(Graduate) Introduction to computer vision

CS 6670

Quick Info

- Instructor - Bharath Hariharan
- TA – Zeqi Gu
- Office – 311 Gates Hall
- Lecture venue - 219 Phillips Hall
- Time: Tu / Thu 1 – 2:15 pm
- OH:
  - Bharath: W / F 1:15 – 2:15 pm
  - Zeqi: TBD
- Course web page: https://www.cs.cornell.edu/courses/cs6670/
- Ed Discussions: TBD

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 research means

- Research involves contributing something new to a body of knowledge
- *Uncharted waters*
- Will talk today about what to expect of research

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
  - How to do research
- Technical content will focus on evolving state-of-the-art

What you will do

- Check out Deliverables in course webpage
- Primarily project-based
  - Project title and one sentence summary (September 16th)
  - Project proposal (September 30th)
  - Extended abstract (November 1st)
  - Review (November 10th)
  - Presentation (December 5th)
  - Final paper (TBD)

- Next up: Decide a project team and get cracking!

- Instructions about project in a bit

What you will do

- Also paper readings
- Not graded, but class will be dull if you don’t read
- Aim also to learn how to read papers quickly and widely

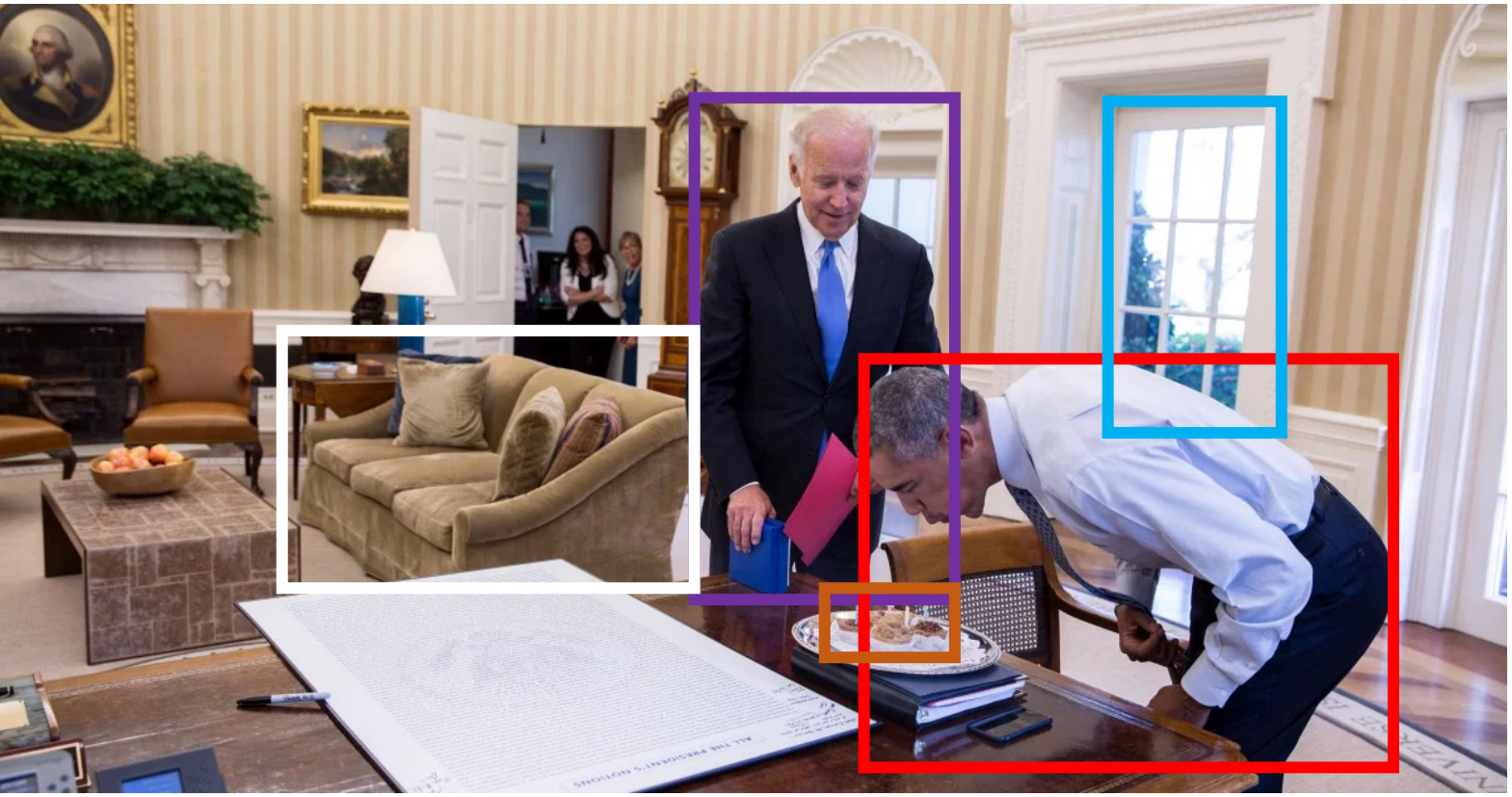
Computer vision overview

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?

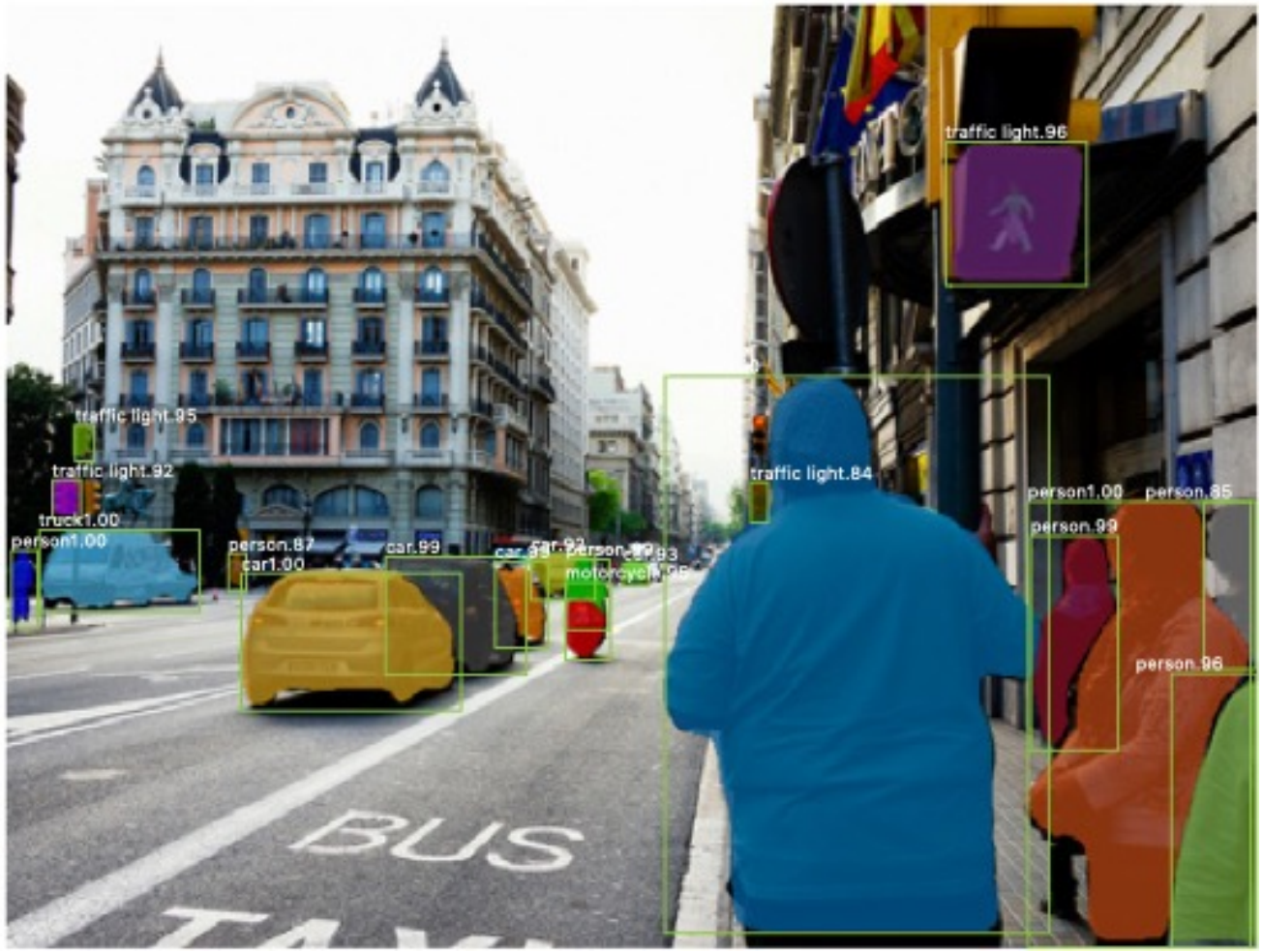
Where is the light coming from?

Where is the camera?

What is the current state-of-the-art?

- *Recognition:*
  - **Can** classify, localize some objects, understand their pose
  - *Caveats:*
    - highly dependent on training data
    - no guarantees of correctness
    - Computationally expensive in general
  - **Cannot**
    - reason about complex interactions (e.g., can the pedestrian cross the street)
    - Recognize things with insufficient training data

What is the current state of the art?

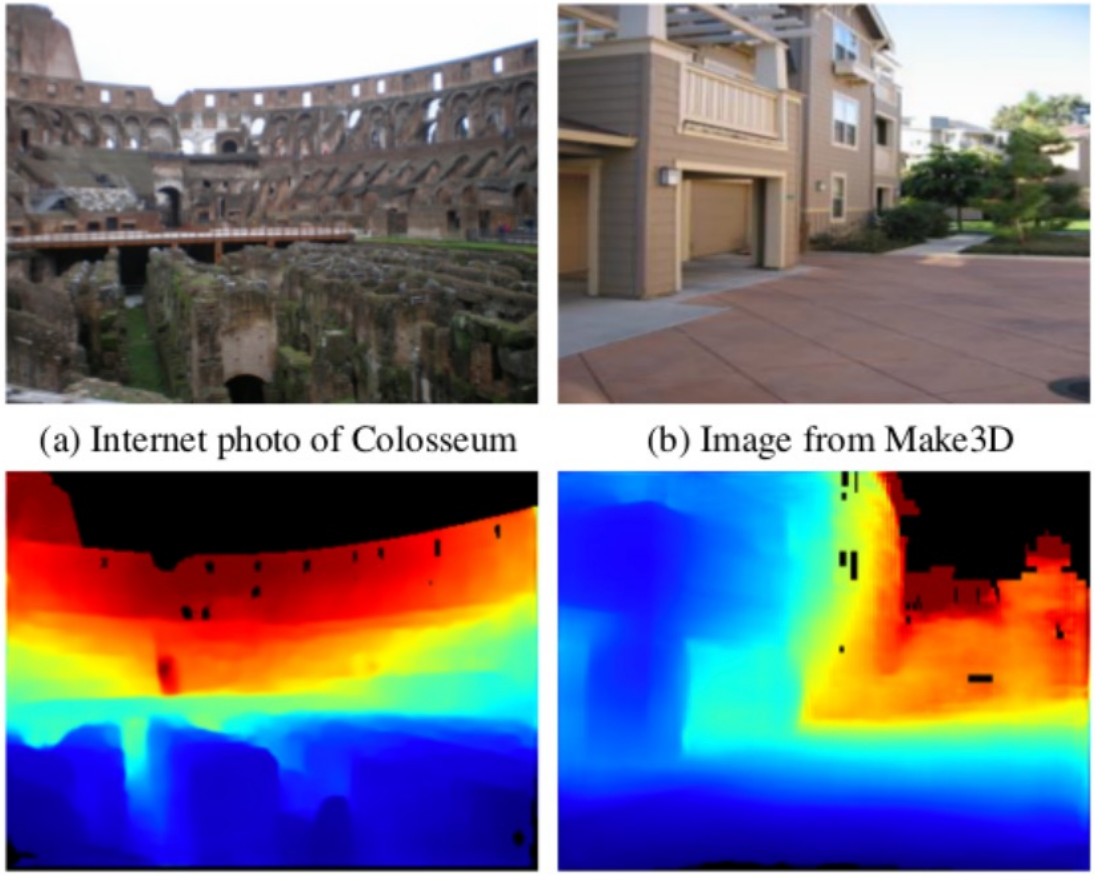


He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.

What is the current state-of-the-art?

- *Reconstruction*
  - **Can:** reconstruct rigid scenes from multiple views
  - **Caveats:**
    - Need at least two views, the more views the better
    - Difficult to reconstruct shapes of untextured, specular or transparent surfaces
    - Difficult to model materials with complex interactions with light (e.g., skin)
  - **Cannot:**
    - Reconstruct scenes from just a few views
    - Reconstruct scenes with deforming or articulating objects

What is the current state of the art



(a) Internet photo of Colosseum (b) Image from Make3D

(c) Our single-view depth prediction (d) Our single-view depth prediction

Li, Zhengqi, and Noah Snavely. "Megadepth: Learning single-view depth prediction from internet photos." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

What is the computer vision community working on now?

- Recognition
  - Training recognition systems with less data, less compute
- Reconstruction
  - Reconstructing complex scenes with fewer images
- Other tasks
  - Computer vision + Robotics (Embodied cognition)
  - Computer vision + Language

Question

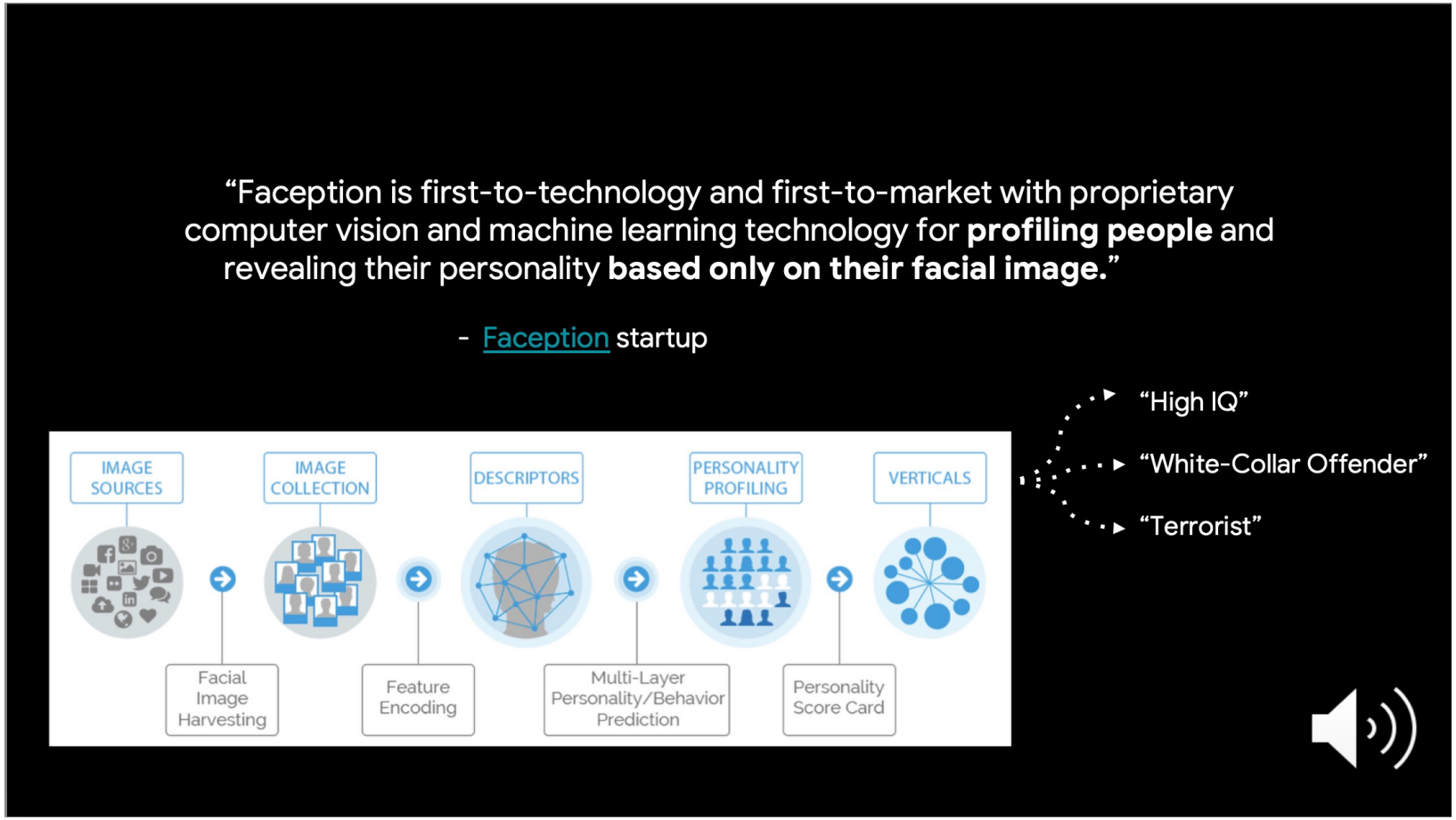
- What is a useful application of computer vision?

- What properties must existing technology satisfy in order to be deployed?

Beyond the research world: where are these deployed?

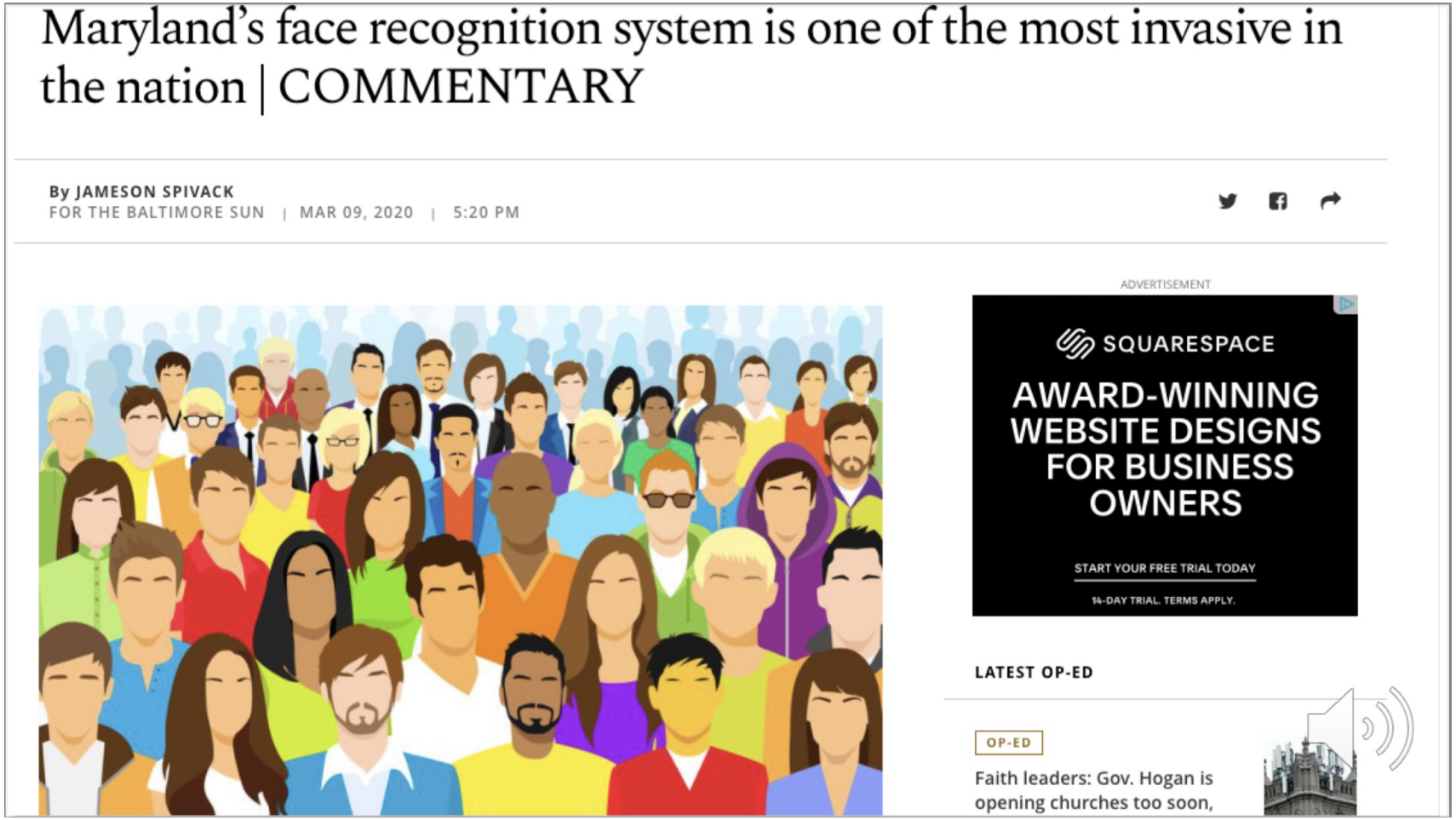
- Face recognition tech
- Self-driving cars
- Benefits and harms?

Potential for who?



Slide credit: Timnit Gebru, Emily Denton
https://sites.google.com/view/fatecv-tutorial/schedule

Potential for who?



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

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

Who is seen? *How* are they seen?

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

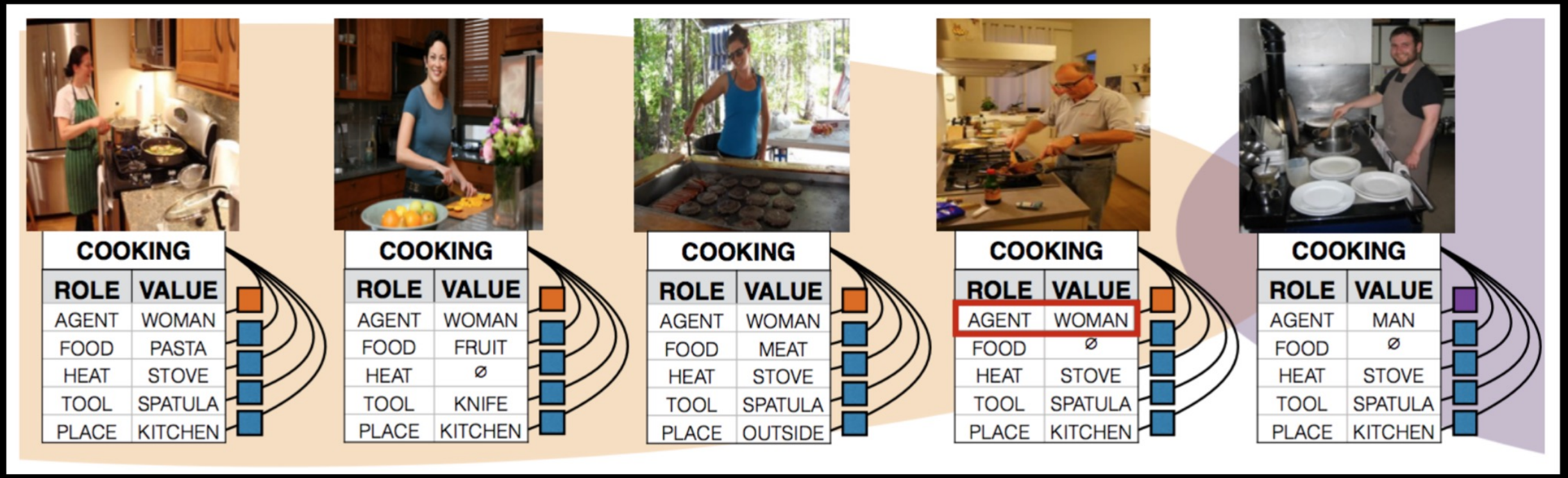
	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TYPE VI
Microsoft	1.7%	1.1%	3.3%	0%	23.2%	25.0%
FACE++	11.9%	9.7%	8.2%	13.9%	32.4%	46.5%
IBM	5.1%	7.4%	8.2%	8.3%	33.3%	46.8%

Buolamwini & Gebru FAT* 2018, Slides from Joy Buolamwini

Slide credit: Timnit Gebru, Emily Denton
https://sites.google.com/view/fatecv-tutorial/schedule

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



[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

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

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.