

Graduate computer vision

Quick Info

- Instructor - Bharath Hariharan
- Office – 311 Gates Hall
- Lecture venue – Rhodes Hall 571
- Time: Tu / Thu 1:00 – 2:15 pm
- OH:
 - Bharath: M / F 1:30 – 3:00 pm
- Course web page: <https://www.cs.cornell.edu/courses/cs6670/>

Course Overview

What graduate introduction means

- This is a course about computer vision + research
- So either:
 - You want to do research in computer vision, or
 - Computer vision is relevant to your research
- If you are not interested in research of any kind, consider the undergraduate version, 4670 (offered in spring)

What I assume you know / can figure out

- Math
 - Linear algebra
 - Calculus
 - Probability and statistics
- Programming
 - Typically Python, although other languages are an option
 - I'll assume you can read tutorials/docs and figure out libraries like pytorch / tensorflow

What you will learn

- Check out Learning Outcomes in course page for full list
- We will talk / learn about
 - Recognition
 - Reconstruction
 - Embodied cognition
 - Synthesis
- Technical content will focus on evolving state-of-the-art

What you will do

- Check out Deliverables in course webpage
- **December 5:** A one pager that answers the following two questions for each of the four sections of the course: Recognition, Reconstruction, Synthesis and Embodied vision
 - What are the open research problems, namely, things that current state-of-the-art cannot do?
 - What are the technical challenges in solving these problems? Brainstorm about possible solutions.
For both of these, be creative! More points for thinking out of the box.

What you will do

- Deliverable 2: A project with three deliverables:
 - **October 7:** A one sentence project idea
 - **November 11:** A two-page project proposal that contains:
 - Introduction that motivates the particular problem you are working on
 - Related work that clearly describes what has been done and what is missing in prior work
 - A section describing your approach and how it addresses the limitations of past work
 - **December 5:** A one-pager final result
 - A preliminary result showing evidence your approach might be successful.

What you will be do

- Deliverable 3: **November 22**: Peer review for 2 project proposals from your peers (Papers will be assigned to reviewers by **November 15**). This peer review should answer:
 - Is the problem well defined?
 - Does the related work clearly identify the holes in the prior work?
 - Does the proposed approach address the limitations of past work?
 - Do you agree with the authors conclusions from the preliminary experiments?
 - What further experiments and modifications to the approach would you suggest the authors do?

Doing research: choosing a
research project

The Heilmeier Catechism

- What are you trying to do? Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you are successful, what difference will it make?
- What are the risks?
- ~~How much will it cost?~~
- ~~How long will it take?~~ Will it be finished in time?
- What are the mid-term ~~and final~~ “exams” to check for success?

Examples of projects

- Solve an existing problem, but better
- Example: using covariance of feature maps for fine-grained recognition
- <http://vis-www.cs.umass.edu/bcnn/>
- But aim for a proof of concept, e.g., on a small dataset with small models

Examples of projects

- Define a new problem
- E.g., can you draw interpolate between two different faces?
- <https://grail.cs.washington.edu/cflow/>
- (Again, aim for a proof of concept)

Examples of projects

- Use computer vision to do a research project in your area.
- Standard: what is the equivalent of a workshop paper in your area?
- Example: can we detect cell organelles in microscopy

Examples of projects

- Evaluate problems and existing solutions in different ways
- E.g., evaluate accuracy of face attribute detection systems on faces of color.
- <http://gendershades.org>
- Example 2: evaluate segmentation algorithms for how well they get object boundaries.

Computer vision overview

What is computer vision?

- Getting a machine to "see" like humans
- But "see" what? What input and what output
- Input: Images, or visual data
 - Typically captured by a camera
 - In principle can also include satellite data, microscopes, go beyond the visible spectrum etc.
- Output: Understanding (?)

Output of computer vision technology

- *Recognition*: Abstract concepts



Barack Obama

Joe Biden

Window

Cupcake

Couch

Output of computer vision technology

- *Reconstruction*: Physical properties



What is the shape of each object in the scene?

What color is each object?

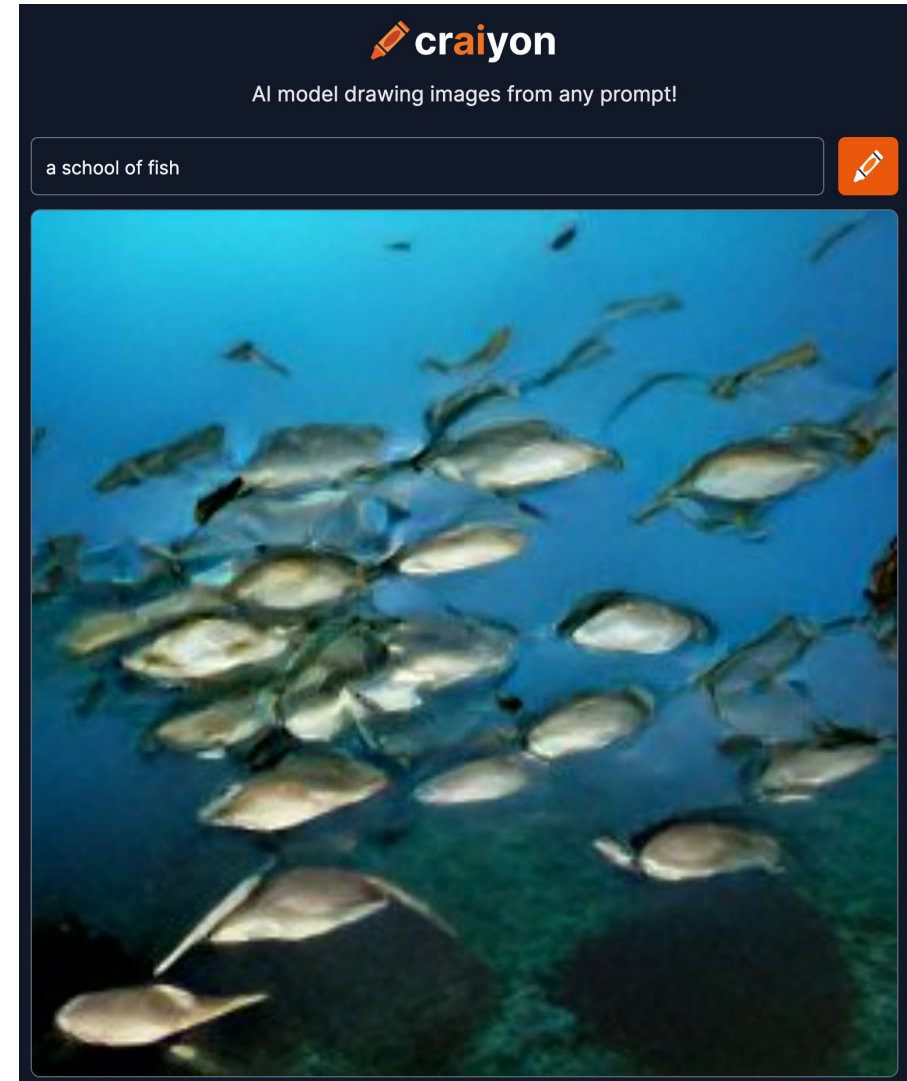
What material is it made of?

Where is the light coming from?

Where is the camera?

Output of computer vision technology

- *Synthesis*: generation of new images
- “A school of fish”
- Generate images
- Stylize images
- Edit images



Output of computer vision technology

- *Embodied vision*
- Perception for robots
 - How do we connect perception to action
 - How can action help for perception.

Introduction to Recognition

What is recognition?



Funes the memorious

“It was not only difficult for him to understand that the generic term *dog* embraced so many unlike specimens of differing sizes and different forms; he was disturbed by the fact that a dog at three-fourteen (seen in profile) should have the same name as the dog at three-fifteen (seen from the front)”

- Jorge Luis Borges

Where do labels come from?



Email: woodworldhv@gmail.com



Where do labels come from?



Where do labels come from?



Email: woodworldhv@gmail.com



Where do labels come from?



Where do labels come from?



Email: woodworldhv@gmail.com



Basic/sub-ordinate/super-ordinate category recognition

- What would you label this image?



Basic/sub-ordinate/super-ordinate category recognition

- What would you label this image?



Living thing

Animal

Bird

Warbler

“Robin”

Basic/sub-ordinate/superordinate category recognition

- Basic category: First thing that comes to mind
- Usually consistent across multiple people
- Subordinate categories: fine-grained recognition
- Super-ordinate categories : coarser grained recognition
- So not all classifications are equal

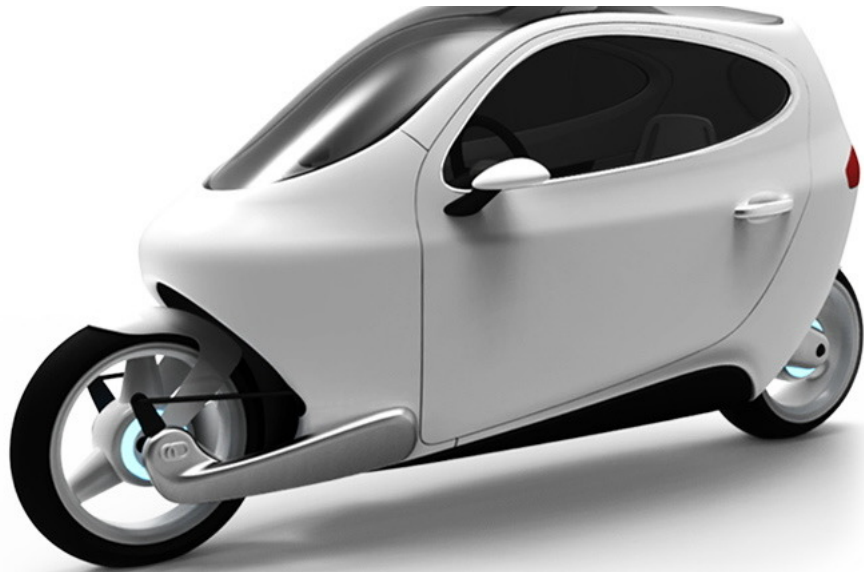
What is a category?

- Fixed set of attributes?
- A dog is a furry animal with four legs and tail and a snout
- Really?
- Not clear people's categories have precise specifications



Class boundaries are fuzzy

Is this a car?



The need for ML

- Classes are difficult to specify
- And have fuzzy boundaries
- Can only ask people to provide labels by example
- Hence specify classes with a training set
 - Really old idea

The need for ML

- We want to perform an operation but don't know how to specify it in code
 - Either known but too complex (e.g., some physical quantity)
 - Or unknown (e.g., “dog”, “cat”)
 - Or fuzzy (e.g., “artistic”)
- Optimize an *approximation* using a training dataset
- Can never guarantee that approximation will be exactly correct (no matter how high the test set accuracy!)

ML: Assumptions and guarantees

- Key assumption: training distribution is similar to the test distribution
- Usually very difficult to get this
 - Especially: robotics
- Can provide only population-level (statistical) guarantees

Why recognition?

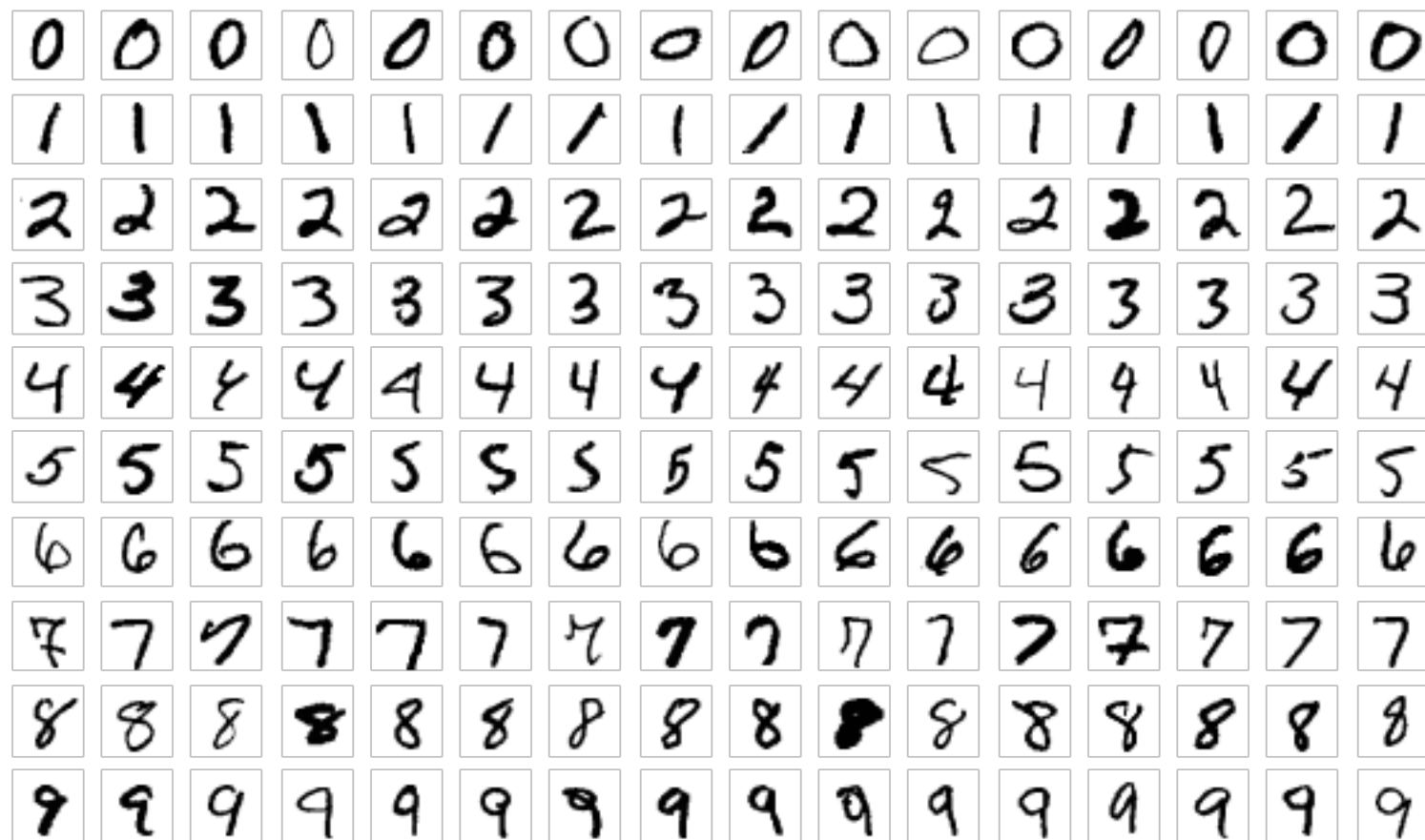
- Humans can do it very well
- Lots of different applications

Early motivating applications I - robotics

- Robots need to identify objects
- Often *instance recognition* – need to identify individual objects and their pose

Early motivating applications II - document recognition

- Automatically parse checks and letters
- Sort envelopes
- Need to recognize digits and letters



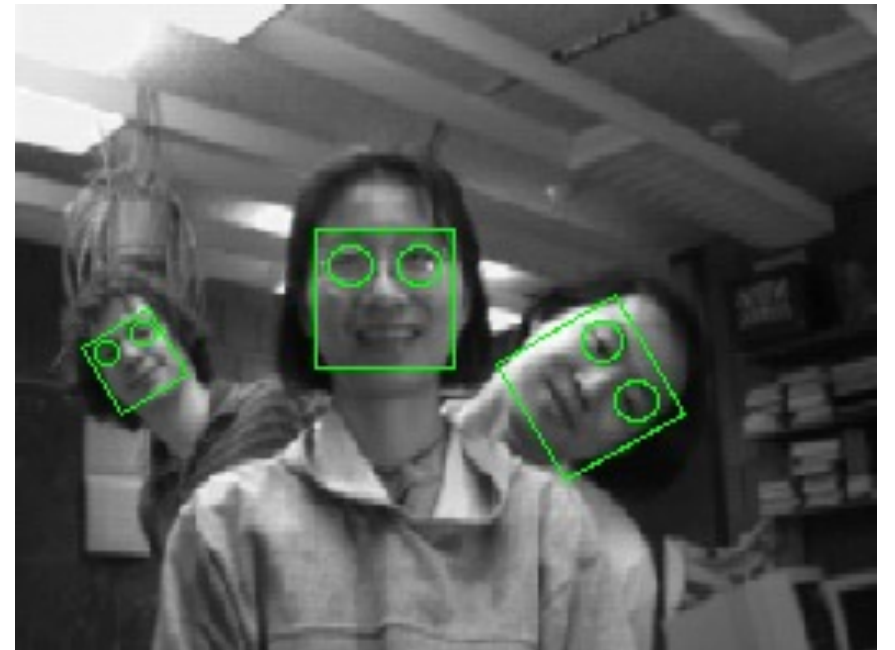
MNIST

Early motivating applications III - Surveillance

- Detect and recognize faces and people
- Important ethical considerations on this

Early applications IV

- Pedestrian detection deployed in smart-assist driving systems
- Face detection deployed in cameras since the 2000s



Thinking through the ethics of recognition

- Why?
 - What is the end task / application? Is it worth doing?
 - Does it even make sense?
 - Who is it designed for?
 - Who will benefit and who will be harmed?

The case of face recognition

Wrongfully Accused by an Algorithm

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

New York Times, June 24, 2020

<https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>



The case of face recognition



The New York Times



The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and “might lead to a dystopian future or something,” a backer says.



The case of face recognition

Maryland's face recognition system is one of the most invasive in the nation | COMMENTARY

By JAMESON SPIVACK
FOR THE BALTIMORE SUN | MAR 09, 2020 | 5:20 PM

Twitter Facebook Share



ADVERTISEMENT

SQUARESPACE
AWARD-WINNING
WEBSITE DESIGNS
FOR BUSINESS
OWNERS

START YOUR FREE TRIAL TODAY
14-DAY TRIAL. TERMS APPLY.

LATEST OP-ED

OP-ED

Faith leaders: Gov. Hogan is opening churches too soon,



Slide credit: Timnit Gebru, Emily Denton

<https://sites.google.com/view/fatecv-tutorial/schedule>

Potential for who?

*Maryland has a complicated history with face recognition. Many praised it after it was used to **identify the Annapolis Capital Gazette shooter**. On the other hand, police in Baltimore County also used face recognition on social media photos to identify people at the **Freddie Gray protests** and target them for **unrelated arrests**. Using face recognition to surveil people at protests and rallies — activities protected by the First Amendment — discourages political participation.*



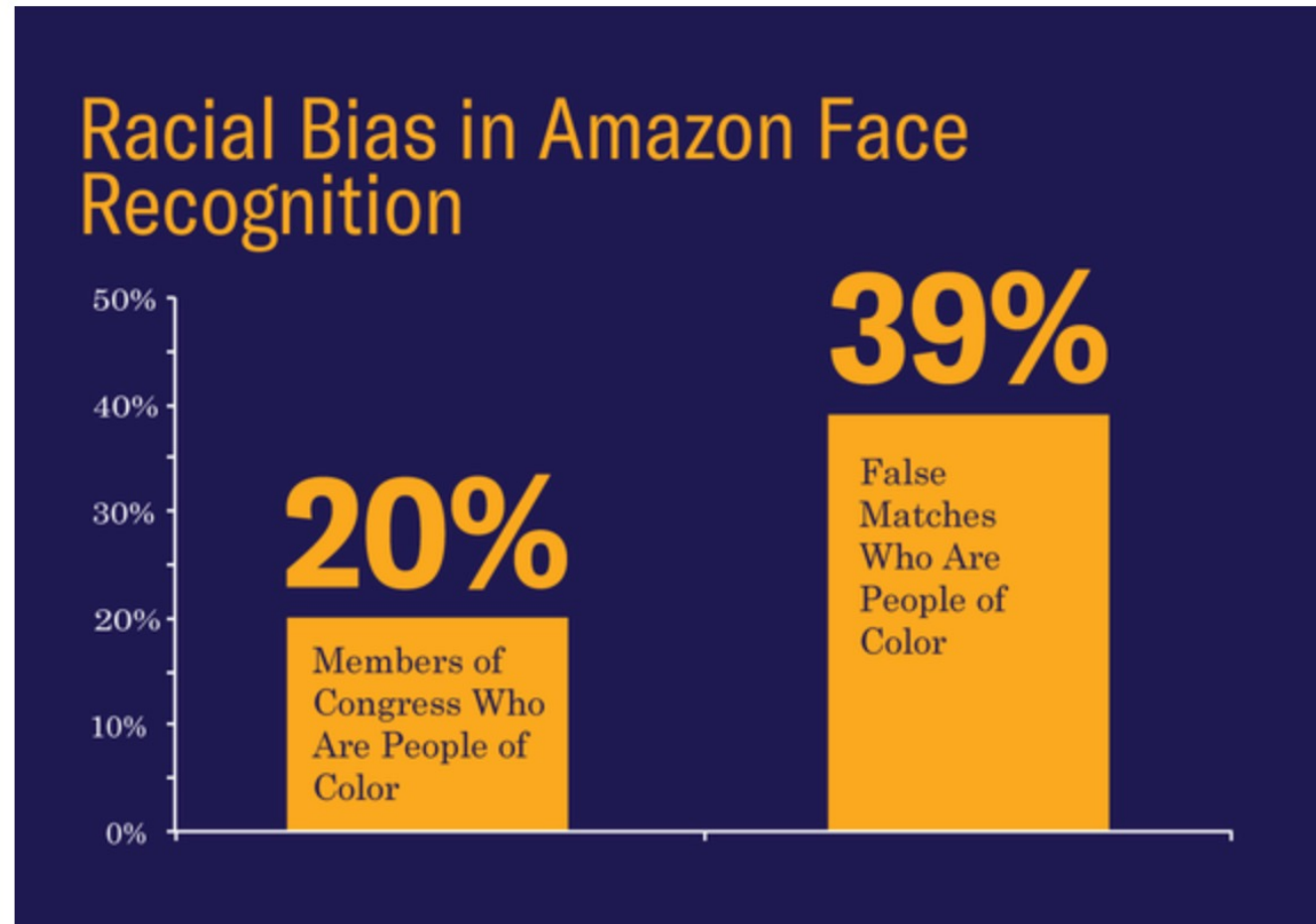
- Other problems with face recognition in policing: using it to match *sketches* rather than actual faces [1].

1. Ruha Benjamin. *Race after technology*

Slide credit: Timnit Gebru, Emily Denton

<https://sites.google.com/view/fatecv-tutorial/schedule>

The case of face recognition

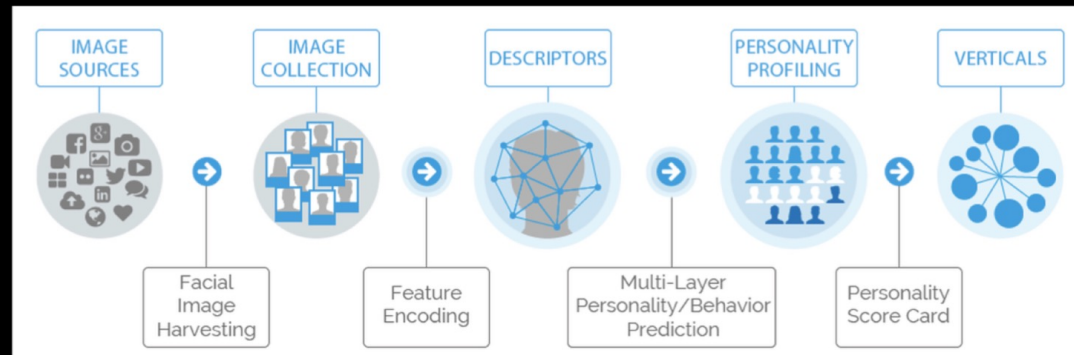


People of color were disproportionately falsely matched in our test.

The case of face analysis

“Faception is first-to-technology and first-to-market with proprietary computer vision and machine learning technology for **profiling people** and revealing their personality **based only on their facial image.**”

- [Faception](#) startup



“High IQ”

“White-Collar Offender”

“Terrorist”







Slide credit: Timnit Gebru, Emily Denton

<https://sites.google.com/view/fatecv-tutorial/schedule>

Who is seen? How are they seen?

Error Rate_(1-PPV) By Female x Skin Type



	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TYPE VI
	1.7%	1.1%	3.3%	0%	23.2%	25.0%
	11.9%	9.7%	8.2%	13.9%	32.4%	46.5%
	5.1%	7.4%	8.2%	8.3%	33.3%	46.8%

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini



Not just about bias

- Privacy
- Safety
- Security
 -

What does this mean for CV research?

- Research is not “value neutral”
 - Value judgements implicit not just in choice of problem, but also in choice of data, evaluation metrics and even model choices
- Existing areas of focus are based on who is framing the problem
 - There’s a diversity crisis in AI
 - We don’t value perspectives of those who may be marginalized / harmed
- Technical fixes alone cannot solve the problem
 - We don’t value interdisciplinary / “social science” work

Face recognition to analysis

- Face attribute recognition
- Offered by Microsoft, Amazon etc.
- E.g., Gender recognition
- Questions:
 - Is this an application we want to build / enable?



Woman, smiling, blonde

Case study 2

- Diagnosing chest X rays
- Question:
 - Is this an application we want to enable?
 - What level of accuracy do we require?
 - What happens if there is an error?



Case study 3

- Self-driving car application: car with a camera on it
- Where are cars in the image?
- Questions:
 - Is this an application we want to build / enable?
 - What form should the output take?
 - What happens if there is an error?
 - What level of accuracy do we require?



Case study 4

- Alt text for the visually impaired
- Questions
 - What kind of output is correct?
 - How do we measure correctness?

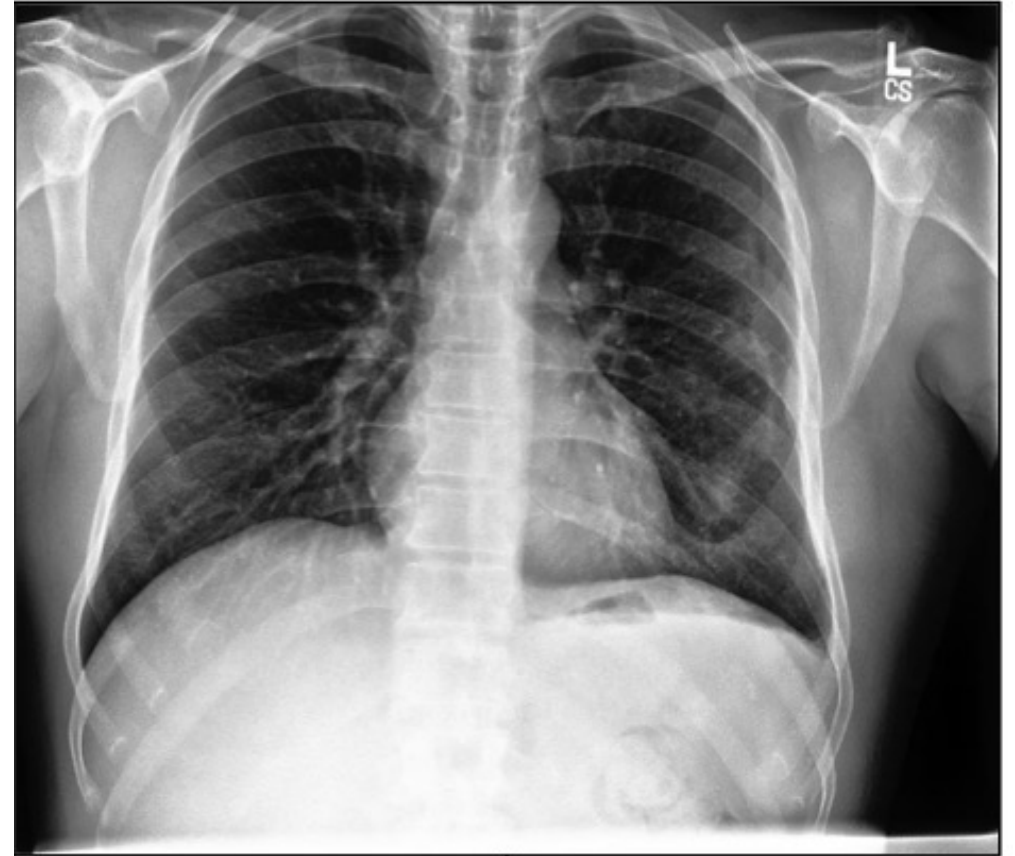


A person in a suit riding a green car with a bicycle on the back

Description automatically generated with low confidence

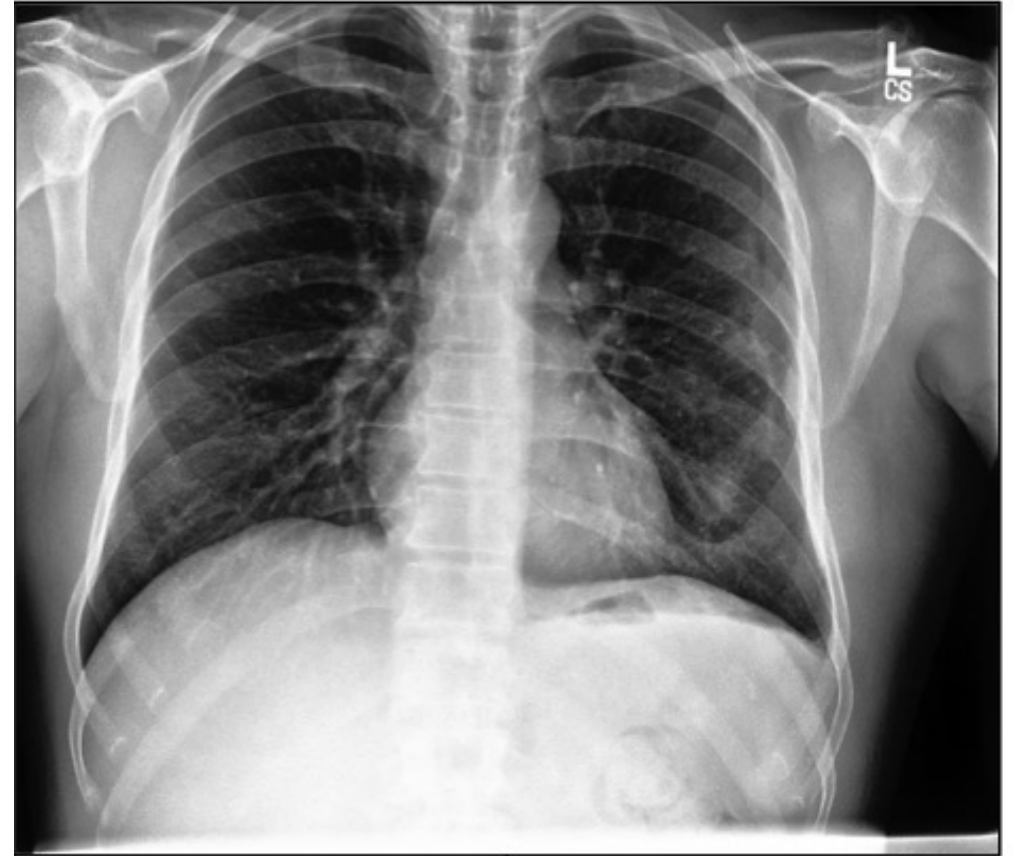
Choosing the right evaluation metric

- What should a metric measure?
- Suppose you are building system to detect Covid in chest x rays
- Accuracy = $P(\text{pred. label} == \text{true label})$
- Accuracy of candidate system = 95%
- Is this good?



Choosing the right evaluation metric

- What should a metric measure?
- Two kinds of errors:
 - False positives: $y_{true} = 1, y_{pred} = 0$
 - False negatives: $y_{true} = 0, y_{pred} = 1$
- Which option is good?
 - High false positive rate, close to 0 false negative rate
 - High false negative rate, close to 0 false positive rate.

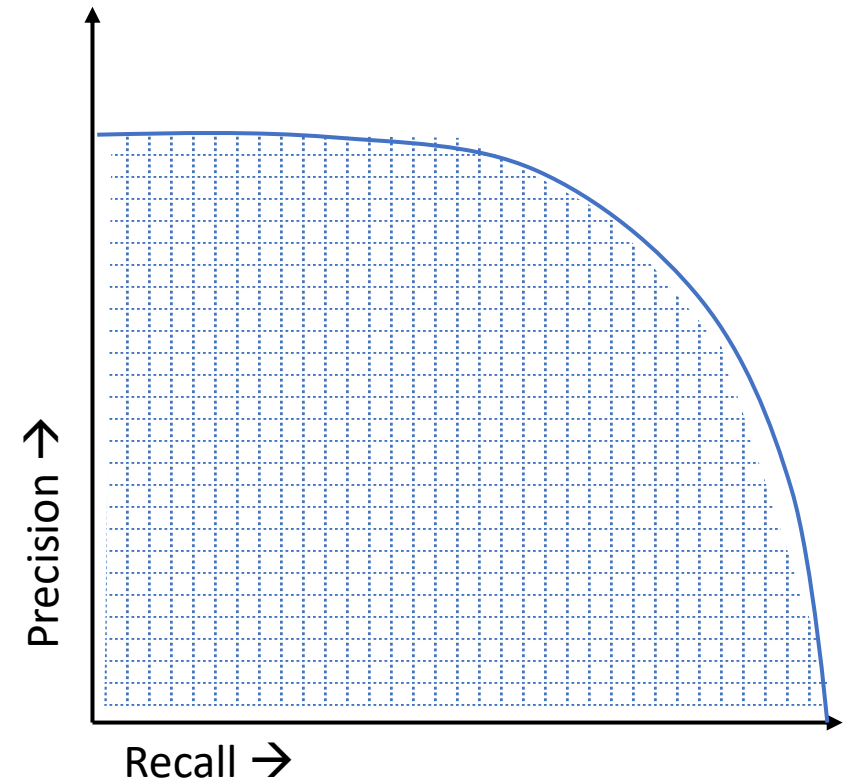


Choosing the right evaluation metric

- Precision = (True positives) / (True positives + False positives)
 - “When the system declares a positive, how often is it correct?”
- Recall = (True positives) / (True positives + False negatives)
 - “When a data point is in fact a positive, how often is it detected by the system?”

Choosing the right evaluation metric

- Systems typically have a way of trading off precision vs recall
- Precision – recall curve
- One measurement: area under PR curve (Average precision or AP)
- But not necessarily interpretable

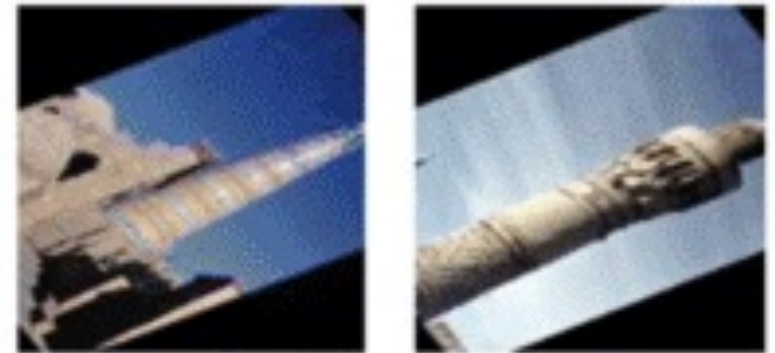


Choosing the right evaluation metric

- How do the errors break down otherwise?
- By race, gender etc.?
- Other kinds of conflating variables?
- Can we tell when the system is likely to be wrong?

Benchmarks

- Benchmark = Test data + evaluation metric
- Questions:
 - Where does the data come from?
 - Is there bias? Spurious correlations?
- Benchmarks serve dual purpose:
 - Measure progress
 - Inspire community, but:
- “When a measure becomes a target, it ceases to be a good measure.”
 - Marilyn Strathern (generalizing Goodhart’s law)



Minaret

Who is seen? How are they seen?

Training data: 33% of cooking images have man in the agent role
Model predictions: 16% cooking images have man in the agent role

The figure displays five images of people cooking, each with a corresponding table of roles and values. The tables are connected to the images by lines. The fourth table has a red box around the 'AGENT' and 'WOMAN' entry.

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	PASTA
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	FRUIT
HEAT	∅
TOOL	KNIFE
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	MEAT
HEAT	STOVE
TOOL	SPATULA
PLACE	OUTSIDE

COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	∅
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN

COOKING	
ROLE	VALUE
AGENT	MAN
FOOD	∅
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN

[Zhao et al. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints]

[Hendricks et al. Women also snowboard: Overcoming bias in captioning models.]

Slide credit: Timnit Gebru, Emily Denton

<https://sites.google.com/view/fatecv-tutorial/schedule>



The true/correct benchmark

- Did it actually help / work?

Artificial intelligence / Machine learning

Hundreds of AI tools have been built to catch covid. None of them helped.

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

by **Will Douglas Heaven**

July 30, 2021

Typical issues that plague deployment

- Images seen during deployment are very different: *domain shift*
- Meaning of classes etc. change: *concept drift*
- Unforeseen circumstances, e.g., new classes: *open world*
- Systems often used not as intended, e.g., output assumed to be unbiased
- ...

Typical issues that plague deployment

Original data



Open world



Domain shift



Concept drift

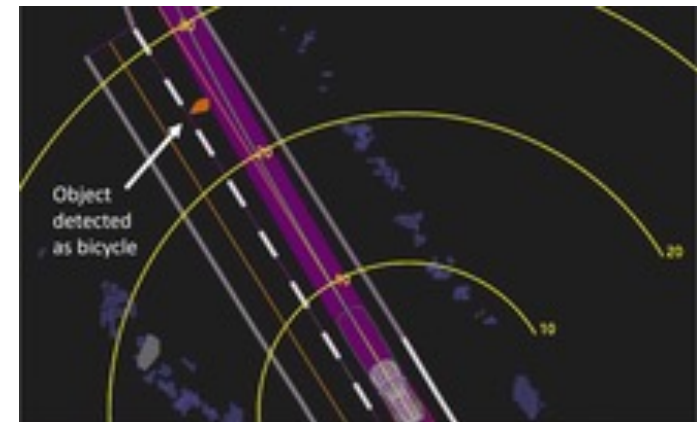


Typical issues that plague deployment

Death of Elaine Herzberg

From Wikipedia, the free encyclopedia

The **death of Elaine Herzberg** (August 2, 1968 – March 18, 2018) was the first recorded case of a pedestrian fatality involving a [self-driving car](#), after a collision that occurred late in the evening of March 18, 2018. Herzberg was pushing a bicycle across a four-lane road in [Tempe, Arizona](#), United States, when she was struck by an [Uber](#) test vehicle, which was operating in self-drive mode with a human safety backup driver sitting in the driving seat. Herzberg was taken to the local hospital where she died of her injuries.^{[2][3][4]}



Using ML: where to get data?

- Uncurated, in-the-wild?
- Yes
 - This will match training domain to test domain
 - ML only works in this scenario
- Not necessarily
 - In-the-wild data has problematic biases, ML may accentuate this
- Issues surrounding consent
 - Did people agree for their data to be used?

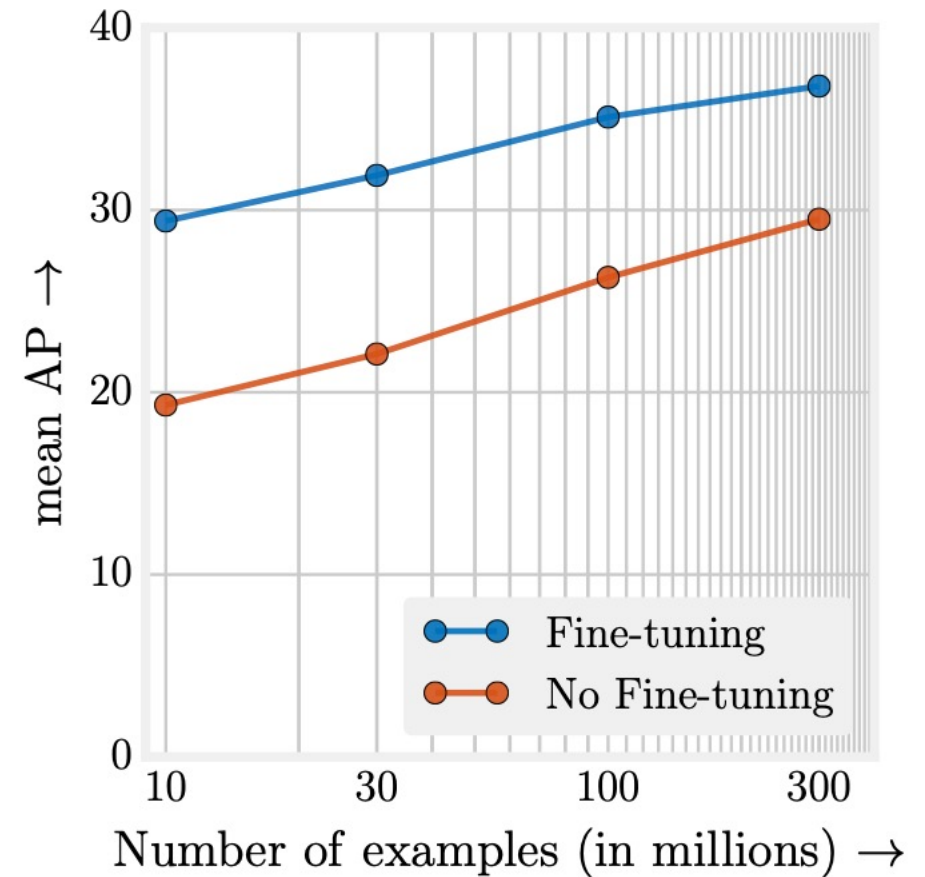
The state of datasets today

- Tiny images
 - 80 million images,
 - Removed due to problematic biases
- ImageNet – 1K
 - 1 million images
 - Has problematic biases
 - Labels not guaranteed to be correct
- MS – Celeb A
 - Face dataset
 - Removed due to potential biases
- *Curation / documentation debt*¹

1. Butters, O., Wilson, R.C., & Burton, P. (2020). Recognizing, reporting and reducing the data curation debt of cohort studies. *International Journal of Epidemiology*, 49, 1067 - 1074.

Using ML: where to get labels

- ML requires not just data but labels
- Usual solution: Amazon mechanical turk
 - Typically less than minimum wage
- Modern solution: labeling companies (e.g., scale.ai)
- Alternatives: techniques that learn from limited labeled data
- But in general: modern techniques always do better with more data
- Whoever has the data has the power!



See also:

<https://sites.google.com/view/fatecv-tutorial/schedule>