

# Welcome to COMP 590/776

## Computer Vision

### ... in 3D World

Instructor: Soumyadip (Roni) Sengupta

ULA: Andrea Dunn, William Li, Liujie Zheng



Course Website:  
Scan Me!

# Soumyadip (Roni) Sengupta, Assistant Professor, UNC Chapel Hill

## Research Directions in My Group

### Research Interest: 3D Vision & Graphics, Computational Photography

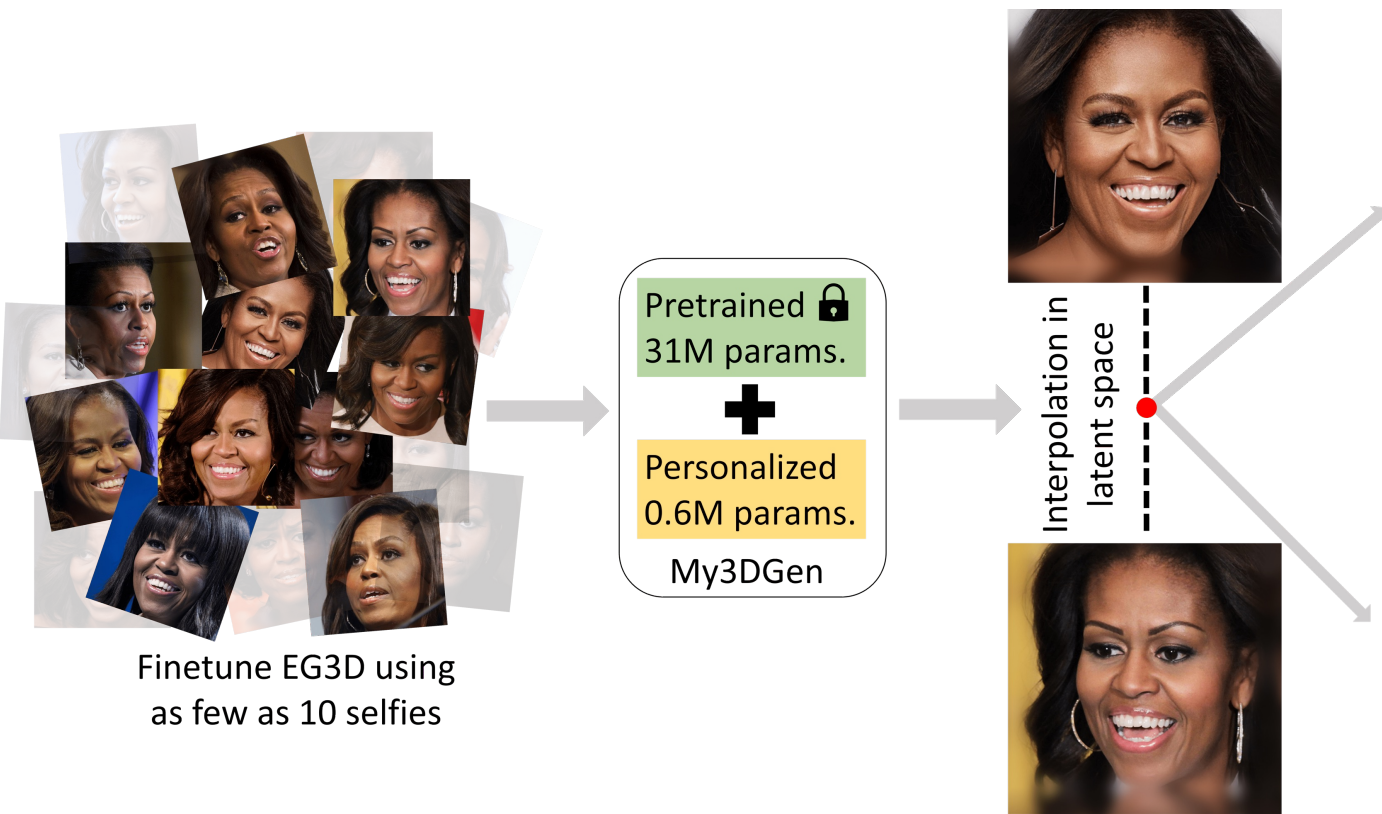
**Solving Inverse Graphics to democratize 3D capture & video production:** How do we represent geometry, reflectance, and illumination of *faces, objects, and scenes*? How do efficiently extract these representations?

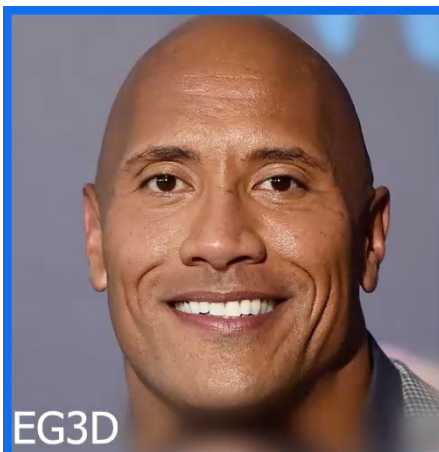
- *Building personalized 3D Generative Face models:*
  - Applications: AR/VR, Continual Learning, Parameter Efficiency
- *Lighting-based 3D reconstruction:*
  - Applications: fast & light weight reconstruction for AR/VR, 3D reconstruction from endoscopy videos
- *Neural Relighting:*
  - Applications: Relightable Zoom calls, Security & Privacy in video calls.
- *Understanding Facial motion:*
  - Applications: Deep Fakes impact to Public Health, Improving camera-based health sensing.

# My3DGen: Building Lightweight Personalized 3D Generative Model

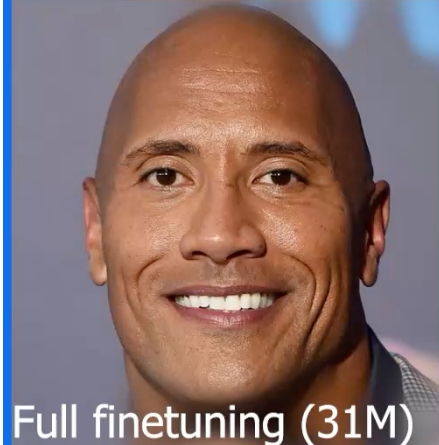
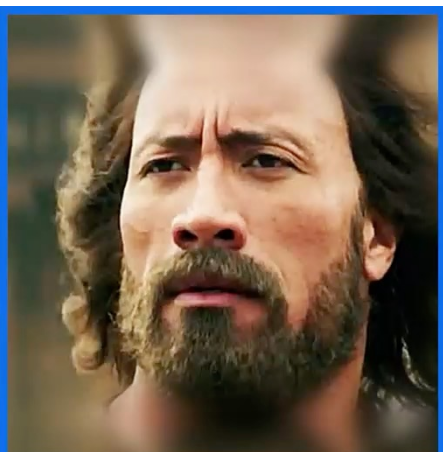
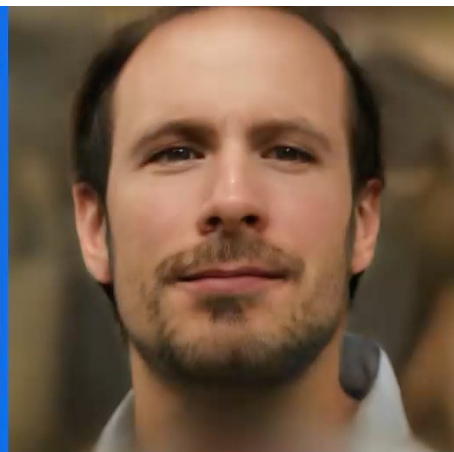
Luchao Qi, Jiaye Wu, Shengze Wang, Soumyadip Sengupta

Arxiv 2023

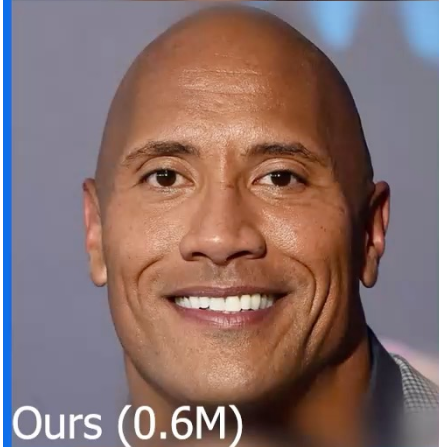
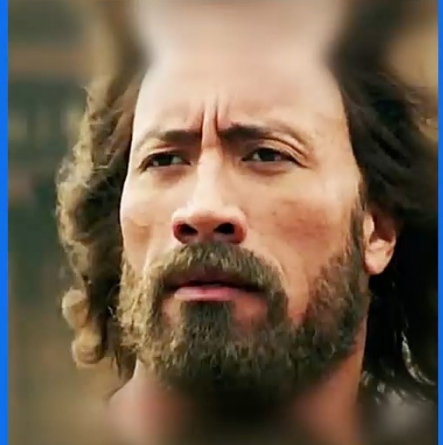
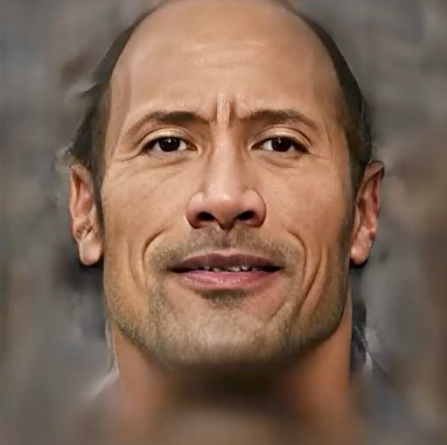
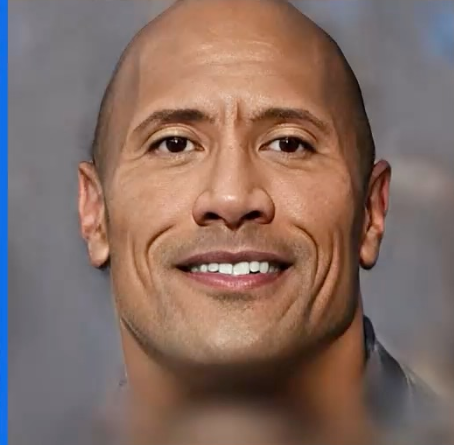




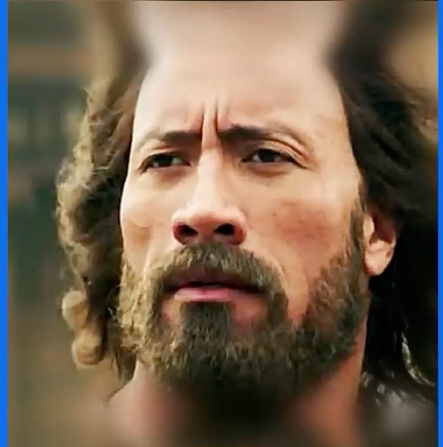
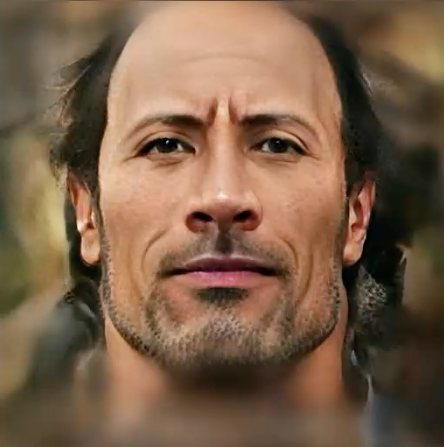
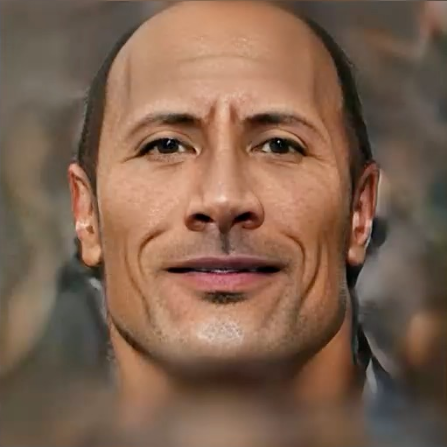
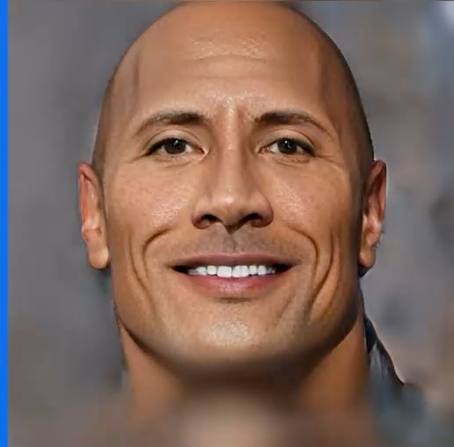
EG3D



Full finetuning (31M)



Ours (0.6M)



# Photometric Stereo + Multi-view Stereo for fast 3D reconstruction

Input



Ground-Truth



PS-NeRF



**Per-scene Optimization**  
Time req. **~12 hours**

Ours



**Multi-view Photometric Stereo**  
Time req. **105 secs**

CasMVSNet



**Multi-view Stereo**  
Time req. **22 secs**

No per-scene optimization

“MVPSNet: Fast Generalizable Multi-view Photometric Stereo”, Dongxu Zhao, Daniel Lichy, Pierre-Nicolas Perrin, Jan-Michael Frahm, Soumyadip Sengupta, ICCV 2023.

# Photometric Stereo + SLAM for colon reconstruction in colonoscopy



Input Video



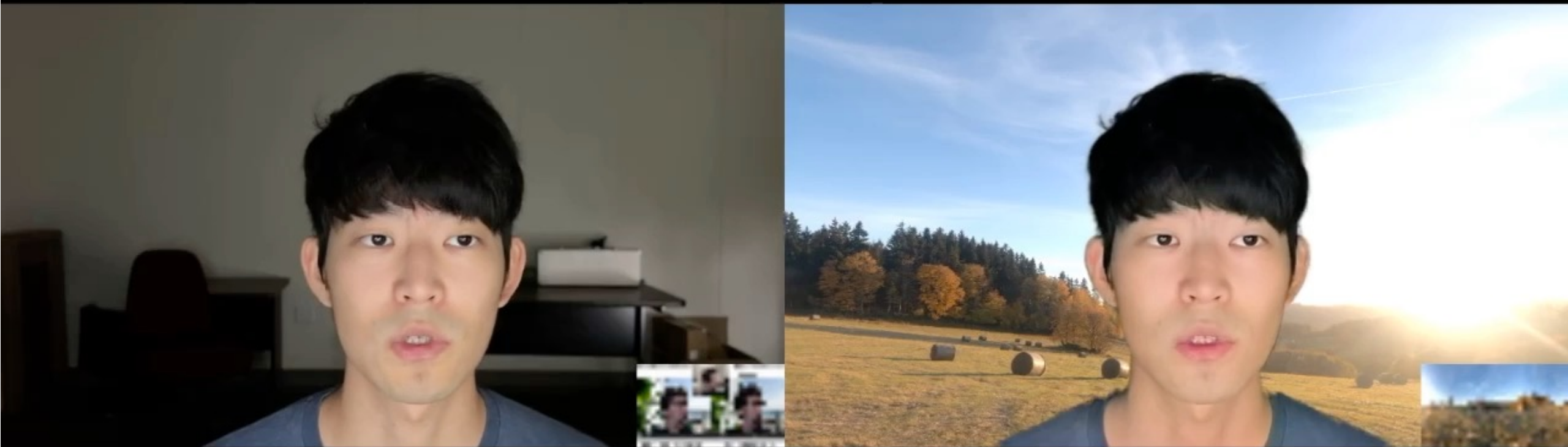
SLAM only



Photometric Stereo +  
SLAM (Ours)

“A Surface-normal Based Neural Framework for Colonoscopy Reconstruction”, Sherry Wang, Yubo Zhang, Sarah McGill, Julian Rosenman, Jan-Michael Frahm, Soumyadip Sengupta, Steve Pizer, IPMI 2023.

# “Relightable Video Calls”, Jun-Myeong Choi, Max Christman, Soumyadip Sengupta



# Today's Plan

- Course Overview
- Why Computer Vision?
- Computer Vision in real world
- Ethics in Computer Vision



# Today's Plan

- **Course Overview**
- Why Computer Vision?
- Computer Vision in real world
- Ethics in Computer Vision

# Schedule

Date	Topic	Details	Special Dates
Intro & Review			
Tues Aug 22	Intro. to Computer Vision & Ethical Concerns		
Thrs Aug 24	Maths Review (Linear Algebra, Probability, Calculus)		HW1: Maths review (assigned)
Colors & Imaging			
Tues Aug 29	Color & Color Spaces		
Thrs Aug 31	In-Camera Imaging Pipeline		HW1: Maths review (due)
Tues Sept 5	No Class (Well-being day)		
Image Processing			
Thrs Sept 7	Filtering - Convolution, Gradients, & Edges		
Tues Sept 12	Frequency domain - Fourier Analysis		HW2: Image Processing (assigned)
Features			
Thrs Sept 14	Feature Detection (Corner & Blob)		
Tues Sept 19	Feature Descriptor & Matching (SIFT)		
2D Transformation			
Thrs Sept 21	2D Transformations & Fitting		HW2: Image Processing (due) HW3: Panorama (assigned)
Tues Sept 26	RANSAC + Image Blending		

Learning & Perception			
Thrs Sept 28	Brief Intro to Deep Learning		
Tues Oct 3	Recognition & Detection		HW3: Panorama (due)
Thrs Oct 5	Segmentation & Matting		HW4: Deep Learning 1 (assigned)
Tues Oct 10	Optical Flow		
Thrs Oct 12	No Class (University Day)		
Tues Oct 17	Generative Models		Online Lecture HW4: Deep Learning 1 (due)
3D Vision			
Thrs Oct 19	No Class (Fall Break)		
Tues Oct 24	Camera Models + Calibration - 1		
Thrs Oct 26	Camera Models + Calibration - 2		HW5: 3D Vision 1 (assigned)
Tues Oct 31	Two-view Geometry-1		
Thrs Nov 2	Two-view Geometry-2		
Tues Nov 7	Stereo		HW5: 3D Vision 1 (due)
Thrs Nov 9	Multi-view Stereo		
Tues Nov 14	Structure from Motion		HW6: 3D Vision 2 (assigned)
Thrs Nov 16	Light & Photometric Stereo		
Tues Nov 21	Deep Learning for MVS, SfM, PS		
Thrs Nov 23	No Class (Thanksgiving)		
Tues Nov 28	NeRFs		HW7: 3D Vision 2 (due)
Thrs Nov 30	Mid-term Review		
Tues Dec 5	Final Exam (in-class and/or take home)		Syllabus: Whole course

# Course Policies

## Grading (for both 590 & 790):

- Assignment 1: [pen & paper] Linear Algebra and Probability Recap: 5%
- Assignment 2: [pen & paper + coding in python] Frequency domain image analysis: 10%
- Assignment 3: [coding in python] Stitching images to generate panorama: 10%
- Assignment 4: [coding in Google Colab] Deep Learning 1: 15%
- Assignment 5: [pen & paper] 3D Vision 1: 10%
- Assignment 6: [pen & + coding in python] 3D Vision 2: 10%
- Final-exam: [pen & paper] in-class + take-home: 30%
- Class Participation & Quizzes: 10%

Grades for 590 students will be curved differently from 790 students.

## Misc.

Final exam: Final exam will be a mixture of in-person & take-home. Students are expected to finish the final exam in-person in 75 minutes. However you can choose to finish some questions by taking it home and submitting it online by the end of the day. You will be graded for only 80% of the total points assigned for each question you submit as take-home.

Class Participation: Class participation points will be provided based on random quizzes throughout the class.

Late Submissions: Late assignments will not be accepted. You lose 5% of total points for that assignment for each late day. No exception allowed.

### Academic Integrity & Collaboration:

- For your assignments, you are allowed to use materials from external sources as long as it helps you to understand the topic and NOT actually solve the assignment problem.
- You MUST NOT use homework solutions from previous years.
- You MUST clearly acknowledge any sources used for solving the assignment.
- You can discuss and brainstorm in groups but the programming and the solution has to be done on an individual basis.
- No copying or replicating of other existing solutions.
- Final-exam is in class. You can bring 1 US Letter size cheat-sheet (1-sided). But collaboration, electronic devices, browsing the web, printed materials etc. are not permitted.
- ANY VIOLATION OF ACADEMIC INTEGRITY WILL BE REPORTED TO THE UNC OFFICE OF STUDENT CONDUCT

590 Grades:

A : > 79% (18/35) -> (4 people: >90%, 10 people: 85%-90%, 4 people: 79%-85%)  
A- : 79%-72% (6/35)  
B+ : 68%-72% (3/35)  
B : 64%-68% (2/35)  
B- : 60%-64% (0/35)  
C+ : 55%-60% (2/35)  
C : 50%-55% (0/35)  
C- : 44%-50% (2/35)  
D+ : 40%-44% (0/35)  
D : 35%-40% (0/35)  
F : < 35% (2/35)

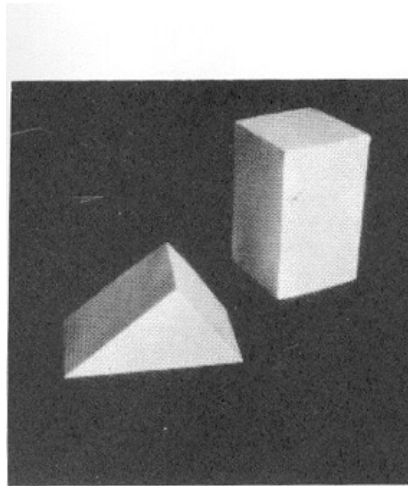
790 grades:

H+ : >90% (3/18)  
H : 79%-90% (10/18)  
H- : 72%-79% (3/18)  
P+ : 67%-72% (1/18)  
P : 60%-67% (1/18)

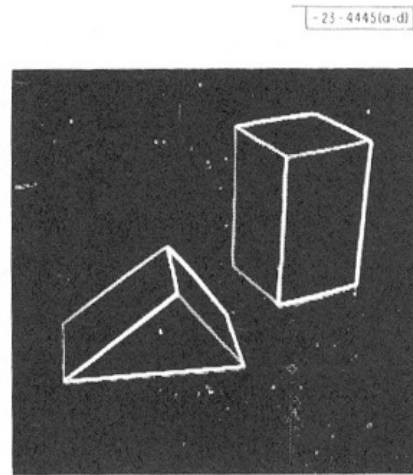
# Today's Plan

- Course Overview
- **Why Computer Vision?**
- Computer Vision in real world
- Ethics in Computer Vision

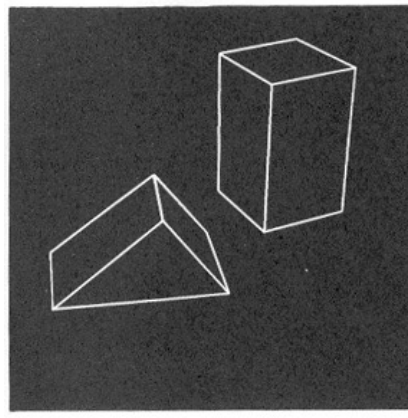
# Origins of computer vision



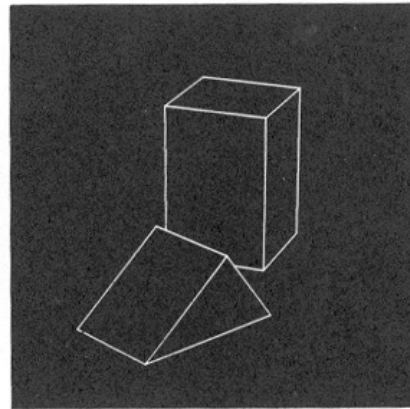
(a) Original picture.



(b) Differentiated picture.



(c) Line drawing.

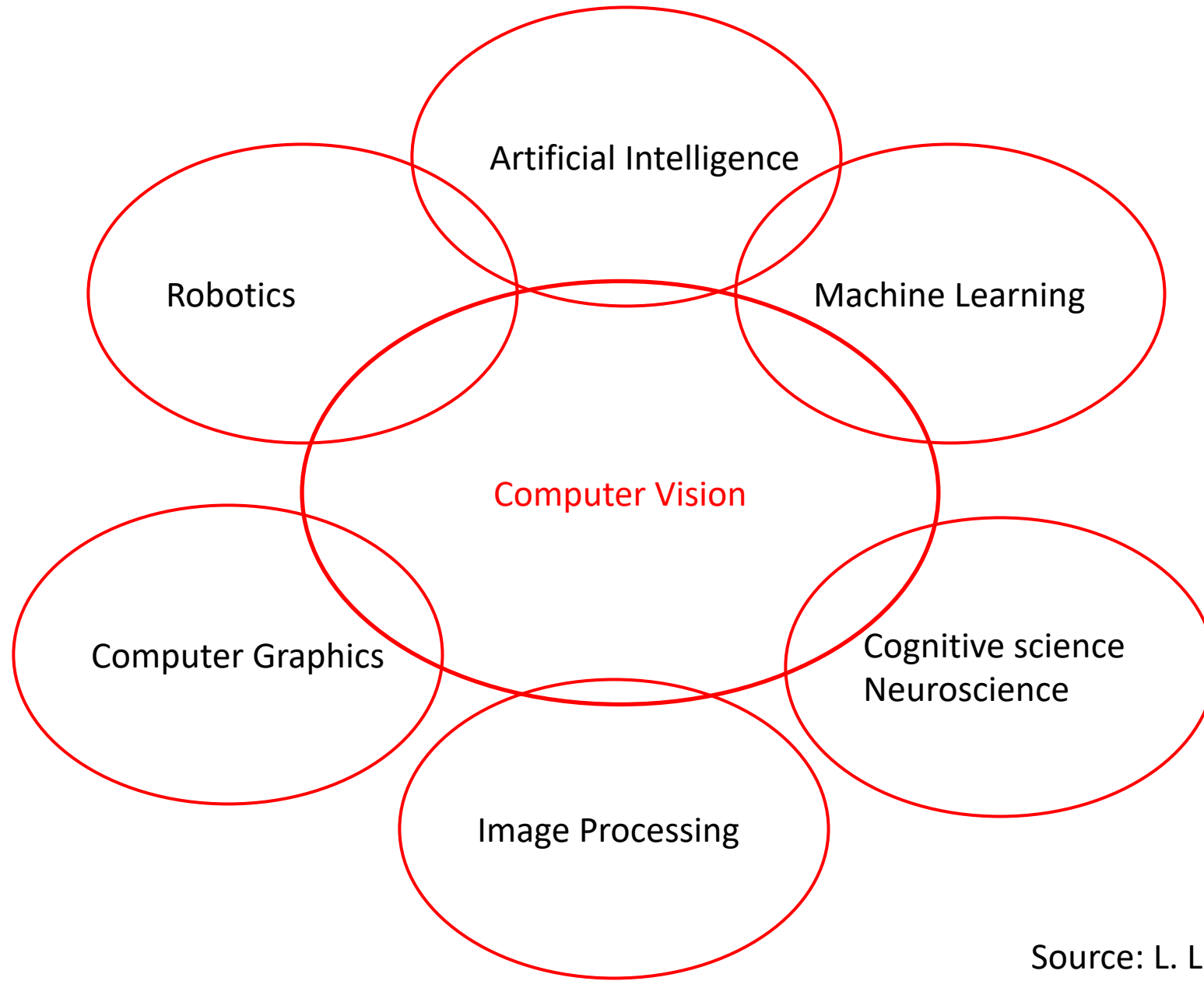


(d) Rotated view.

L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.



# Connections to other disciplines





0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

# Every image tells a story



- Goal of computer vision: perceive the “story” behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

# Can computers match human perception?



- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are better at “hard” things
- But huge progress
  - Accelerating in the last five years due to deep learning
  - What is considered “hard” keeps changing

# Human perception has its shortcomings



[Sinha and Poggio, \*Nature\*, 1996](#)  
("The Presidential Illusion")

But humans can tell a lot about a scene from a little information...



# The goal of computer vision

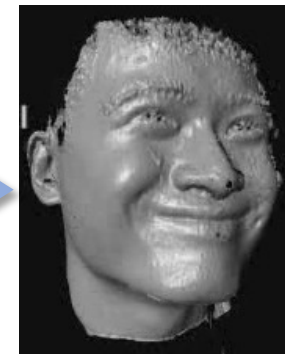
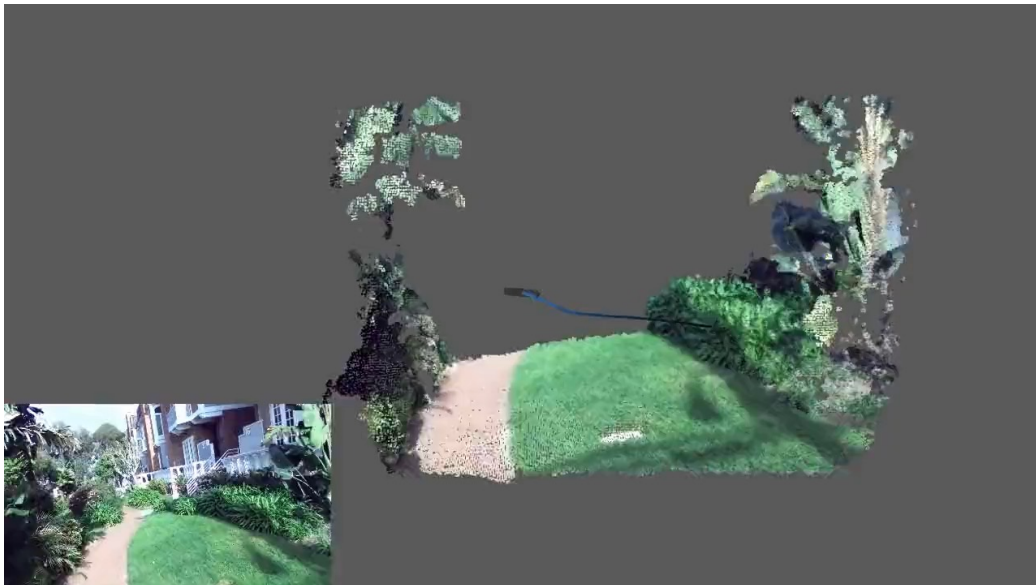
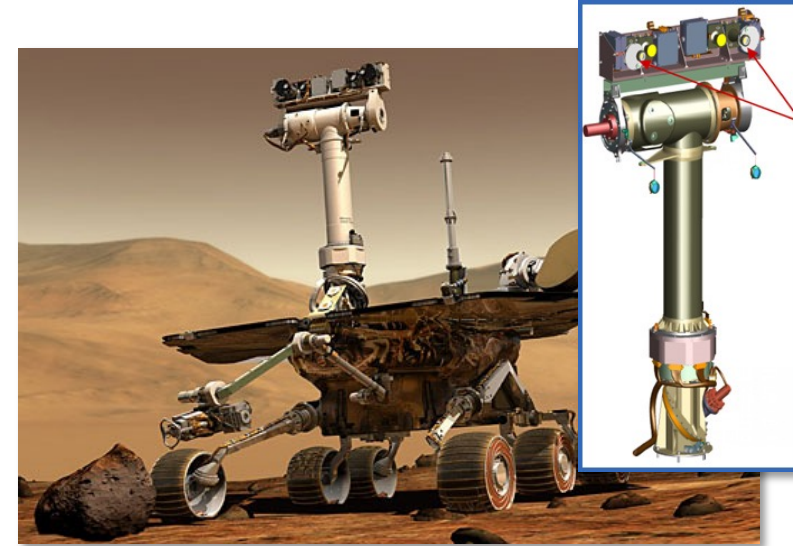


# The goal of computer vision

- Compute the 3D shape of the world (3D Vision)



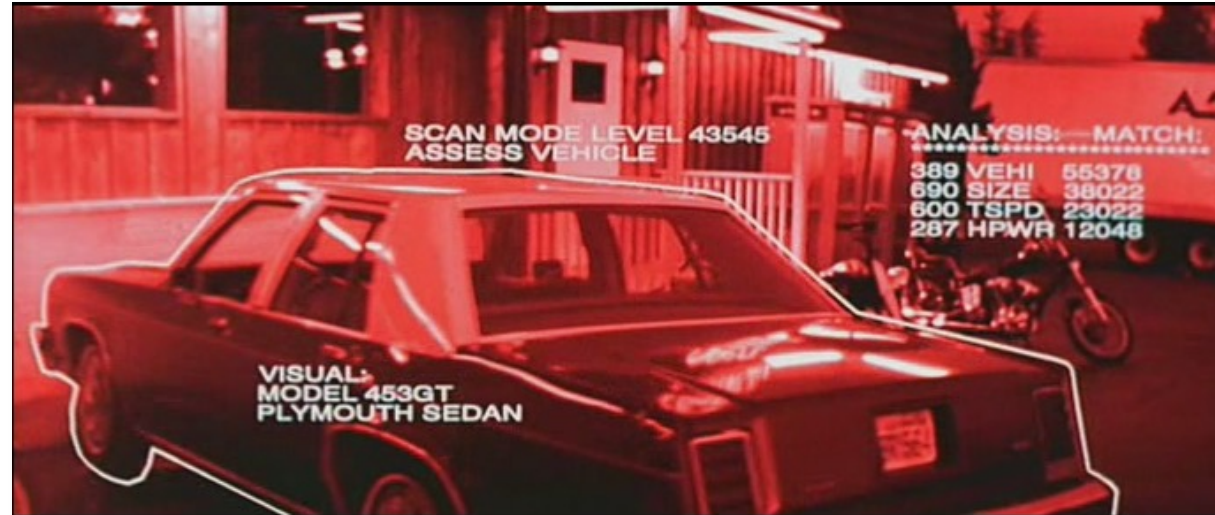
ZED 2i Camera





# The goal of computer vision

- Recognize objects and people (Learning and Perception)



*Terminator 2, 1991*



sky

building

flag

face

banner

wall

street lamp

bus

bus

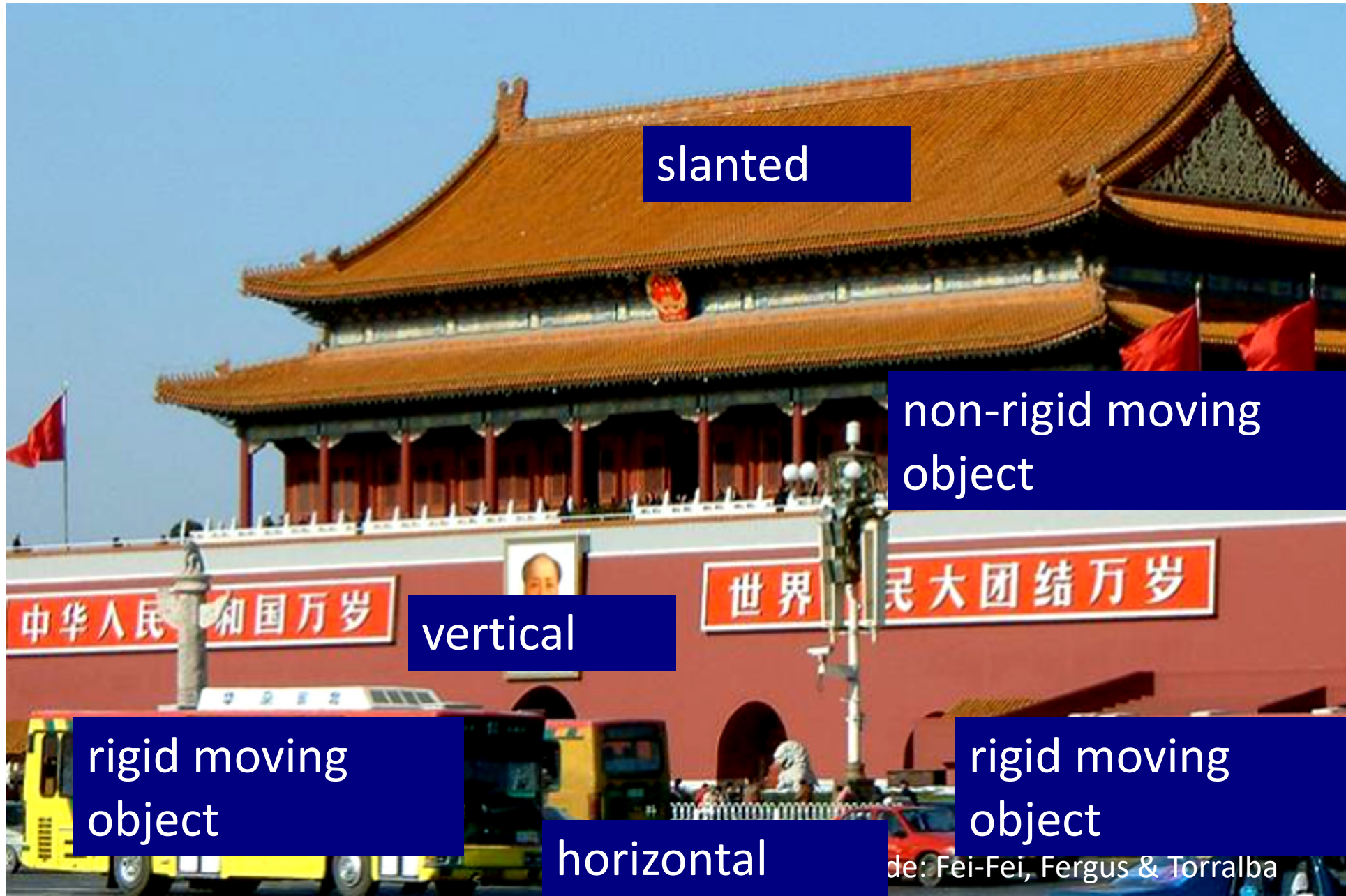
cars

# Scene and context categorization

- outdoor
- city
- traffic
- ...



# Qualitative spatial information



# The goal of computer vision

- Improve photos (“Computational Photography”)



Super-resolution (source: 2d3)



Low-light photography  
(credit: [Hasinoff et al., SIGGRAPH ASIA 2016](#))

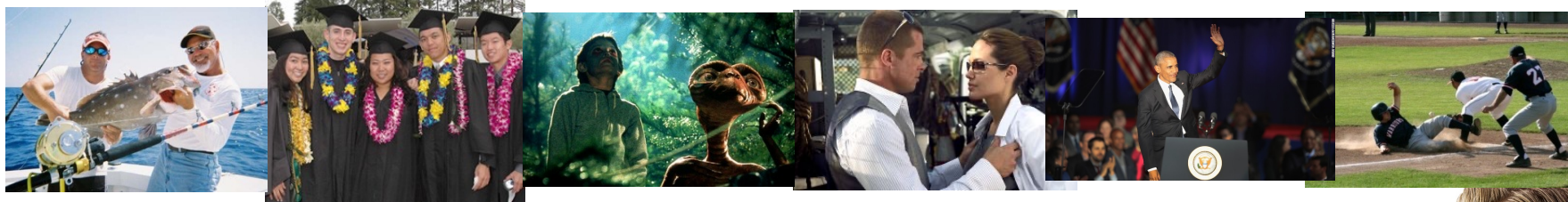


Depth of field on cell phone camera  
(source: [Google Research Blog](#))



Inpainting / image completion  
(image credit: Hays and Efros)

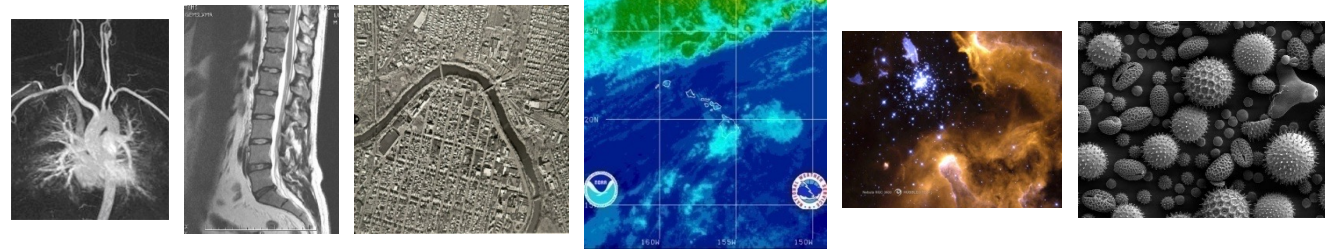
- Billions of images/videos captured per day



flickr



YouTube



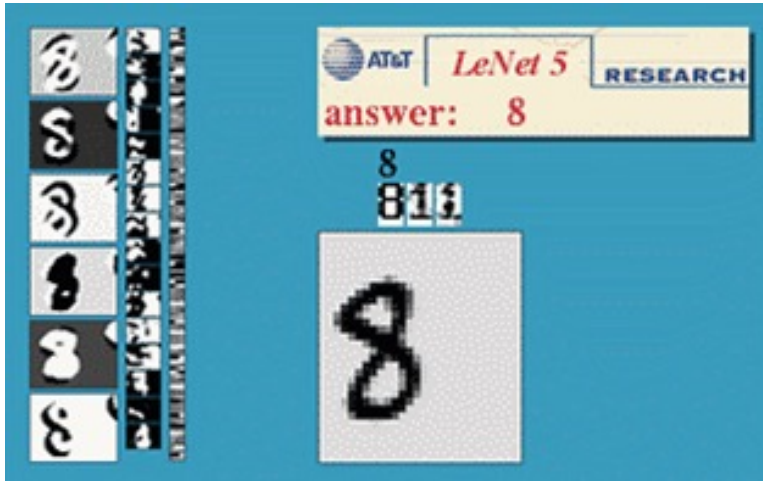
- Huge number of potential applications

# Today's Plan

- Course Overview
- Why Computer Vision?
- **Computer Vision in real world**
- Ethics in Computer Vision

# Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs (1990's)  
<http://yann.lecun.com/exdb/lenet/>



License plate readers

[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



Automatic check processing



Sudoku grabber

<http://sudokugrab.blogspot.com/>



# Face detection



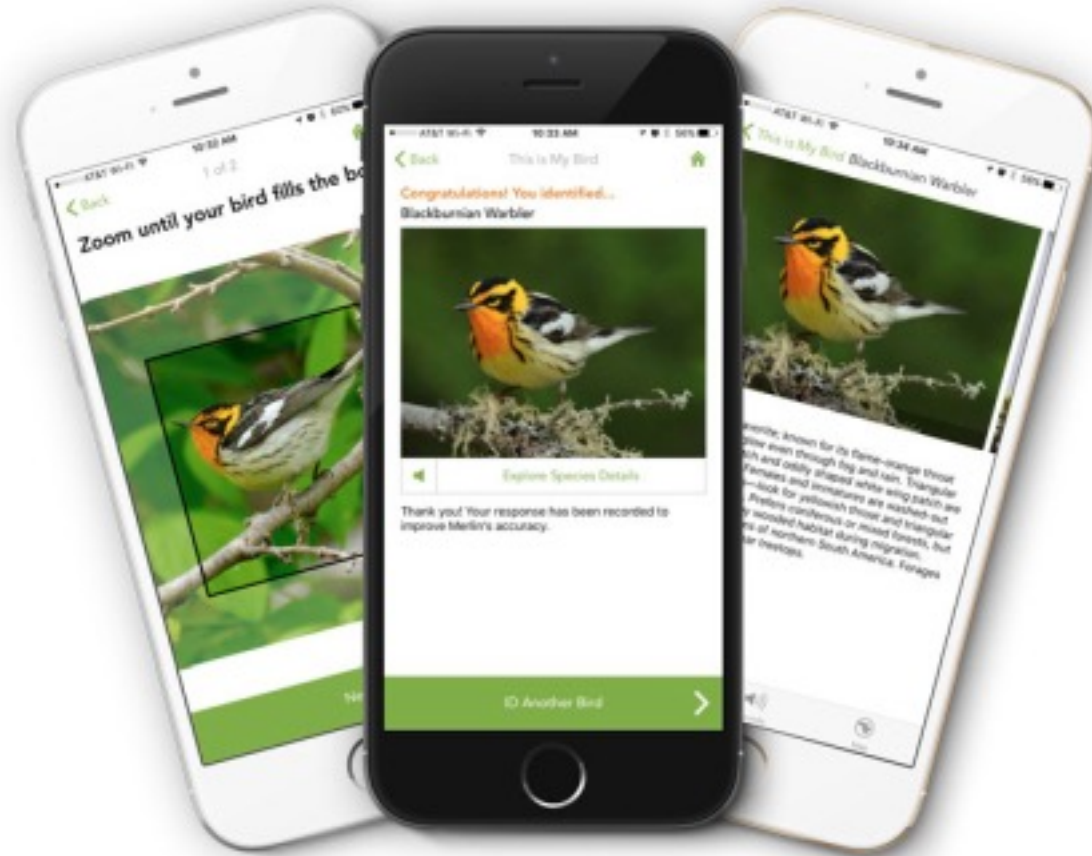
- Nearly all cameras detect faces in real time

# Object Recognition

- Item recognition by Evolution Robotics

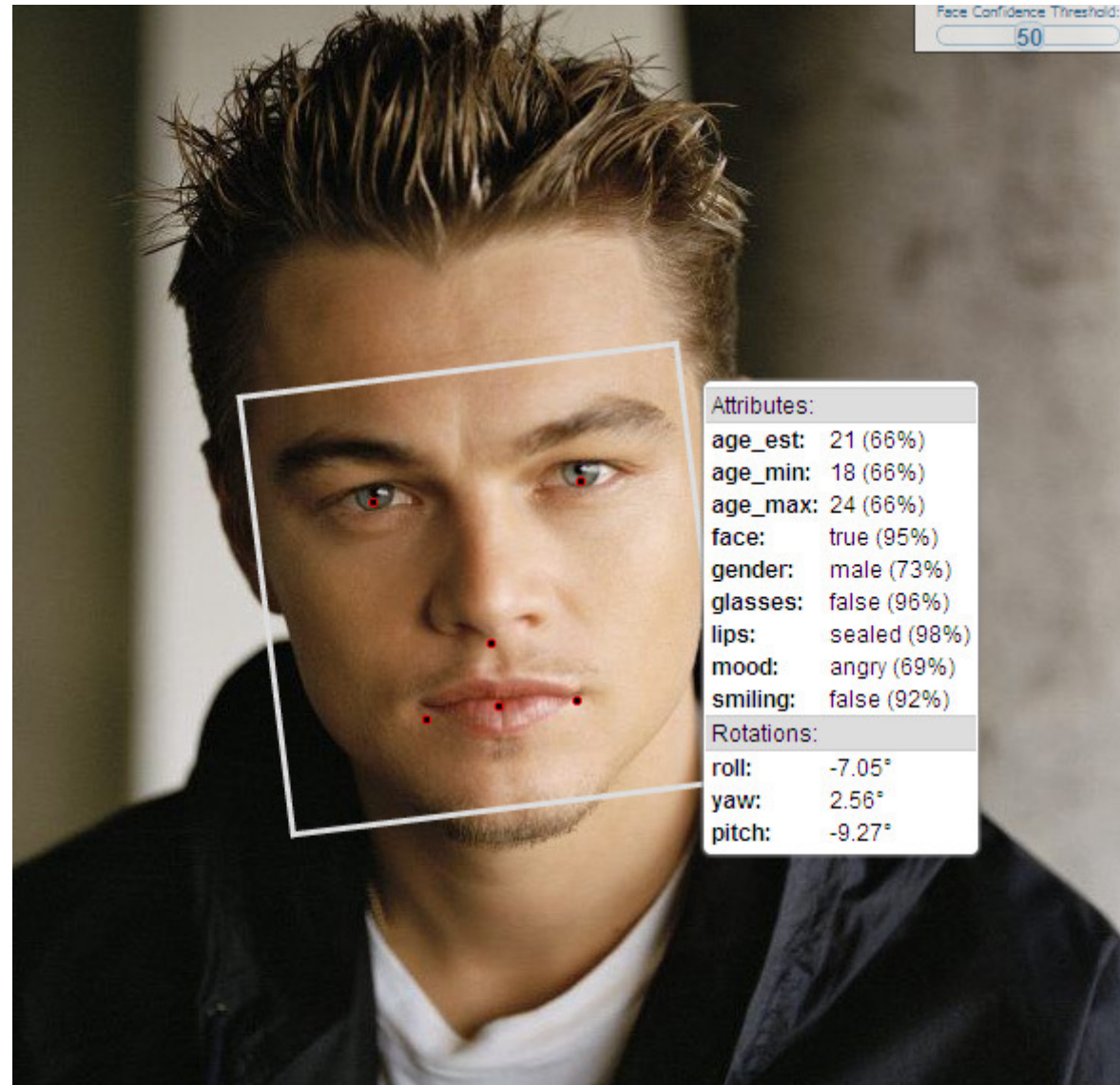


# Bird identification

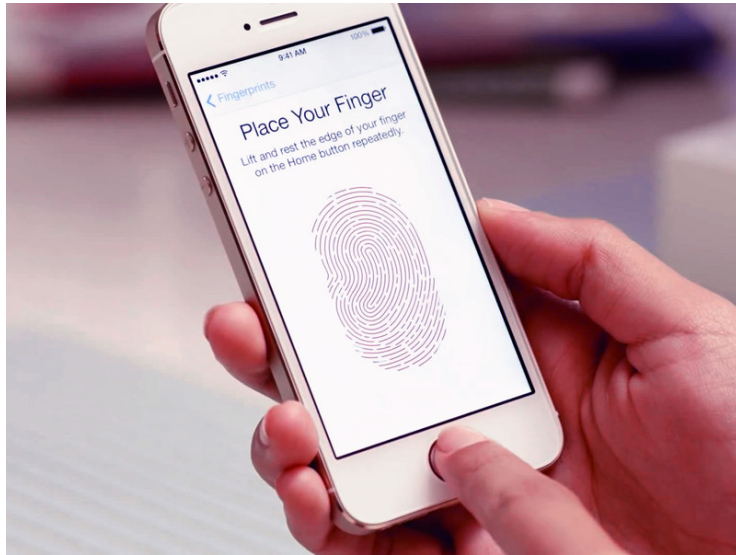


Merlin Bird ID (based on Cornell Tech technology!)

# Face analysis and recognition



# Login without a password



Fingerprint scanners on many new smartphones and other devices



Face unlock on Apple iPhone X  
See also <http://www.sensiblevision.com/>

# Special effects: shape capture



*The Matrix* movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

# Special effects: motion capture



*Pirates of the Caribbean*, Industrial Light and Magic

Source: S. Seitz

MOVIES



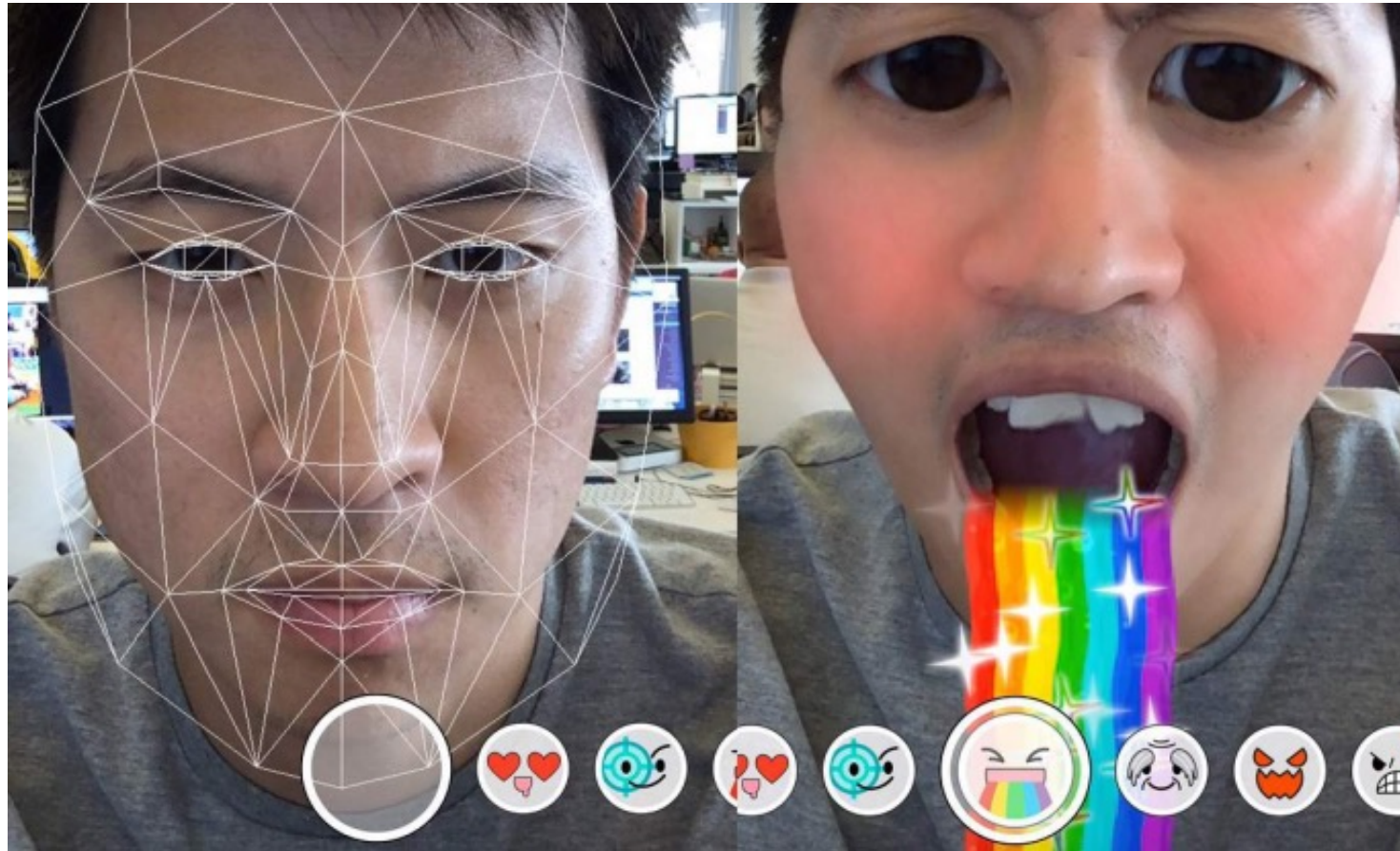
## Robert De Niro said no green screen. No face dots. How ‘The Irishman’s’ de-aging changes Hollywood



Makeup and wig work got Robert De Niro partway to his character, Frank Sheeran, at 41, left. It took a specially built camera and visual artists to get all the way there, as before-and-after images show. (Netflix)



# 3D face tracking w/ consumer cameras



Snapchat Lenses

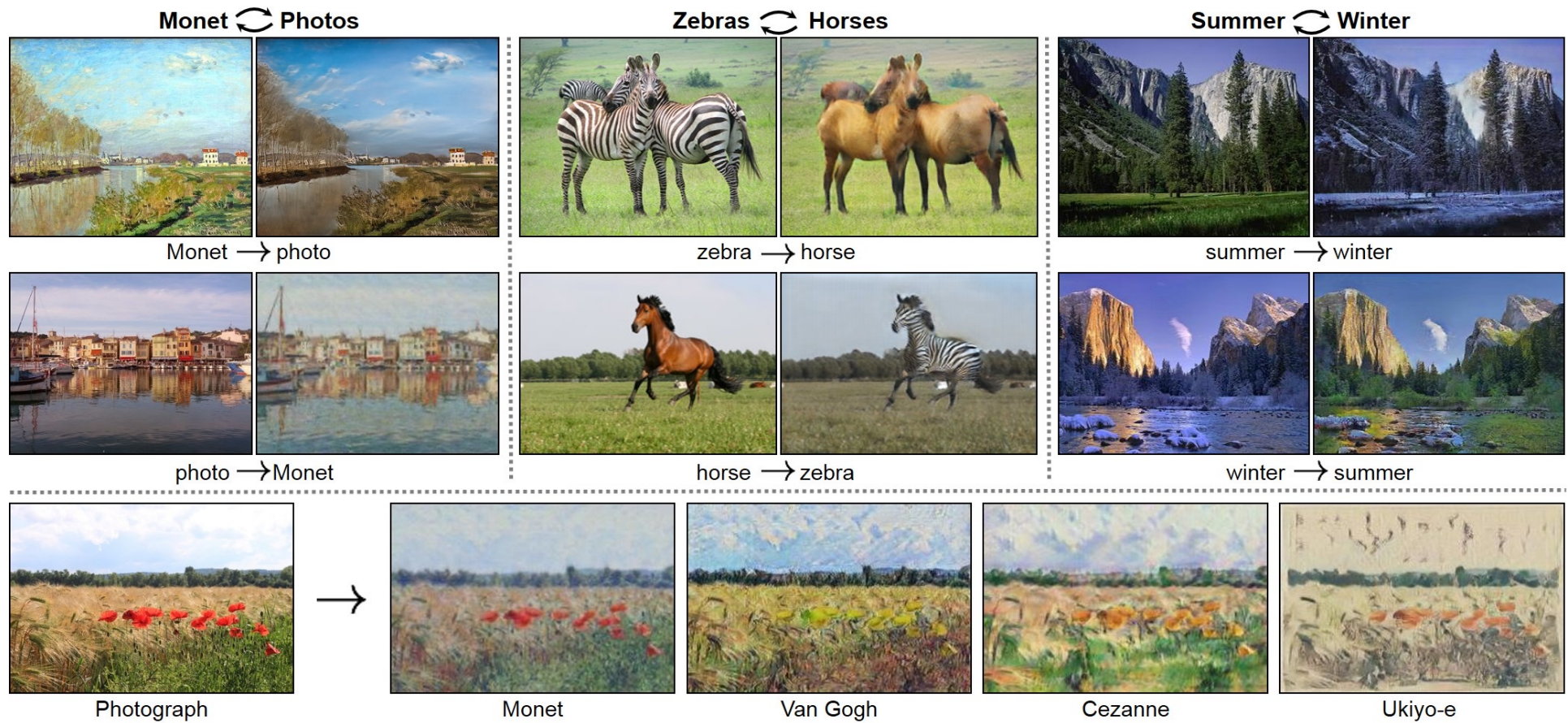
# 3D face tracking w/ consumer cameras



# Image synthesis



# Image synthesis



# Sports

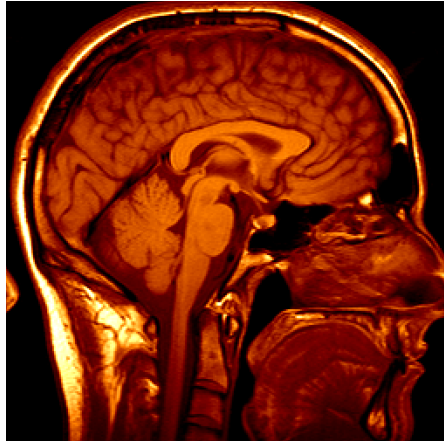


*Sportvision* first down line  
[Explanation](http://www.howstuffworks.com) on [www.howstuffworks.com](http://www.howstuffworks.com)

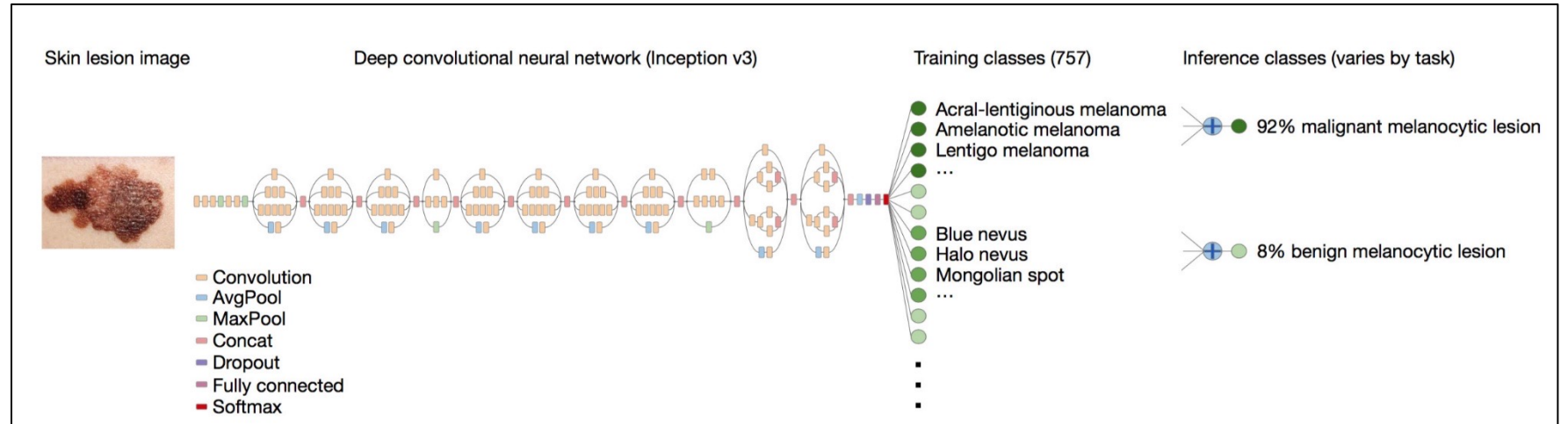


Highlights of the men's 4x200m relay final on Day 5.

# Medical imaging



3D imaging  
(MRI, CT)



Skin cancer classification with deep learning

<https://cs.stanford.edu/people/esteva/nature/>

# Smart cars

The banner features a central image of a car from a top-down perspective, with four yellow beams representing camera views: rear, forward, and two side views. The text 'Our Vision. Your Safety.' is prominently displayed above the car. Navigation tabs for 'manufacturer products' and 'consumer products' are at the top. Below the car, three product highlights are shown: 'EyeQ Vision on a Chip' with a chip image, 'Vision Applications' with a pedestrian detection image, and 'AWS Advance Warning System' with a dashboard display image. Each highlight includes a 'read more' link.

manufacturer products    consumer products

## Our Vision. Your Safety.

rear looking camera    forward looking camera

side looking camera

↳ **EyeQ** Vision on a Chip [read more](#)

↳ **Vision Applications** Road, Vehicle, Pedestrian Protection and more [read more](#)

↳ **AWS** Advance Warning System [read more](#)

**News**

↳ Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System

↳ Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end ... [read more](#)

[all news](#)

**Events**

↳ [Mobileye at Equip Auto, Paris, France](#)

↳ [Mobileye at SEMA, Las Vegas, NV](#)

[read more](#)

- [Mobileye](#)
- Tesla Autopilot
- Safety features in many cars

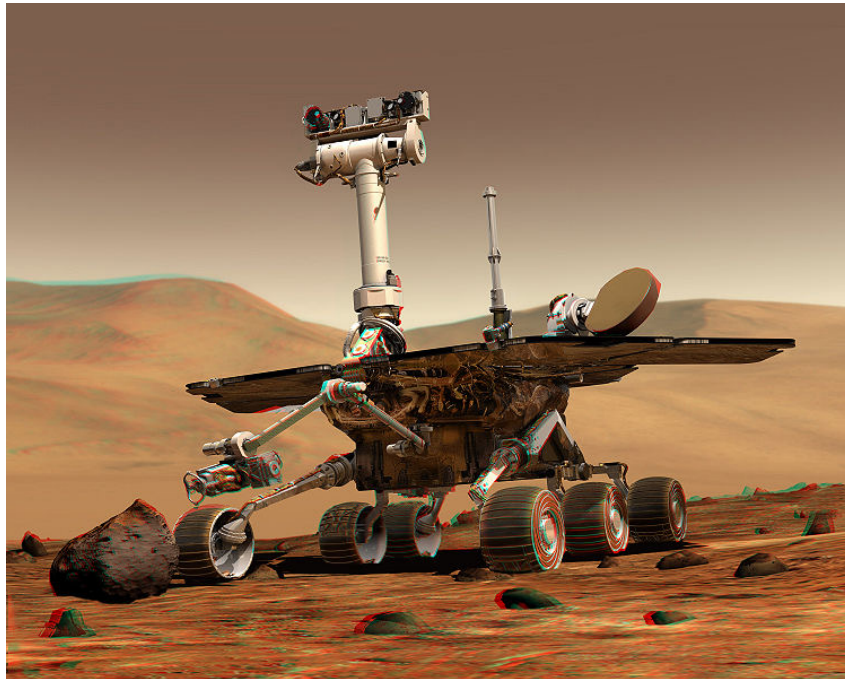
# Night Vision Systems

- Daimler Night Vision system with pedestrian detection

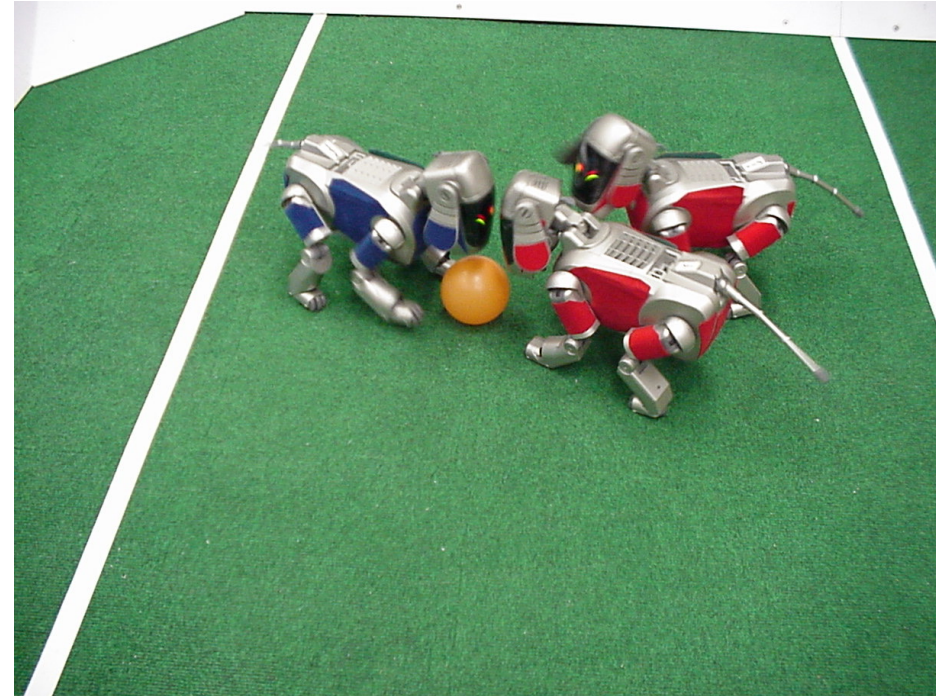




# Robotics



NASA's Mars Spirit Rover  
[http://en.wikipedia.org/wiki/Spirit\\_rover](http://en.wikipedia.org/wiki/Spirit_rover)



<http://www.robocup.org/>

# Vision in space



[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

## Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “[Computer Vision on Mars](#)” by Matthies et al.

# Robotics



Amazon Prime Air



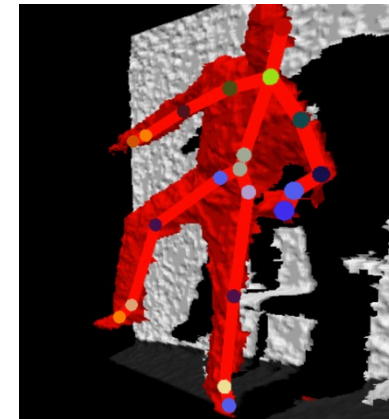
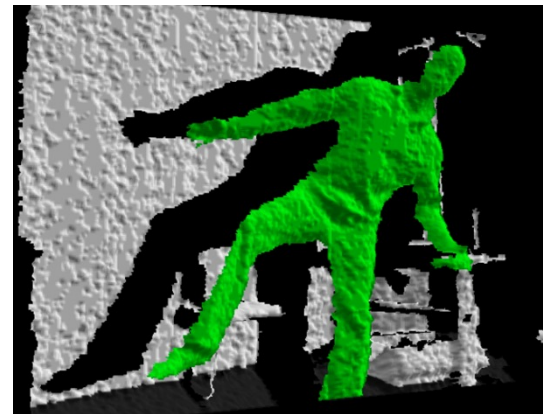
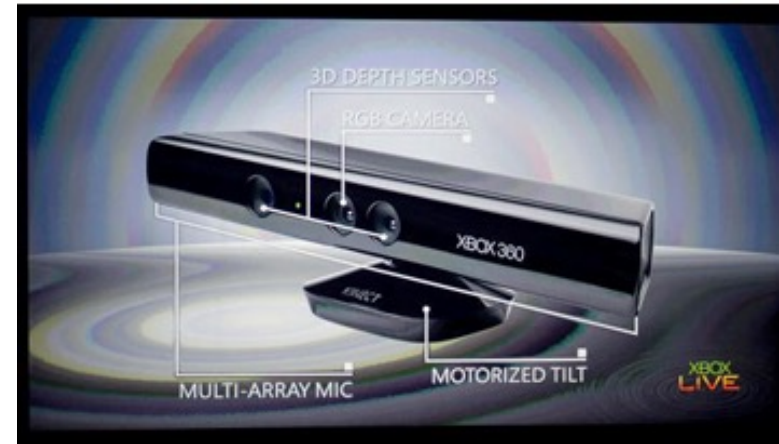
Amazon Picking Challenge

<http://www.robocup2016.org/en/events/amazon-picking-challenge/>



Amazon Scout

# Vision-based interaction: Xbox Kinect



<http://blogs.howstuffworks.com/2010/11/05/how-microsoft-kinect-works-an-amazing-use-of-infrared-light/>

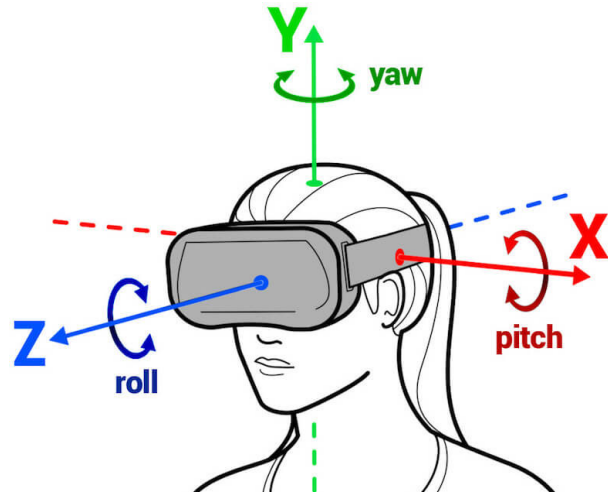
<http://www.xbox.com/en-US/Live/EngineeringBlog/122910-HowYouBecometheController>

<http://electronics.howstuffworks.com/microsoft-kinect.htm>

<http://www.ismashphone.com/2010/12/kinect-hacks-more-interesting-than-the-devices-original-intention.html>

Source: L. Lazebnik

# Virtual & Augmented Reality



6DoF head tracking



Hand & body tracking



3D scene understanding



3D-360 video capture

# Current state of the art

- You just saw many examples of current systems.
  - Many of these are less than 5 years old
- Computer vision is an active research area, and rapidly changing
  - Many new apps in the next 5 years
  - Deep learning powering many modern applications
- Many startups across a dizzying array of areas
  - Deep learning, robotics, autonomous vehicles, medical imaging, construction, inspection, VR/AR, ...

# Today's Plan

- Course Overview
- Why Computer Vision?
- Computer Vision in real world
- **Ethics in Computer Vision**
  - Examples of bias in computer vision and beyond
  - Datasets and unintended consequences
  - Algorithms (Discriminative & Generative)

0.0B

Public

🔗 16



# Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

**Humans only 60% correct,  
but an algorithm is 80% correct!**

Contributors: [Yilun Wang](#), [Michal Kosinski](#)

Date created: 2017-02-15 11:37 AM | Last Updated: 2020-05-25 06:11 PM

Identifier: DOI [10.17605/OSF.IO/ZN79K](https://doi.org/10.17605/OSF.IO/ZN79K)

Category:  Project

Description: We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.



# Why do we build ML systems?

Automate decision making, so machines can make decision instead of people.

**Ideal:** Automated decisions can be cheaper, more accurate, more impartial, improve our lives

**Reality:** If we aren't careful, automated decisions can encode bias, harm people, make lives worse

# Advances in computer vision

- Sometimes we think of technological development as a uniform positive
- But computer vision exists in a societal context, and can have both good and bad consequences – need to be mindful of both
- Example: as computer vision gets better, our privacy gets worse (e.g., through improved face recognition)

# Questions

- Should I be working on this problem at all?
- Does a given vision task even make sense?
- What are the implications if it doesn't work well?
- What are the implications if it does work well?
- What are the implications if it works well for some people, but not others?
- Who benefits and who is harmed?
- (About datasets) How was it collected? Is it representative?
- (For any technology) Who is it designed for?

# More questions

- Does the application align with your values?
- Does the task specification / evaluation metric reflect the things you care about?
- For recognition:
  - Does the collected training / test set match your true distribution?
- Are the algorithm's errors biased?
- Are you being honest in public descriptions of your results?
- Is the accuracy/correctness sufficient for public release?

# Case study – classifying sexual orientation

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

0.0B Public 16 ...

Contributors: [Yilun Wang](#), [Michal Kosinski](#)

Date created: 2017-02-15 11:37 AM | Last Updated: 2020-05-25 06:11 PM

Identifier: DOI [10.17605/OSF.IO/ZN79K](https://doi.org/10.17605/OSF.IO/ZN79K)

Category:  Project

Description: We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.

“We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain... Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. ... Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles ... our findings expose a threat to the privacy and safety of gay men and women.”

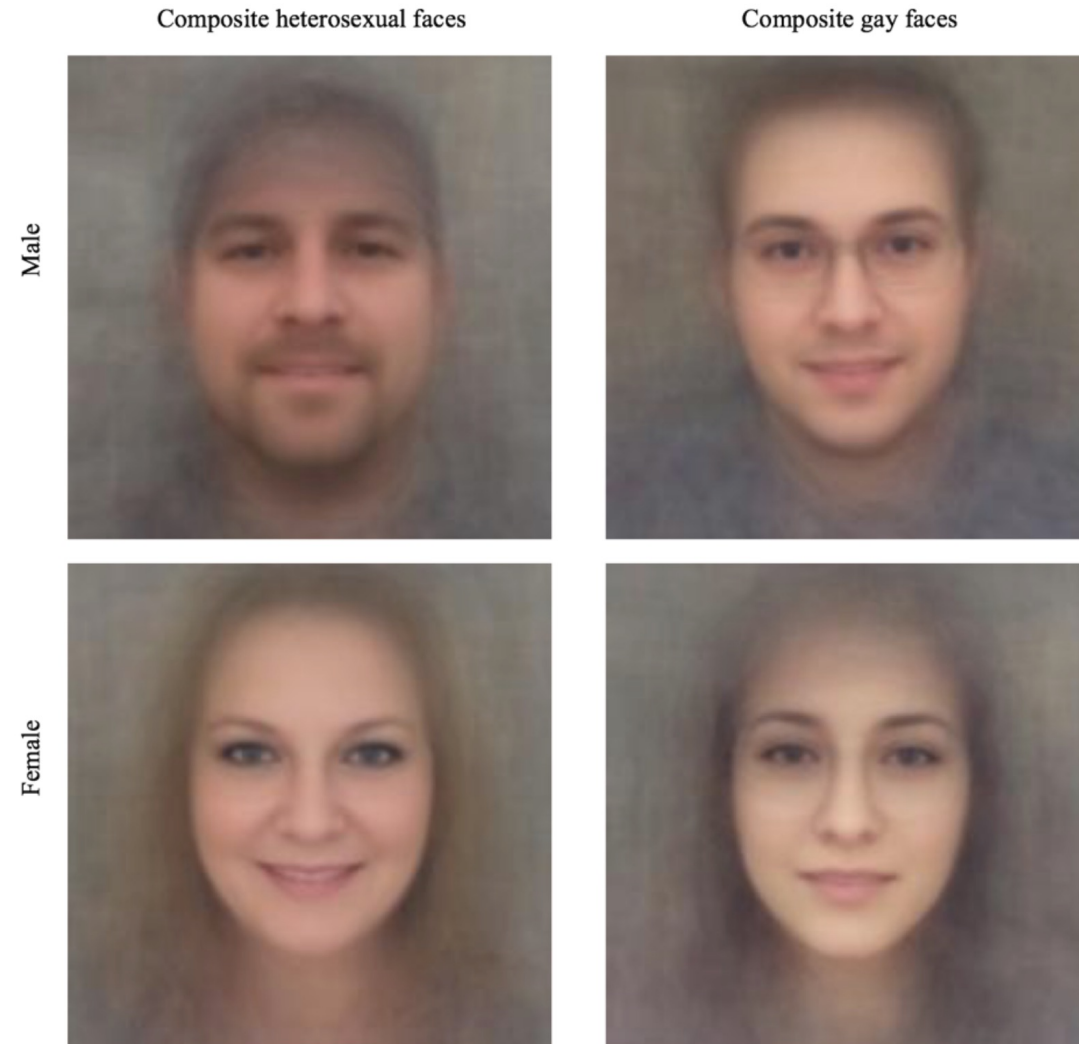
Wang & Kosinski 2017

# More questions

- **Does the application align with your values?**
- Does the task specification / evaluation metric reflect the things you care about?
- For recognition:
  - **Does the collected training / test set match your true distribution?**
- Are the algorithm's errors biased?
- **Are you being honest in public descriptions of your results?**
- Is the accuracy/correctness sufficient for public release?

# Answers

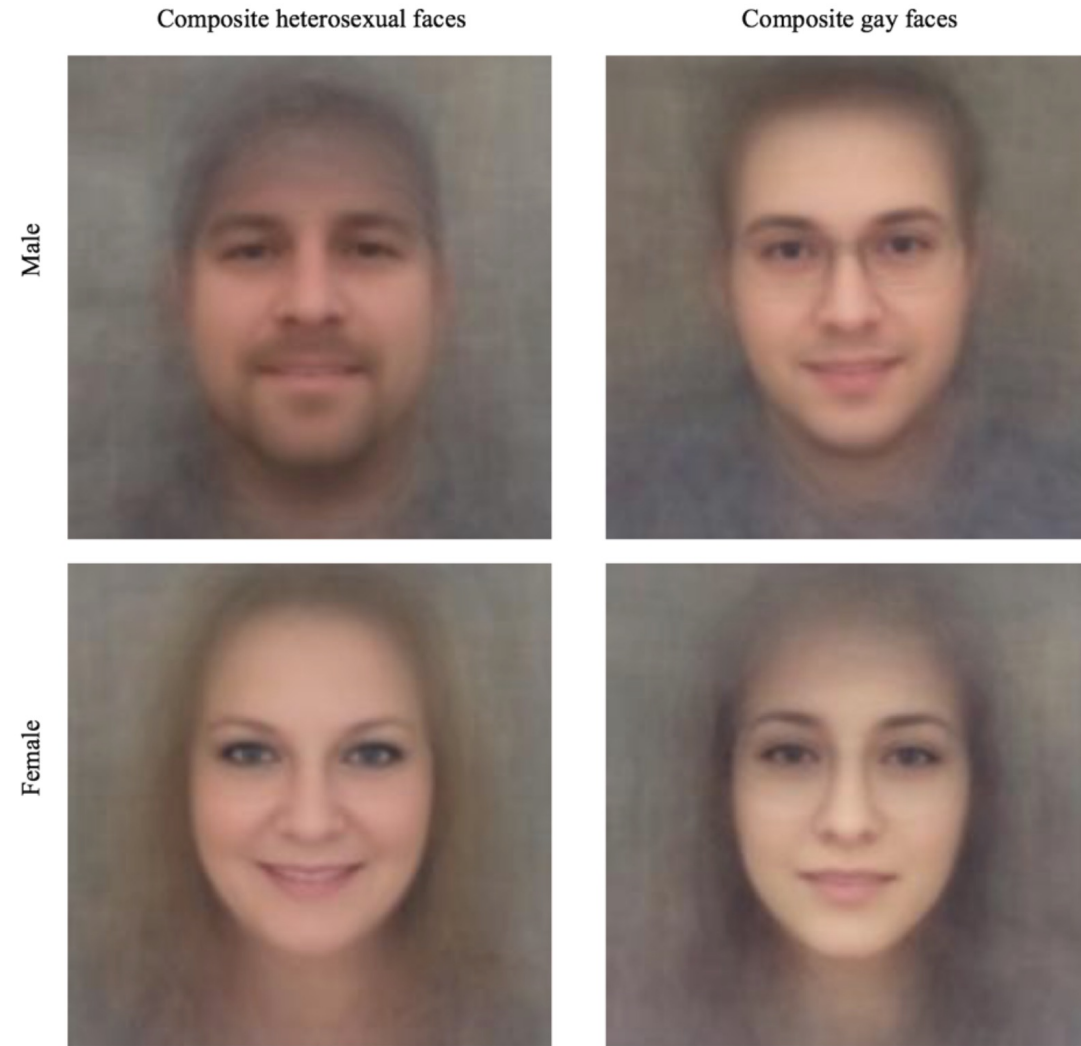
- Training / test set?
  - 35,326 images from public profiles on a **US dating website**
- “average” images of straight/gay people:
- Question:
  - Are differences caused by actual differences in faces?
  - Or how people choose to present themselves in dating websites?





# Answers

- Goal: raise privacy concerns.
- Side-effects?
  - Reinforces potentially harmful stereotypes
  - Provides ostensibly “objective” criteria for discrimination





<https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477>

# Today's Plan

- Course Overview
- Why Computer Vision?
- Computer Vision in real world
- **Ethics in Computer Vision**
  - Examples of bias in computer vision and beyond
  - Datasets and unintended consequences
  - Algorithms (Discriminative & Generative)

# Bias in computer vision and beyond

- What follows are a number of examples of bias from the last 100 years

# Shirley cards



Example Kodak Shirley Card,  
1950s and beyond



Kodak's Multiracial Shirley Card, North  
America. 1995.

How Kodak's Shirley Cards Set Photography's Skin-Tone Standard

<https://www.npr.org/2014/11/13/363517842/for-decades-kodak-s-shirley-cards-set-photography-s-skin-tone-standard>

The Racial Bias Built Into Photography

<https://www.nytimes.com/2019/04/25/lens/sarah-lewis-racial-bias-photography.html>

# Face recognition

- Probably the most controversial vision technology
- Three different versions:
  - Face verification: “Is this person Roni?” (e.g., Apple’s Face Unlock)
  - Face clustering: “Who are all the people in this photo collection”? (e.g., Google Photos search)
  - Face recognition: “Who is this person”? (e.g., identify a person from surveillance footage of a crime scene)
- Applications can suffer from bias (working well for some populations but not others) and misuse

# *Many Facial-Recognition Systems Are Biased, Says U.S. Study*

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



NYT, 12/20/19

<https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html>

Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images



# *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>







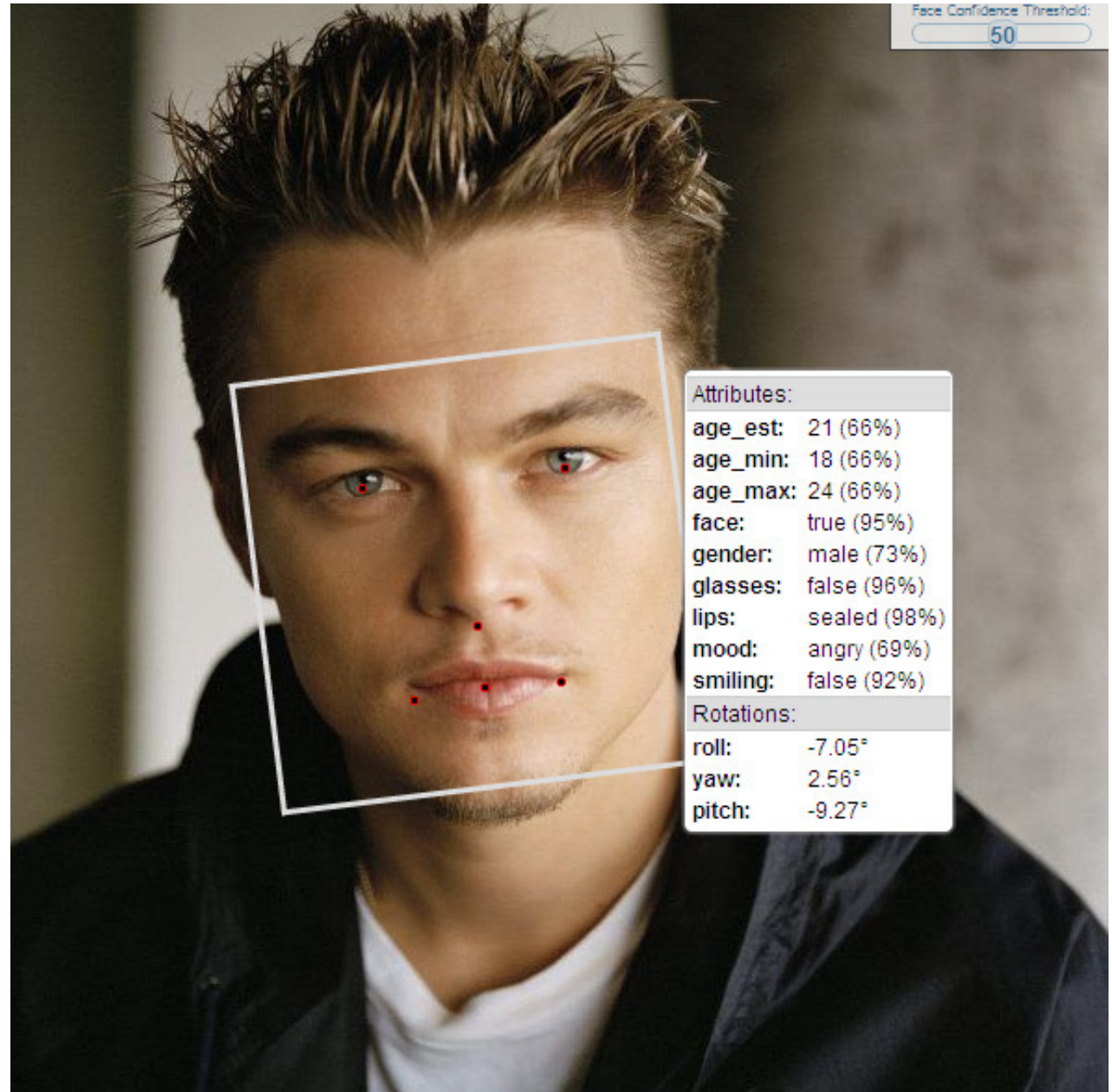
# The Secretive Company That Might End Privacy as We Know It

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.



# Face analysis

- Gender classification
- Age regression
- Expression classification
- Ethnicity classification

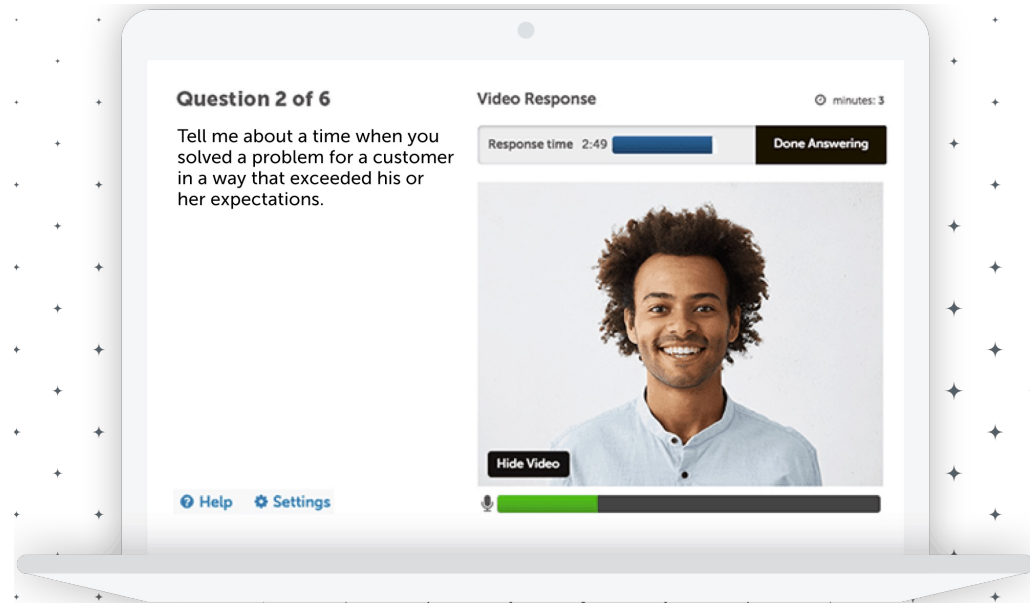


# Example: Video Interviewing

## Technology

# A face-scanning algorithm increasingly decides whether you deserve the job

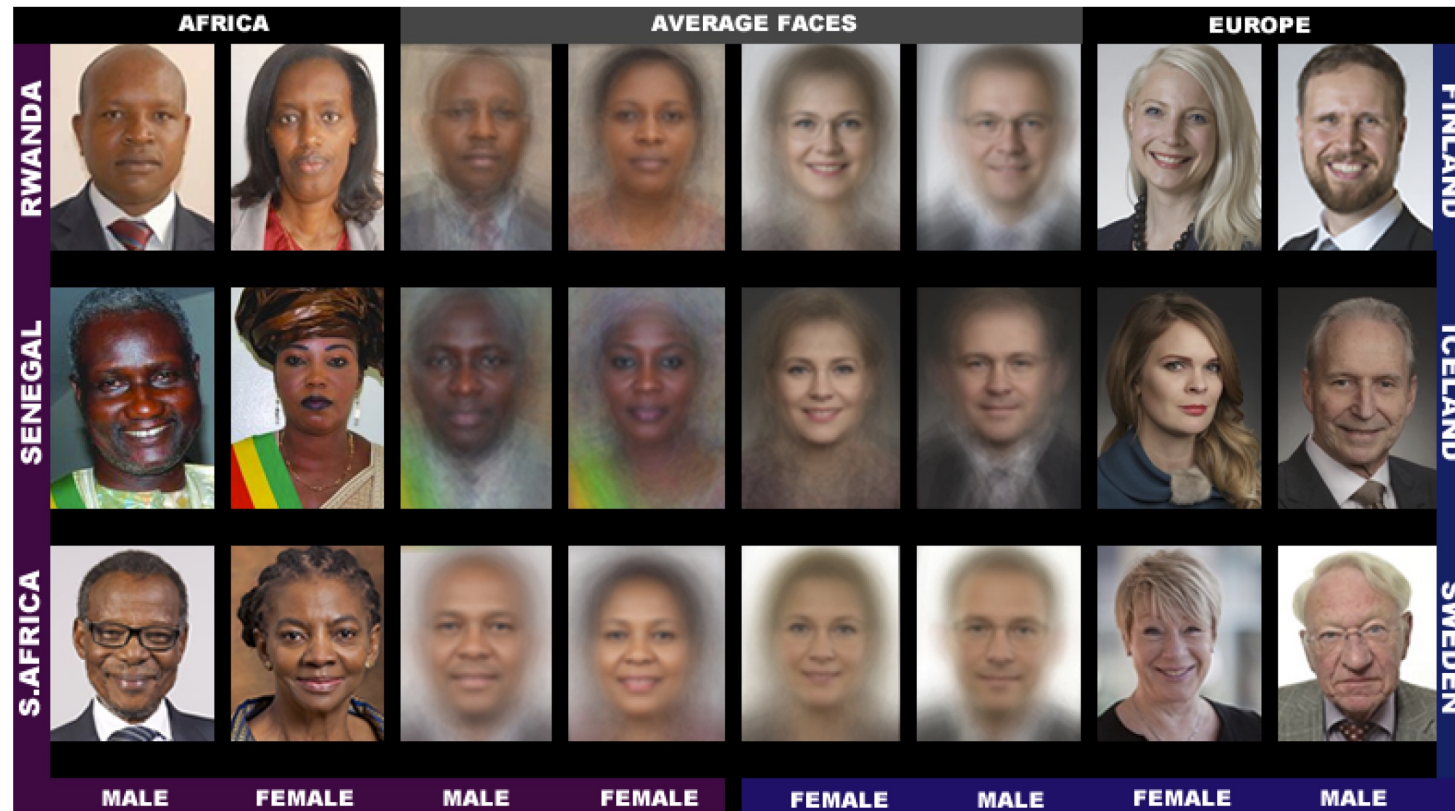
HireVue claims it uses artificial intelligence to decide who's best for a job. Outside experts call it 'profoundly disturbing.'



Source: <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>  
<https://www.hirevue.com/platform/online-video-interviewing-software>

Example Credit: Timnit Gebru

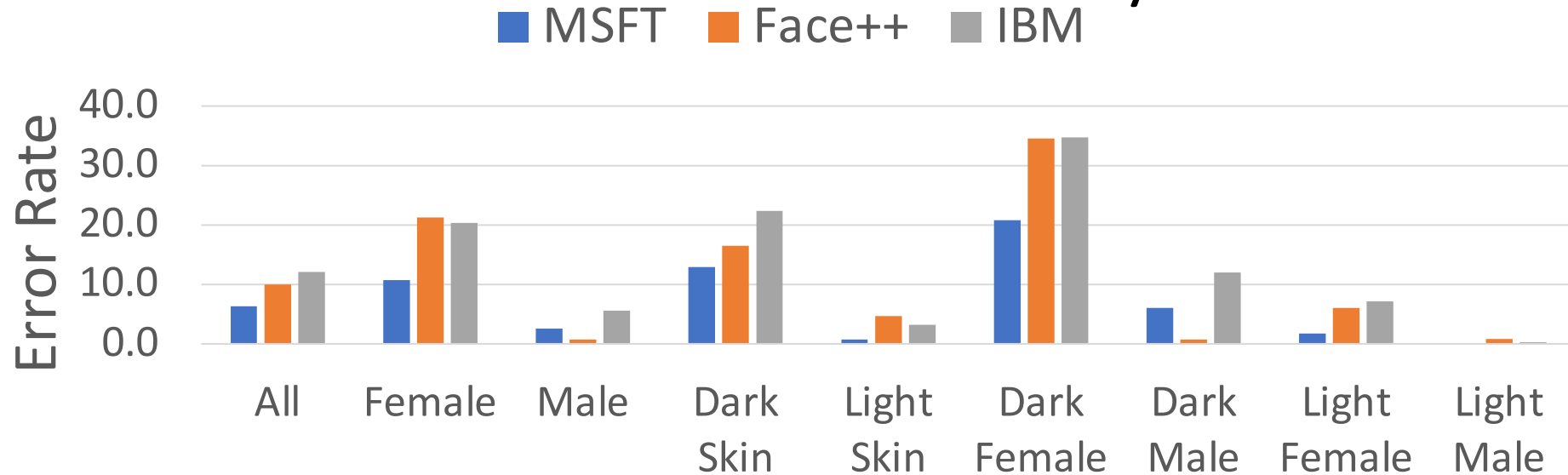
# Gender Shades – Evaluation of bias in Gender Classification



Images from the Pilot Parliaments Benchmark

Joy Buolamwini and Timnit Gebru. **Gender shades: Intersectional accuracy disparities in commercial gender classification.** Conference on Fairness, Accountability and Transparency. 2018.

# Gender Shades: Intersectionality



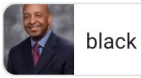
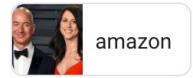
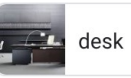
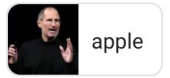
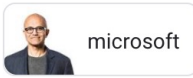
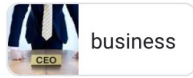
**Problem:** Much higher error rate for dark-skinned women

**Bigger Problem:** Why are we classifying gender at all?  
Why does an automated system care? If it does, ask!



First woman: CEO Barbie =(

Source: <https://www.bbc.com/news/newsbeat-32332603>



Chief executive officer - Wikipedia  
en.wikipedia.org



CEO vs. Owner: The Key Differences ...  
onlinemasters.ohio.edu



How to use 'CEO magic' when tryi...  
europeanceo.com



Odilon Almeida as President ...  
businesswire.com



You are the CEO of Your Life - Person...  
personalexcellence.co



Harvard study: What CEOs do all day  
cnbc.com



CEO doesn't believe in CX ...  
heartofthecustomer.com



7 Personality Traits Every CEO Shoul...  
forbes.com



Roeland Baan new CEO of Haldor T...  
blog.topsoe.com



Wartime CEOs are not the ideal leaders ...  
ft.com



# Racial Profiling: Uighur Classification



Source: <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>



# Today's Plan

- Course Overview
- Why Computer Vision?
- Computer Vision in real world
- **Ethics in Computer Vision**
  - Examples of bias in computer vision and beyond
  - **Datasets and unintended consequences**
  - Algorithms (Discriminative & Generative)

# Datasets – Potential Issues

- Licensing and ownership of data
- Consent of photographer and people being photographed
- Offensive content
- Bias and underrepresentation
  - Including amplifying bias
- Unintended downstream uses of data

# Think Critically about Datasets

CelebA Dataset: 202k images labeled with 40 binary attributes



Liu et al, "Deep Learning Face Attributes in the Wild", ICCV 2015

# Think Critically about Datasets

CelebA Dataset: 202k images labeled with 40 binary attributes

5_o_Clock_Shadow	Double_Chin	Pointy_Nose
Arched_Eyebrows	Eyeglasses	Receding_Hairline
Attractive	Goatee	Rosy_Cheeks
Bags_Under_Eyes	Gray_Hair	Sideburns
Bald	Heavy_Makeup	Smiling
Bangs	High_Cheekbones	Straight_Hair
Big_Lips	Male	Wavy_Hair
Big_Nose	Mouth_Slightly_Open	Wearing_Earrings
Black_Hair	Mustache	Wearing_Hat
Blond_Hair	Narrow_Eyes	Wearing_Lipstick
Blurry	No_Beard	Wearing_Necklace
Brown_Hair	Oval_Face	Wearing_Necktie
Bushy_Eyebrows	Pale_Skin	Young
Chubby		

# Think Critically about Datasets

CelebA Dataset: 202k images labeled with 40 binary attributes

5\_o\_Clock\_Shadow

Arched\_Eyebrows

**Attractive**

Bags\_Under\_Eyes

Bald

Bangs

**Big\_Lips**

**Big\_Nose**

Black\_Hair

Blond\_Hair

Blurry

Brown\_Hair

**Bushy\_Eyebrows**

**Chubby**

**Double\_Chin**

Eyeglasses

Goatee

Gray\_Hair

**Heavy\_Makeup**

High\_Cheekbones

Male

Mouth\_Slightly\_Open

Mustache

Narrow\_Eyes

No\_Beard

Oval\_Face

**Pale\_Skin**

**Pointy\_Nose**

Receding\_Hairline

Rosy\_Cheeks

Sideburns

Smiling

Straight\_Hair

Wavy\_Hair

Wearing\_Earrings

Wearing\_Hat

Wearing\_Lipstick

Wearing\_Necklace

Wearing\_Necktie

**Young**

Many attributes seem subjective. Who chose the attributes?  
Why? How are they defined? Who labeled the images?

# Think Critically about Datasets

CelebA Dataset: 202k images labeled with 40 binary attributes

5\_o\_Clock\_Shadow    **Double\_Chin**    **Pointy\_Nose**  
Arched\_Eyebrows    Eyeglasses    Receding\_Hairline  
**Attractive**    **Almost no detail in the paper**    **Beary\_Cheeks**  
Pale\_Under\_Eyes    Gray\_Hair    Sidburne

images of 5,749 identities. Each image in CelebA and LFWA is annotated with forty face attributes and five key points by a professional labeling company. CelebA and LFWA have over eight million and five hundred thousand attribute labels, respectively.

Blurry    No\_Beard    Wearing\_Necklace  
Brown\_Hair    Oval\_Face    Wearing\_Necktie  
**Bushy\_Eyebrows**    **Pale\_Skin**    **Young**  
**Chubby**

Liu et al, "Deep Learning Face Attributes in the Wild", ICCV 2015

Many attributes seem subjective. Who chose the attributes?  
Why? How are they defined? Who labeled the images?

# Problem: Datasets are Biased

\*This analysis conflates gender with sex, and assumes that it is binary.

## Example: COCO Dataset



Multilabel  
Classification

Person

Umbrella

Cat

Define “gender bias” of object category  $C$  as:

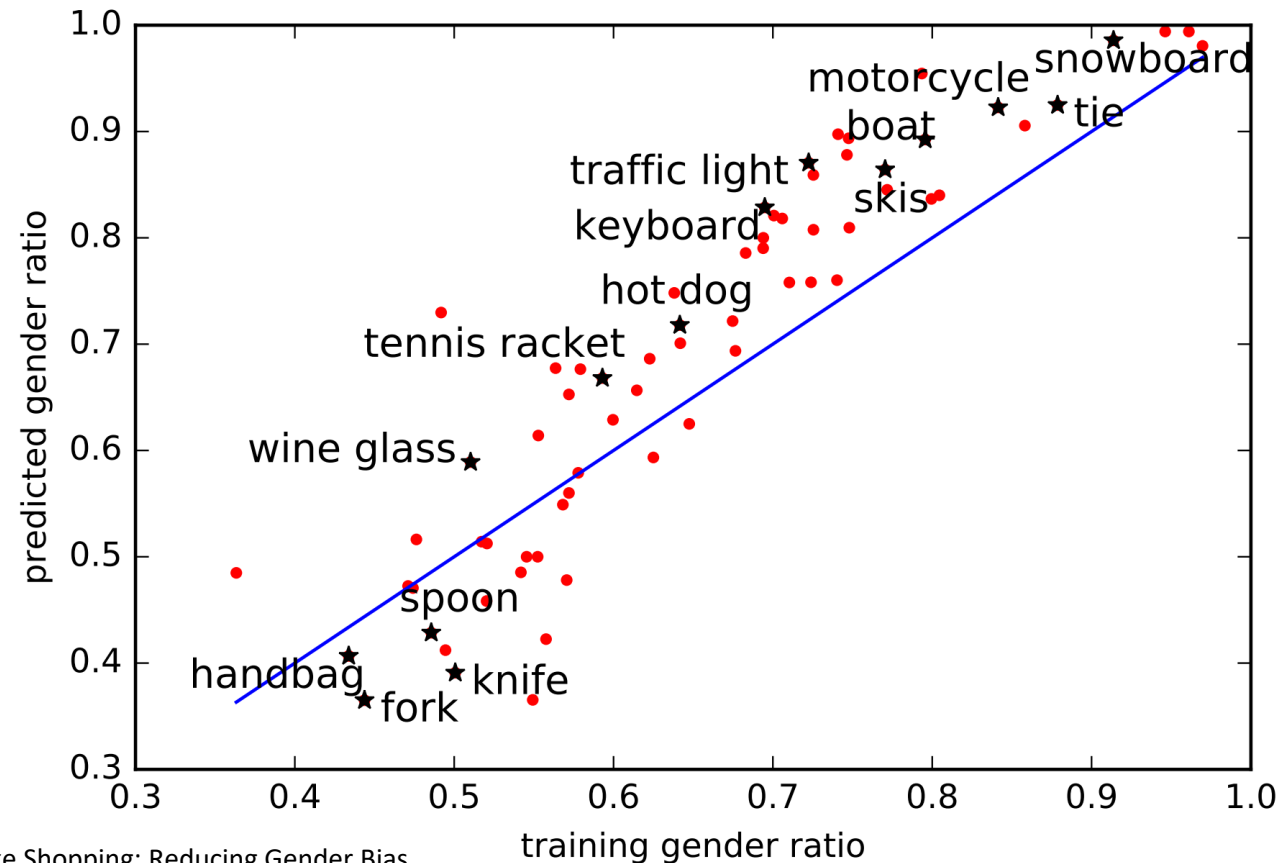
$$\frac{\#(C, Man)}{\#(C, Man) + \#(C, Woman)}$$

**Example: “Snowboards” are 90% biased towards men**

# Problem: Bias Amplification

CNN predictions are **more biased** than their training data!

Reducing bias in datasets is **not enough**





# Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy

**Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, Olga Russakovsky**

## Abstract

Computer vision technology is being used by many but remains representative of only a few. People have reported misbehavior of computer vision models, including offensive prediction results and lower performance for underrepresented groups. Current computer vision models are typically developed using datasets consisting of manually annotated images or videos; the data and label distributions in these datasets are critical to the models' behavior. In this paper, we examine ImageNet, a large-scale ontology of images that has spurred the development of many modern computer vision methods. We consider three key factors within the person subtree of ImageNet that may lead to problematic behavior in downstream computer vision technology: (1) the stagnant concept vocabulary of WordNet, (2) the attempt at exhaustive illustration of all categories with images, and (3) the inequality of representation in the images within concepts. We seek to illuminate the root causes of these concerns and take the first steps to mitigate them constructively.

# Datasheets for Datasets

Idea: A standard list of questions to answer when releasing a dataset. Who created it? Why? What is in it? How was it labeled?

**A Database for Studying Face Recognition in Unconstrained Environments**

**Labeled Faces in the Wild**

## Motivation

**For what purpose was the dataset created?** Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

Labeled Faces in the Wild was created to provide images that can be used to study face recognition in the unconstrained setting where image characteristics (such as pose, illumination, resolution, focus), subject demographic makeup (such as age, gender, race) or appearance (such as hairstyle, makeup, clothing) cannot be controlled. The dataset was created for the specific task of pair matching: given a pair of images each containing a face, determine whether or not the images are of the same person.<sup>1</sup>

**Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

The initial version of the dataset was created by Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, most of whom were researchers at the University of Massachusetts Amherst at the time of the dataset's release in 2007.

**Who funded the creation of the dataset?** If there is an associated grant, please provide the name of the grantor and the grant name and number.

The construction of the LFW database was supported by a United States National Science Foundation CAREER Award.

The dataset does not contain all possible instances. There are no known relationships between instances except for the fact that they are all individuals who appeared in news sources on line, and some individuals appear in multiple pairs.

**What data does each instance consist of?** "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

Each instance contains a pair of images that are 250 by 250 pixels in JPEG 2.0 format.

**Is there a label or target associated with each instance?** If so, please provide a description.

Each image is accompanied by a label indicating the name of the person in the image.

**Is any information missing from individual instances?** If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included in the dataset.

**Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?** If so, please describe how these relationships are made explicit.

There are no known relationships between instances except for the fact that they are all individuals who appeared in news sources

# Today's Plan

- Course Overview
- Why Computer Vision?
- Computer Vision in real world
- **Ethics in Computer Vision**
  - Examples of bias in computer vision and beyond
  - Datasets and unintended consequences
  - **Algorithms (Discriminative & Generative)**

# Economic Bias in Visual Classifiers



**Ground-Truth:** Soap

**Source:** UK, \$1890/month

**Azure:** toilet, design, art, sink

**Clarifai:** people, faucet, healthcare, lavatory, wash closet

**Google:** product, liquid, water, fluid, bathroom accessory

**Amazon:** sink, indoors, bottle, sink faucet

**Watson:** gas tank, storage tank, toiletry, dispenser, soap dispenser

**Tencent:** lotion, toiletry, soap dispenser, dispenser, after shave



**Ground-Truth:** Soap

**Source:** Nepal, \$288/month

**Azure:** food, cheese, bread, cake, sandwich

**Clarifai:** food, wood, cooking, delicious, healthy

**Google:** food, dish, cuisine, comfort food, spam

**Amazon:** food, confectionary, sweets, burger

**Watson:** food, food product, turmeric, seasoning

**Tencent:** food, dish, matter, fast food, nutriment

# Model Cards

Idea: A standard list of questions to answer when releasing a trained model. Who created it? What data was it trained on? What should it be used for? What should it **not** be used for?

## Model Card

- **Model Details.** Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use.** Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors.** Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors

- Evaluation factors
- **Metrics.** Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data.** Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data.** May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

# Model Cards

# Adopted by Google, OpenAI

## Object Detection

Model Card v0 Cloud Vision API

Overview

- Limitations
- Performance
- Test your own images
- Provide feedback

Explore


- Face Detection
- About Model Cards

### Object Detection

The model analyzed in this card detects one or more physical objects within an image, from apparel and animals to tools and vehicles, and returns a box around each object, as well as a label and description for each object.

On this page, you can learn more about how the model performs on different classes of objects, and what kinds of images you should expect the model to perform well or poorly on.

#### MODEL DESCRIPTION



**Input:** Photo(s) or video(s)


**Output:** The model can detect 550+ different object classes. For each object detected in a photo or video, the model outputs:

- Object bounding box coordinates
- Knowledge graph ID ("MID")
- Label description
- Confidence score

**Model architecture:** Single shot detector model with a Resnet 101 backbone and a feature pyramid network feature map.

[View public API documentation](#)

#### PERFORMANCE



PRECISION 100%

RECALL 100%

Legend: Open Images (blue), Google Internal (green)

Performance evaluated for specific object classes recognized by the model (e.g. shirt, muffin), and for categories of objects (e.g. apparel, food).

Two performance metrics are reported:

- Average Precision (AP)
- Recall at 60% Precision

Performance evaluated on two datasets distinct from the training set:

- Open Images Validation set, which contains ~40k images and 600 object classes, of which the model can recognize 518.
- An internal Google dataset of ~5,000 images of consumer products, containing 210 object classes, all of which model can recognize.

[Go to performance](#)

<https://modelcards.withgoogle.com/object-detection>

## Model Card: CLIP

Inspired by [Model Cards for Model Reporting \(Mitchell et al.\)](#) and [Lessons from Archives \(Jo & Gebru\)](#), we're providing some accompanying information about the multimodal model.

### Model Details

The CLIP model was developed by researchers at OpenAI to learn about what contributes to robustness in computer vision tasks. The model was also developed to test the ability of models to generalize to arbitrary image classification tasks in a zero-shot manner. It was not developed for general model deployment - to deploy models like CLIP, researchers will first need to carefully study their capabilities in relation to the specific context they're being deployed within.

### Model Date

January 2021

### Model Type

The base model uses a ResNet50 with several modifications as an image encoder and uses a masked self-attention Transformer as a text encoder. These encoders are trained to maximize the similarity of (image, text) pairs via a contrastive loss. There is also a variant of the model where the ResNet image encoder is replaced with a Vision Transformer.

### Model Version

Initially, we've released one CLIP model based on the Vision Transformer architecture equivalent to ViT-B/32, along with the RN50 model, using the architecture equivalent to ResNet-50.

As part of the staged release process, we have also released the RN101 model, as well as RN50x4, a RN50 scaled up 4x according to the [EfficientNet](#) scaling rule.

Please see the paper linked below for further details about their specification.

### Documents

- [Blog Post](#)
- [CLIP Paper](#)

### Model Use

#### Intended Use

The model is intended as a research output for research communities. We hope that this model will enable researchers to better understand and explore zero-shot, arbitrary image classification. We also hope it can be used for interdisciplinary studies of the potential impact of such models - the CLIP paper includes a discussion of potential downstream impacts to provide an example for this sort of analysis.

<https://github.com/openai/CLIP/blob/main/model-card.md>

# DeepFakes



<https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher>

# DeepFakes





# DeepFakes

- Active research on both better and better image/video generation and detection of fake images
- Representative work:

CNN-generated images are surprisingly easy to spot...for now

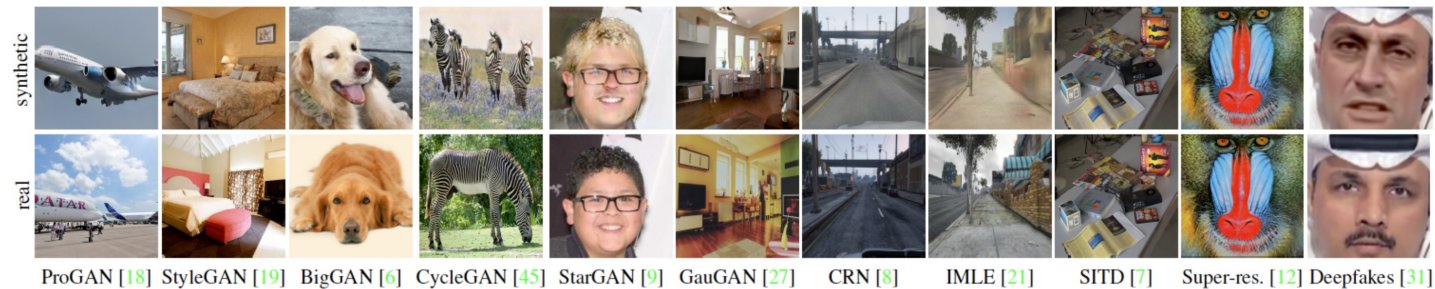
Sheng-Yu Wang<sup>1</sup> Oliver Wang<sup>2</sup> Richard Zhang<sup>2</sup> Andrew Owens<sup>1</sup> Alexei A. Efros<sup>1</sup>

<sup>1</sup>UC Berkeley

<sup>2</sup>Adobe Research

Code [GitHub]

CVPR 2020 (Oral) [Paper]



Are CNN-generated images hard to distinguish from real images? We show that a classifier trained to detect images generated by only one CNN (ProGAN, far left) can detect those generated by many other models (remaining columns).

<https://peterwang512.github.io/CNNDetection/>

# Some tools

- Policy and regulation
  - e.g., a number of cities have banned the use of face recognition by law enforcement
- Transparency
  - e.g., studies on bias in face recognition have led to reforms by tech companies themselves
  - e.g., datasheets can help downstream users of datasets
- Awareness (when you conceive of or build a technology, be aware of the questions we've discussed)

# Questions

- Should I be working on this problem at all?
- Does a given vision task even make sense?
- What are the implications if it doesn't work well?
- What are the implications if it does work well?
- What are the implications if it works well for some people, but not others?
- Who benefits and who is harmed?
- (About datasets) How was it collected? Is it representative?
- (For any technology) Who is it designed for?

# Takeaways

- Thinking about bias and fairness in automated systems goes far beyond computer vision
- People in many fields are thinking about these issues, not just CS
- It's important that the next generation of engineers and scientists (you all!) spend some time thinking about the implications of their work on people and society

# Slide Credits

- CS 5670, Cornell, Prof. Noah Snavely
- COMP 776, UNC, Prof. Jan-Michael Frahm