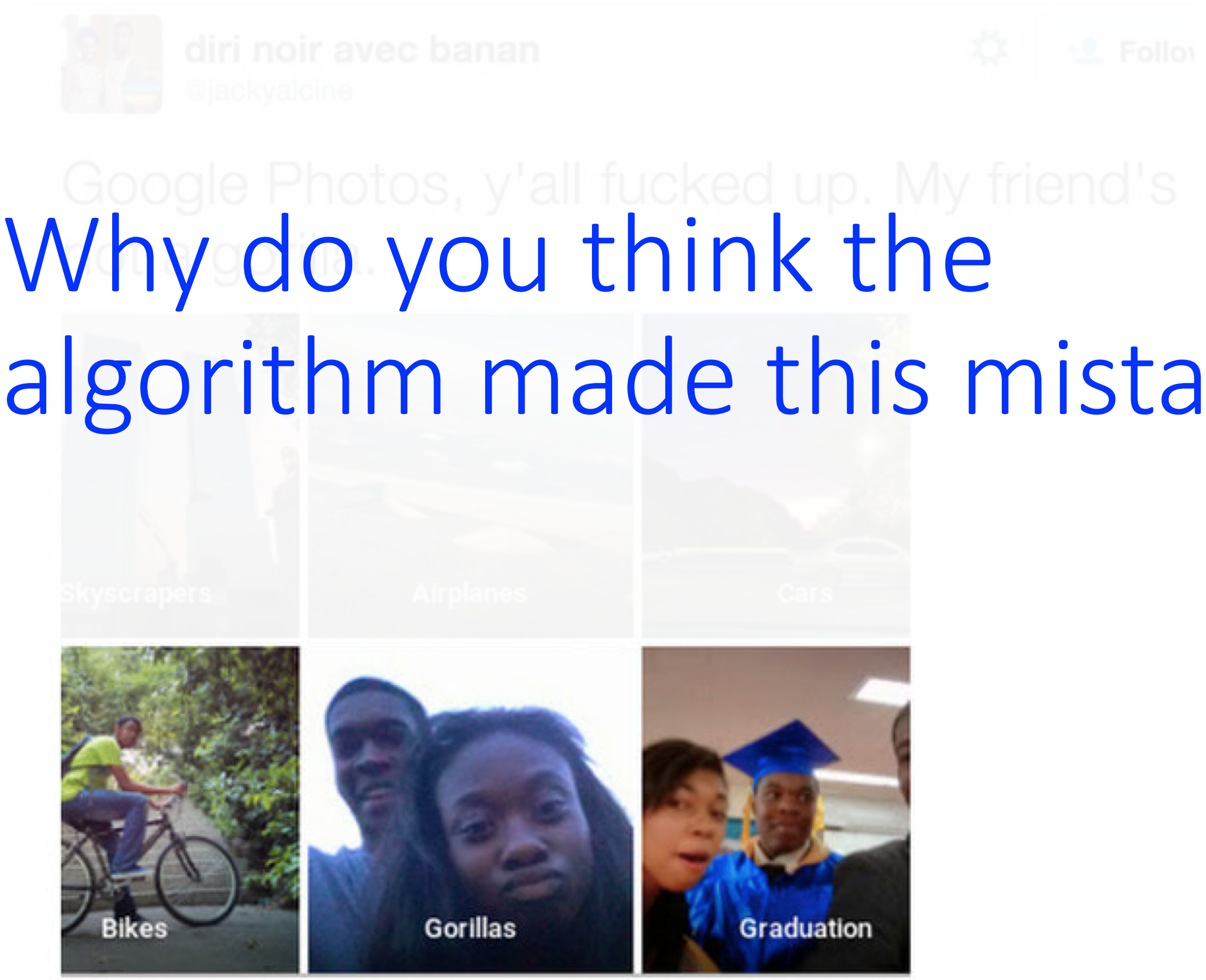
Datasets That Have Led to Unethical AI

Using Twitter to call out Google's algorithmic bias

https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas

Datasets That Have Led to Unethical AI

Why do you think the algorithm made this mistake?

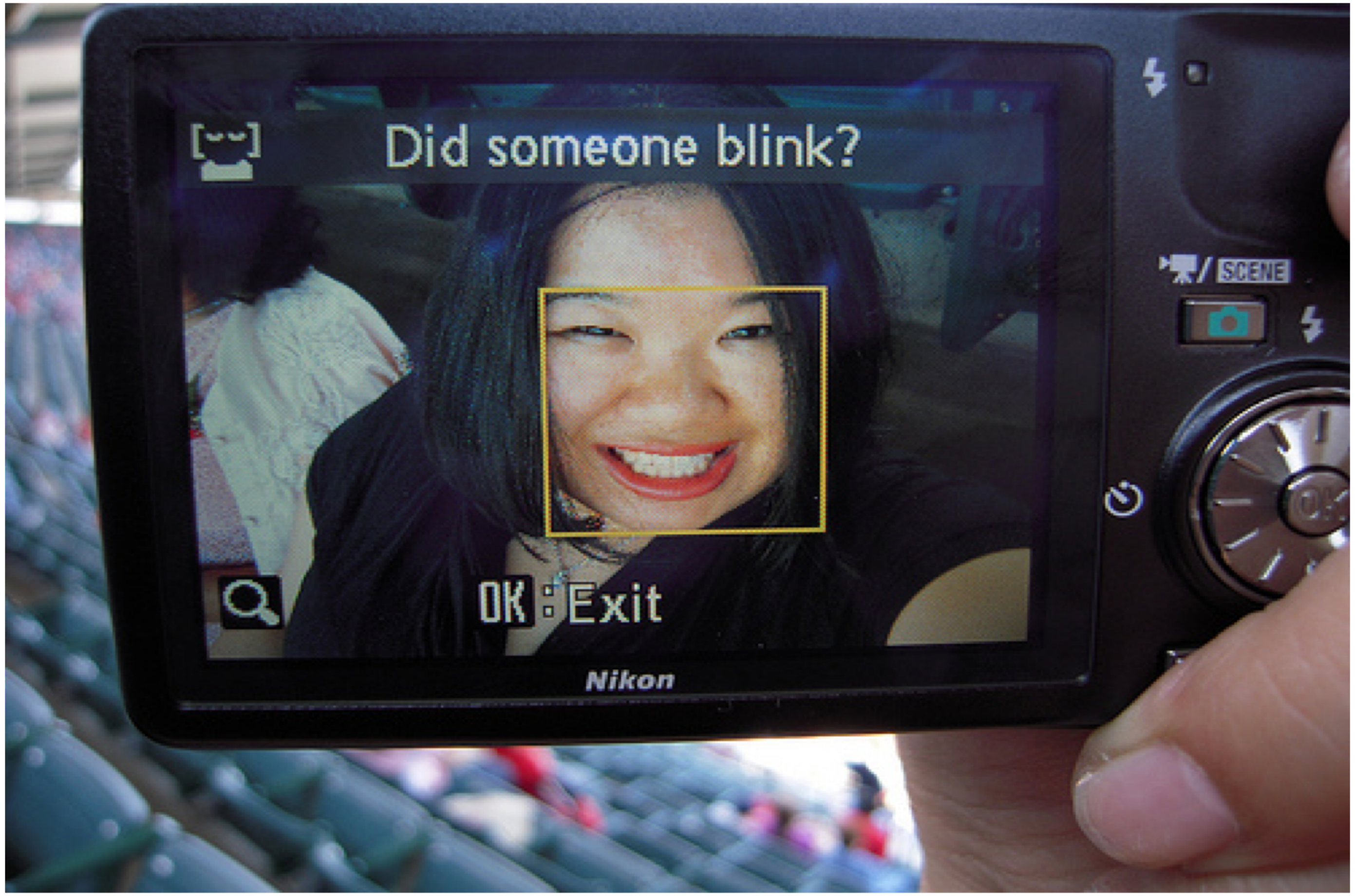


Using Twitter to call out Google's algorithmic bias

https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas

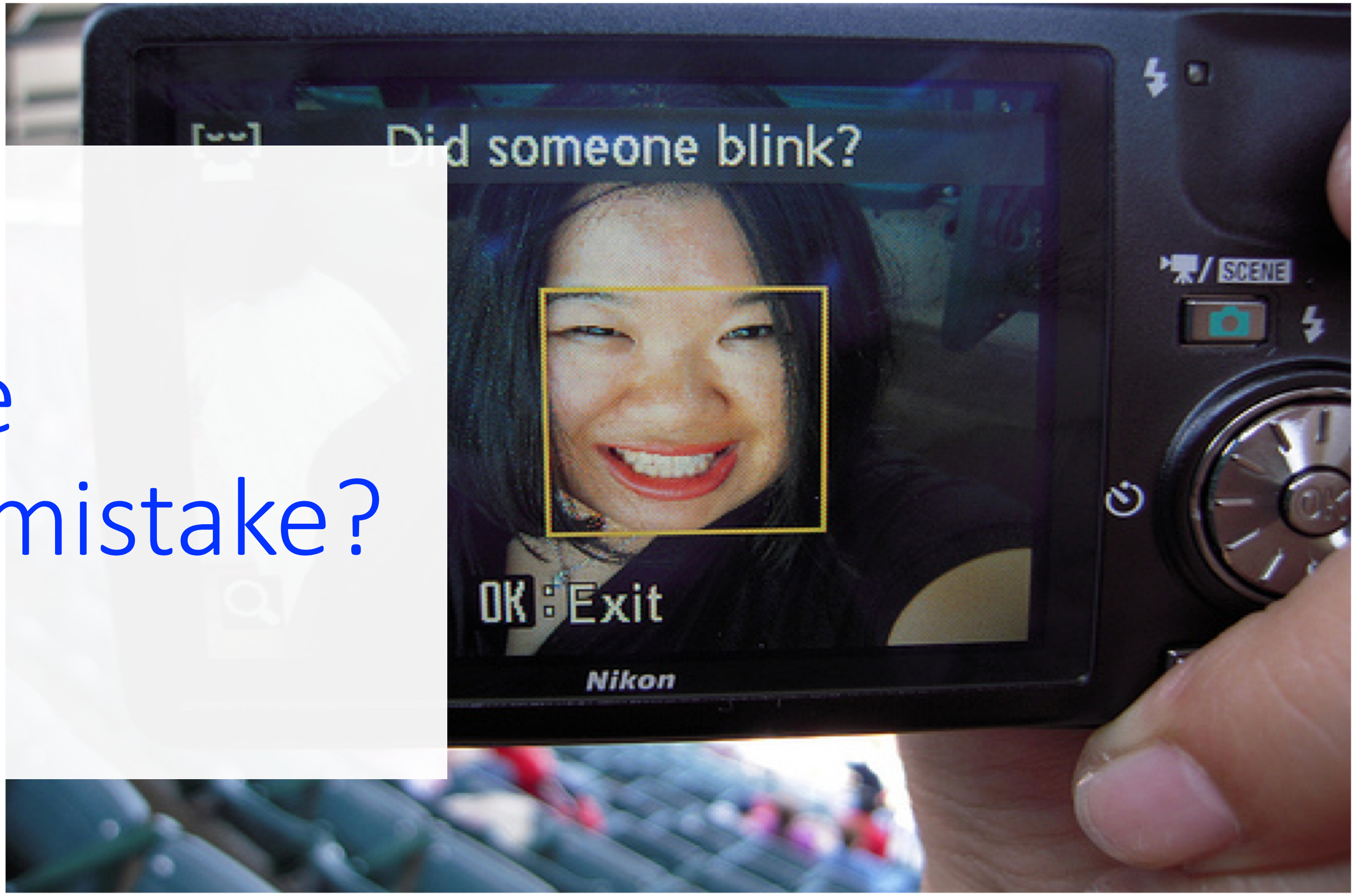
Datasets That Have Led to Unethical AI

Two kids bought their mom a Nikon Coolpix S630 digital camera for Mother's Day… when they took portrait pictures of each other, a message flashed across the screen asking, "Did someone blink?"



http://content.time.com/time/business/article/0,8599,1954643,00.html

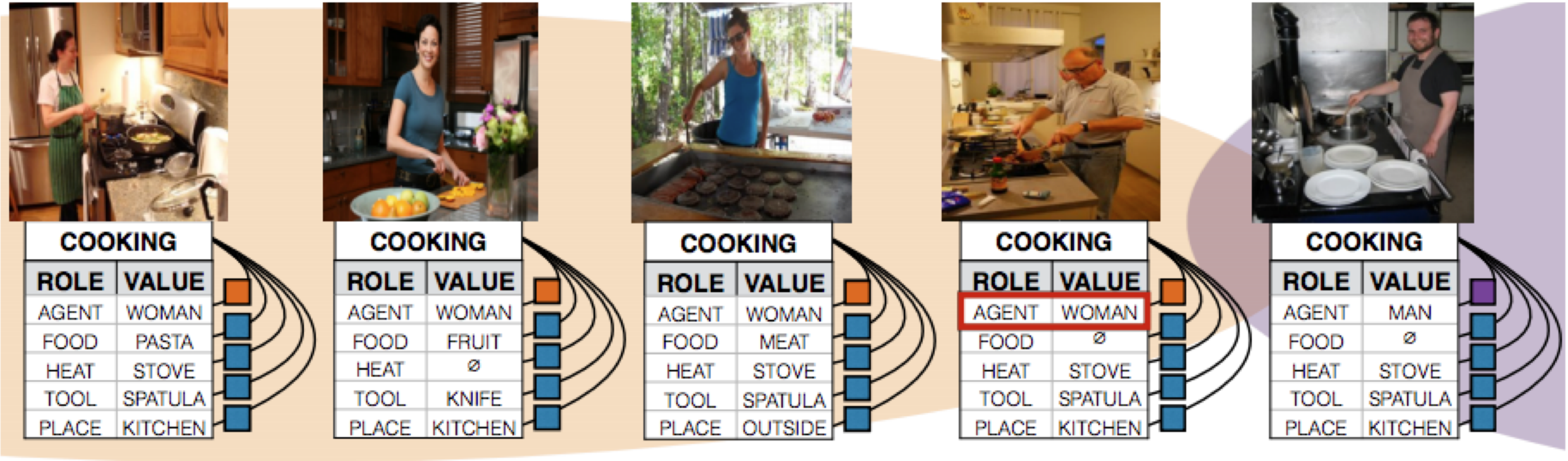
Datasets That Have Led to Unethical AI



Why do you think the algorithm made this mistake?

http://content.time.com/time/business/article/0,8599,1954643,00.html

Datasets That Have Led to Unethical AI



Algorithm identifies men in kitchens as women. Learned this example from given dataset. (Zhao, Wang, Yatskar, Ordonez, Chang, 2017)

https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/ç

Datasets That Have Led to Unethical AI

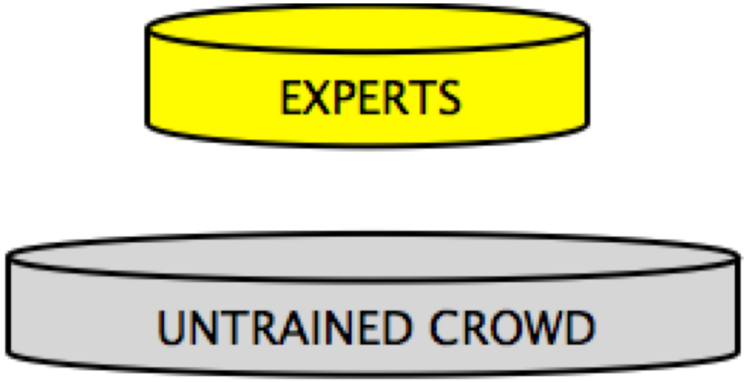
Why do you think the algorithm made this mistake?

Algorithm identifies men in kitchens as women. Learned this example from given dataset. (Zhao, Wang, Yatskar, Ordonez, Chang, 2017)

https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/ç

Crowdsourcing for Computer Vision

- Crowds offer a scalable, less expensive alternative to the gold standard of “experts”



- Crowdsourcing challenge: how to **efficiently** collect **high quality** data from an anonymous crowd that leads to **ethical AI**?

Today’s Topics

- Class logistics
- Introduction to computer vision
- Introduction to crowdsourcing for computer vision
- Lab: web page creation

References

- Nice summaries of the history of computer vision
  - https://hackernoon.com/a-brief-history-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3
  - https://www.epicsysinc.com/blog/machine-vision-history
  - https://sevenshinestudios.wordpress.com/computer-vision-and-deep-learning/
  - https://cds.cern.ch/record/400313/files/p21.pdf
- Slide 12: Image Credits
  - https://www.theverge.com/2018/10/23/18010022/amazon-go-cashier-less-store-san-francisco-location-opens
  - https://en.wikipedia.org/wiki/File:NASA_Mars_Rover.jpg
- Slide 14-22: Image Credit
  - http://www.minnesotavisiontherapy.com/what-is-vision-therapy
- Slide 16: Image Credit
  - https://en.wikipedia.org/wiki/ENIAC#/media/File:Eniac.jpg