CSE 252D: Advanced Computer Vision

Manmohan Chandraker

Lecture 0: Introduction

Virtual classrooms

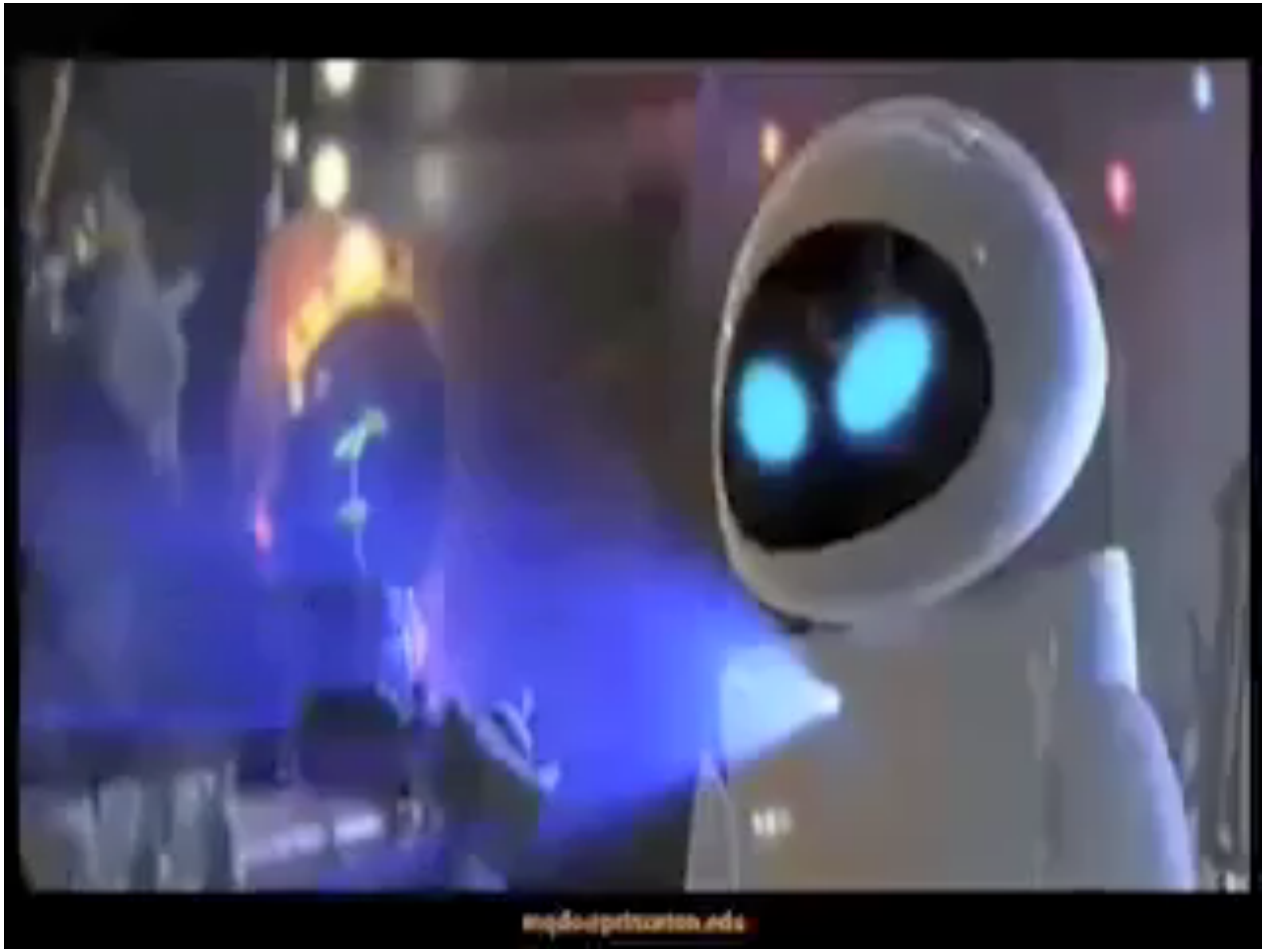
- Virtual lectures on Zoom
  - Only host shares the screen
  - Keep video off and microphone muted
  - But please do speak up (remember to unmute!)
  - Slides uploaded on webpage just before class
- Virtual interactions on Zoom
  - Ask and answer plenty of questions
  - “Raise hand” feature on Zoom when you wish to speak
  - Post questions on chat window
  - Happy to try other suggestions!
- Lectures recorded and upload on Kaltura
  - Available under “My Media” on Canvas

Enrollment logistics

- To enroll if you are on the waitlist
  - Send **"Request to enroll"** email to instructor if on waitlist
  - Include CV, courses, project experience relevant to computer vision
- While on the waitlist
  - You are welcome to attend lectures even if on waitlist
  - To limit TA workload, we can grade only enrolled students
  - Most should be able to enroll eventually
- Canvas
  - All enrolled and waitlisted students should have access
- All announcements will be posted on Piazza
  - Send email to TA (CC instructor) if cannot access Piazza

Computer Vision

Defining computer vision



Wall-E: Fact and Fiction (Minh Do, Princeton University)

Studying computer vision

- Vision is a fundamental interface to the world

Humans

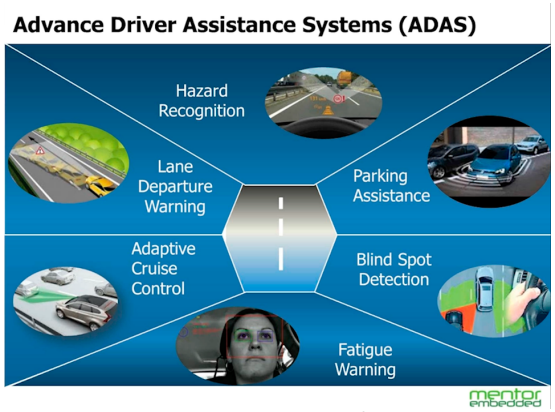
Machines

Virtual
(Cyborgs!)

Source: 2Pmedia

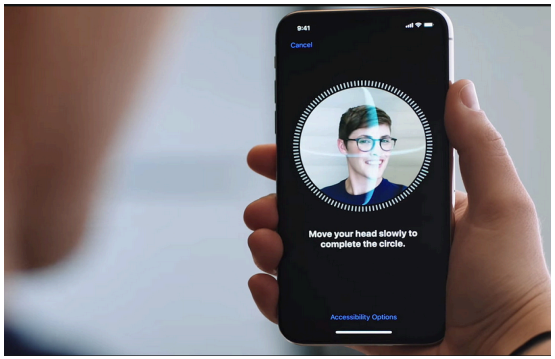
Studying computer vision

- Vision is a fundamental interface to the world
- We are all users of computer vision



Transportation

Design

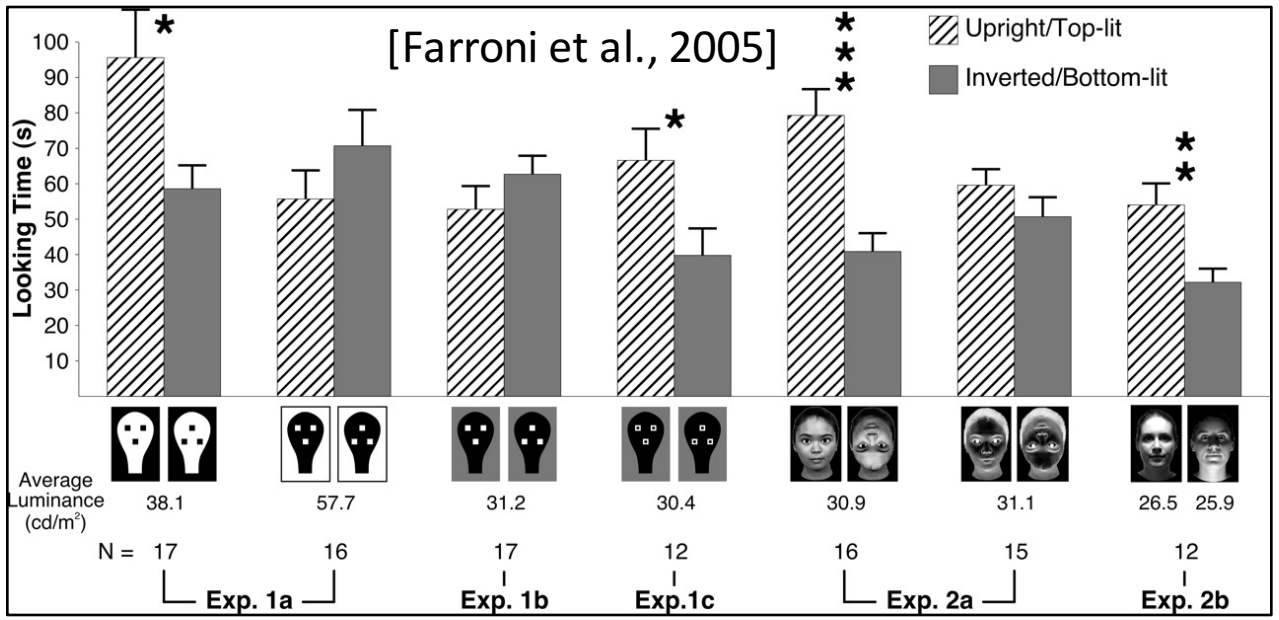


Communication

E-Commerce

Studying computer vision

- Vision is a fundamental interface to the world
- We are all users of computer vision
- Deep and attractive scientific problems
  - How do we recognize objects?
  - Why do newborn babies respond to face-like shapes?



Defining computer vision

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

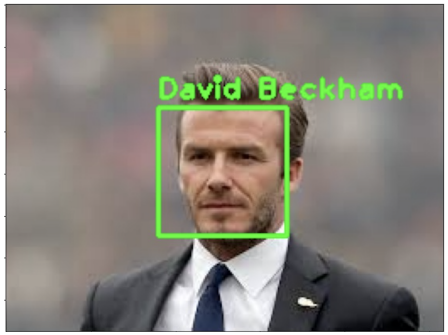
Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

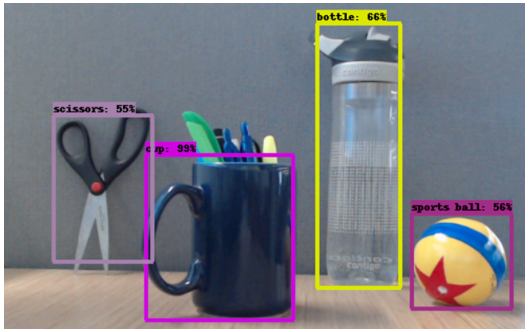
THE SUMMER VISION PROJECT

Seymour Papert

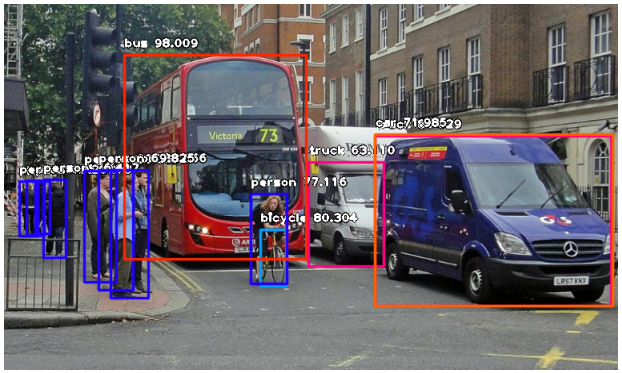
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Defining computer vision

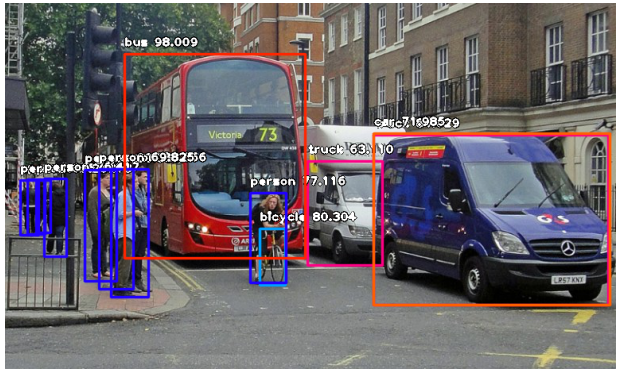




Our problems became more complex

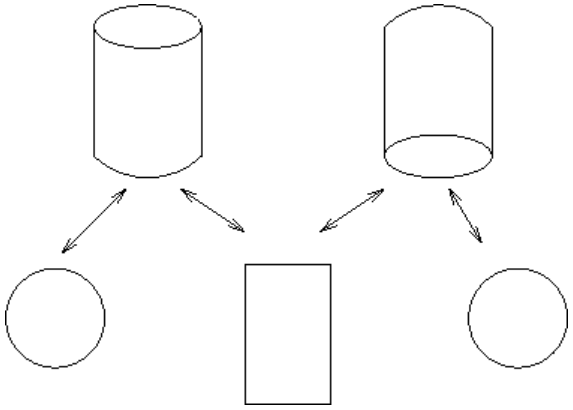


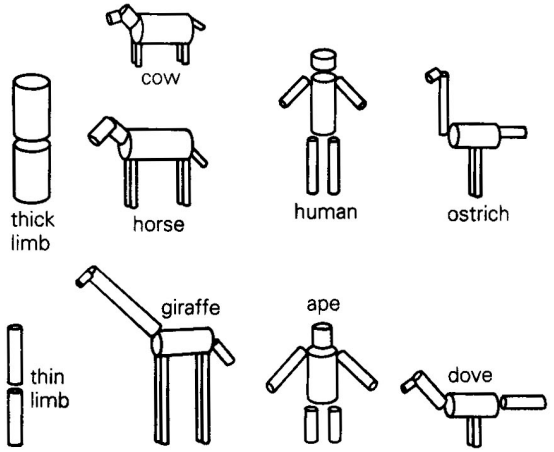
Defining computer vision



Our ambitions became higher

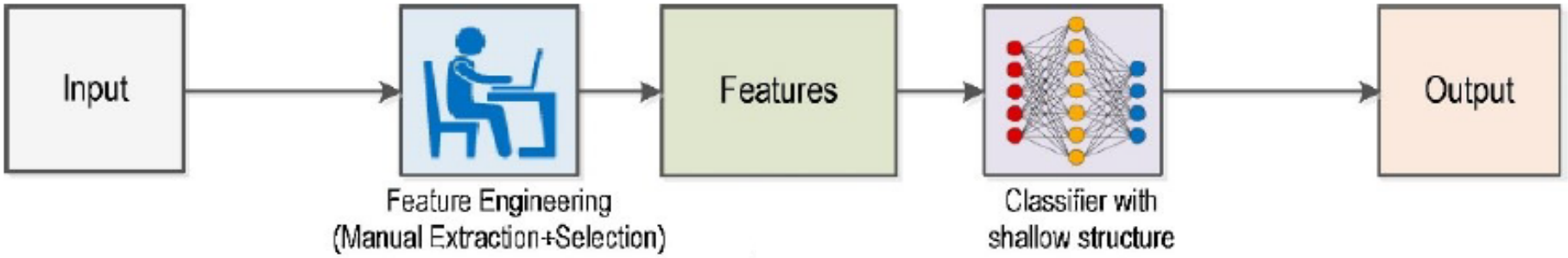
Defining computer vision



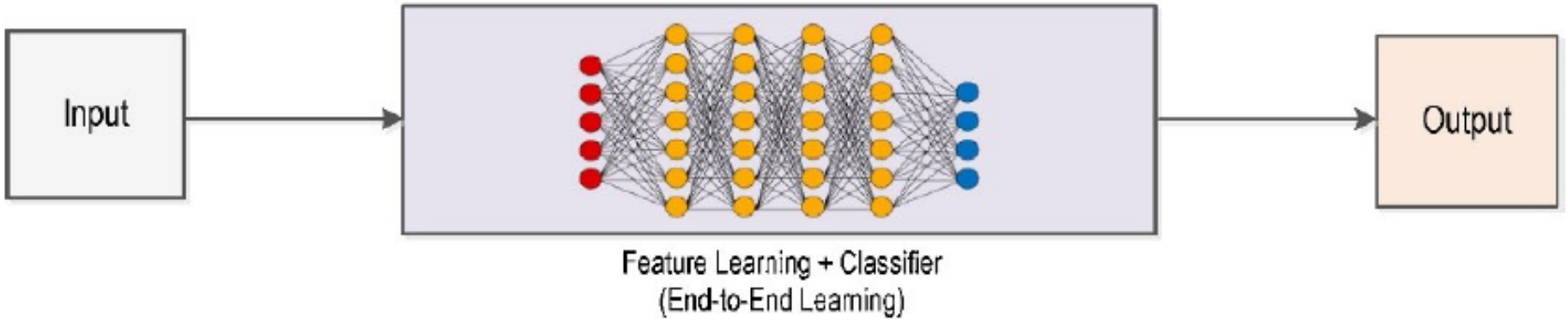


Our concepts changed over time

Defining computer vision



Our tools advanced over time



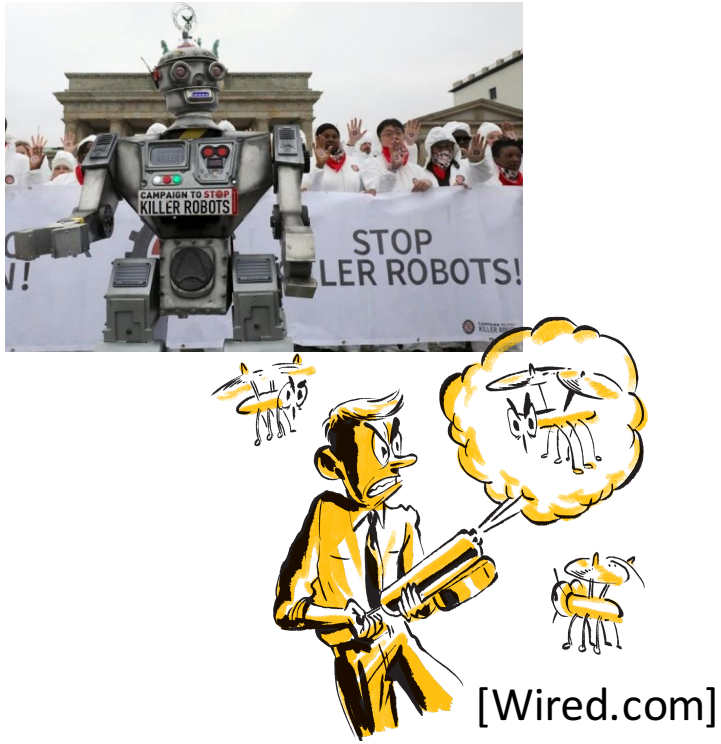
[Mahony et al., 2019]

Gaining perspective on computer vision

Important for Autopilot:

Do not hit a *police* car!

Gaining perspective on computer vision



[Wired.com]

Important for Autopilot:
Do not hit a *police* car!

Important for Robots:
Stay clear of humans!

Gaining perspective on computer vision

[Wired.com]

Important for Autopilot:
Do not hit a *police* car!

Important for Robots:
Stay clear of humans!

Autonomous driving

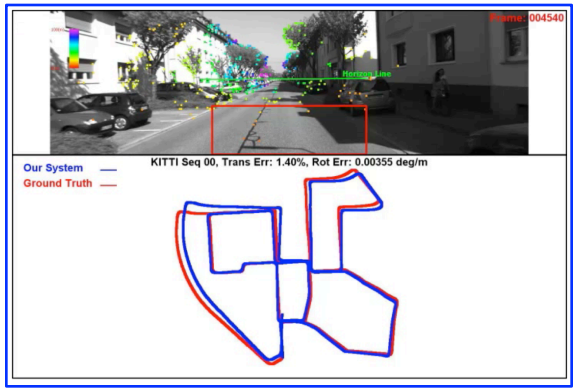
Source: Wired

Autonomous driving

Where is our car?

Structure from Motion

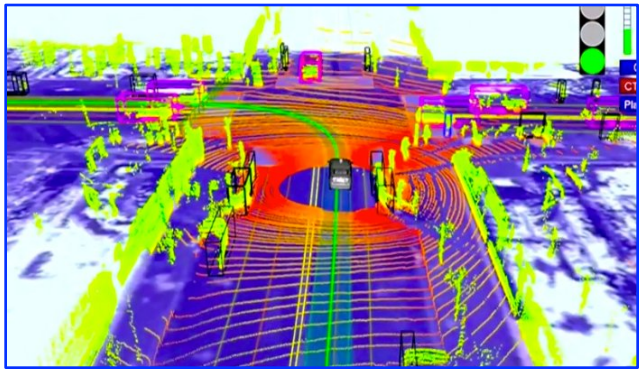
Visual SLAM



What is a safe path?

Behavior prediction

Path planning



Where are other agents?

Object detection

3D localization

Where are scene elements?

Semantic segmentation

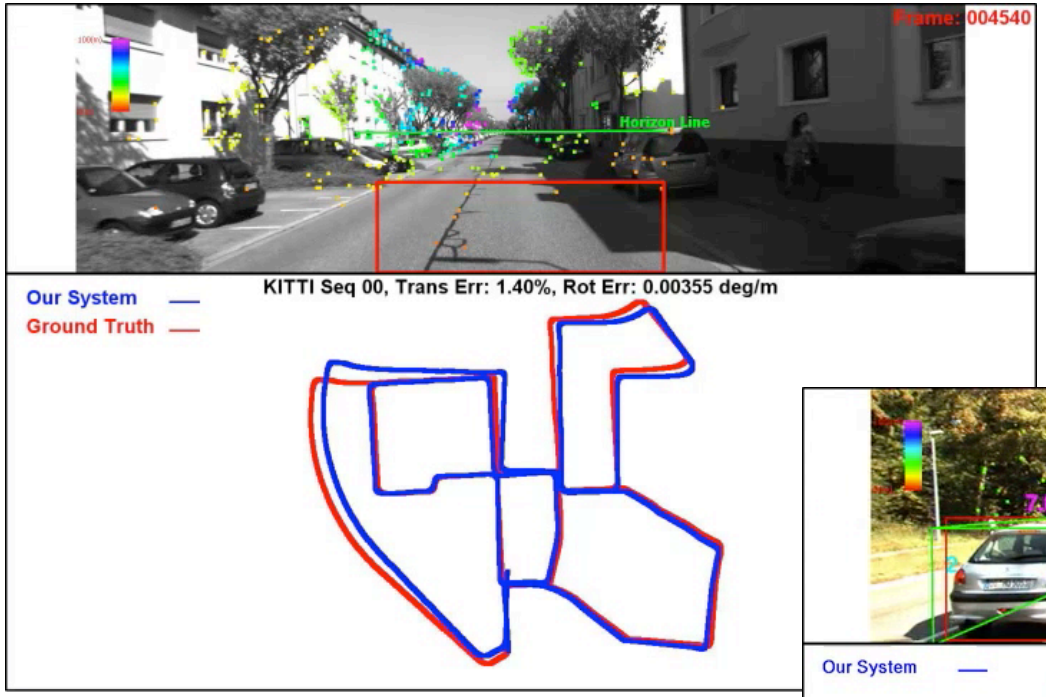
A few topics in this course

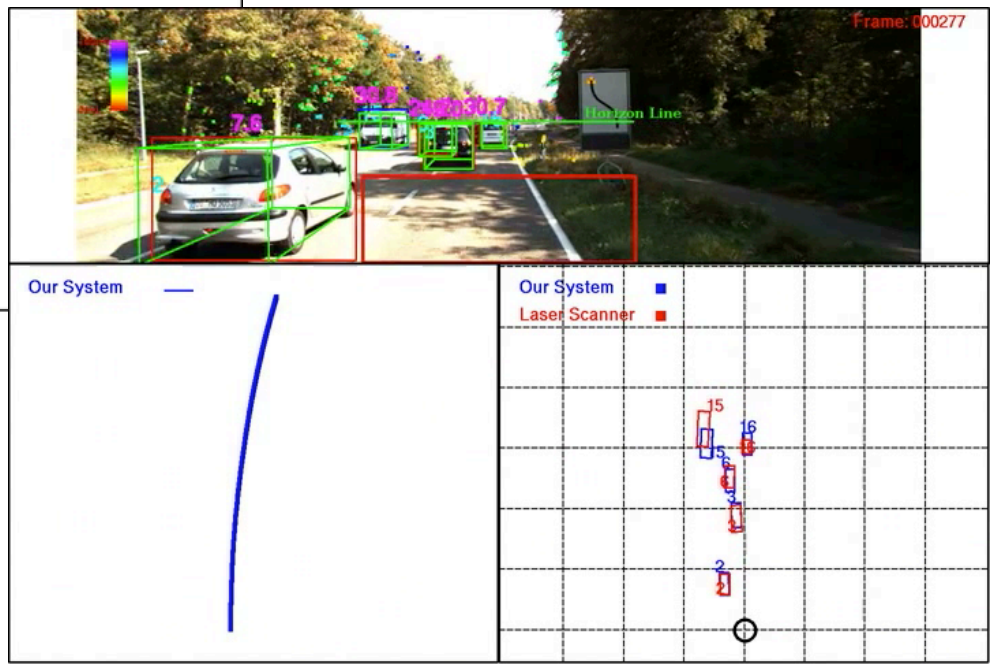
- Structure and Motion
- Faces and Humans
- Objects and Stuff
- Material and Lighting
- Bias and Adaptation

Structure and Motion

[Agarwal et al., ICCV 2009]

Structure and Motion

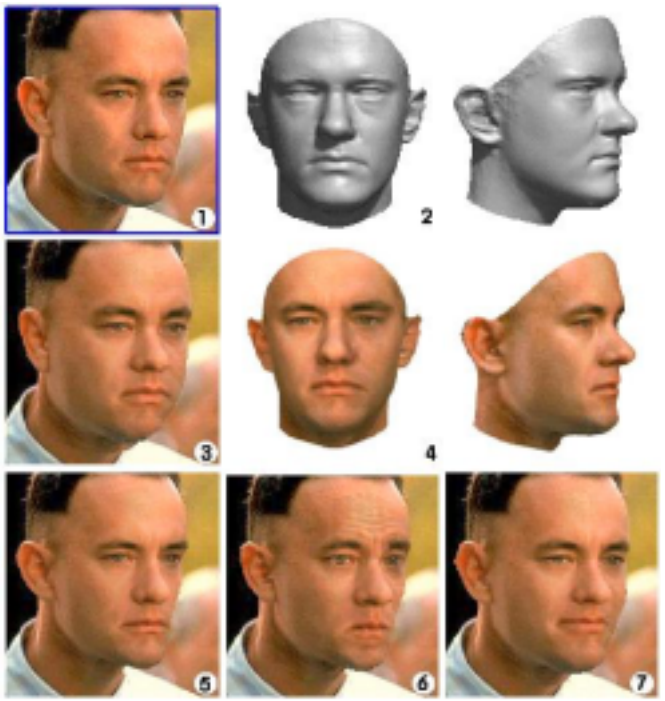




 [Song and Chandraker, CVPR 2015]

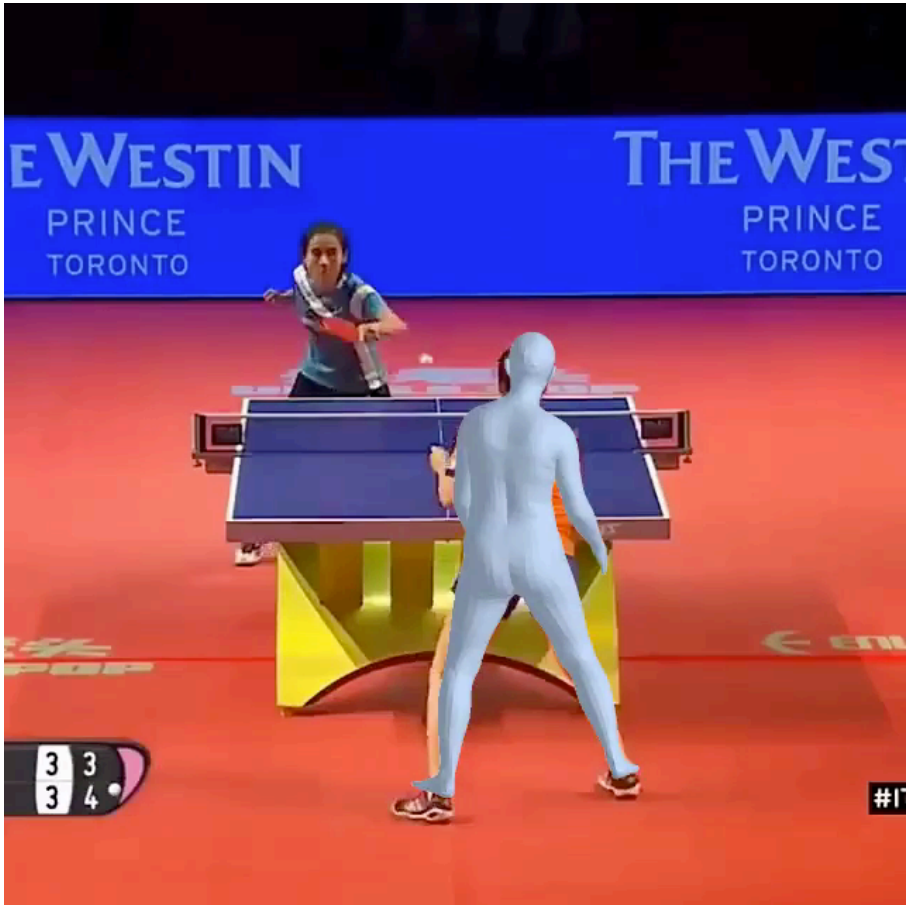
Faces and Humans

[Microsoft blog, 2018]



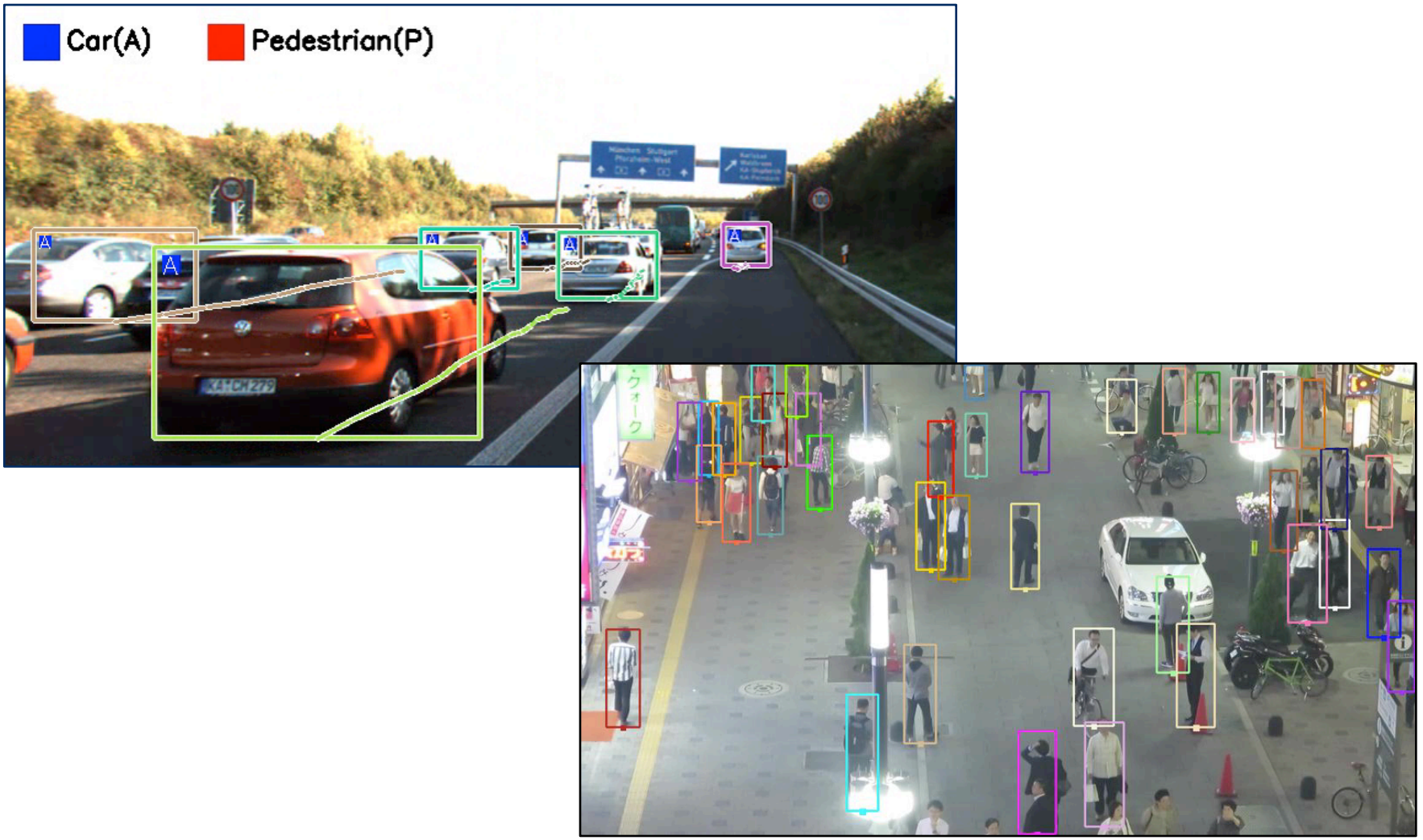
[Blanz and Vetter]

Faces and Humans



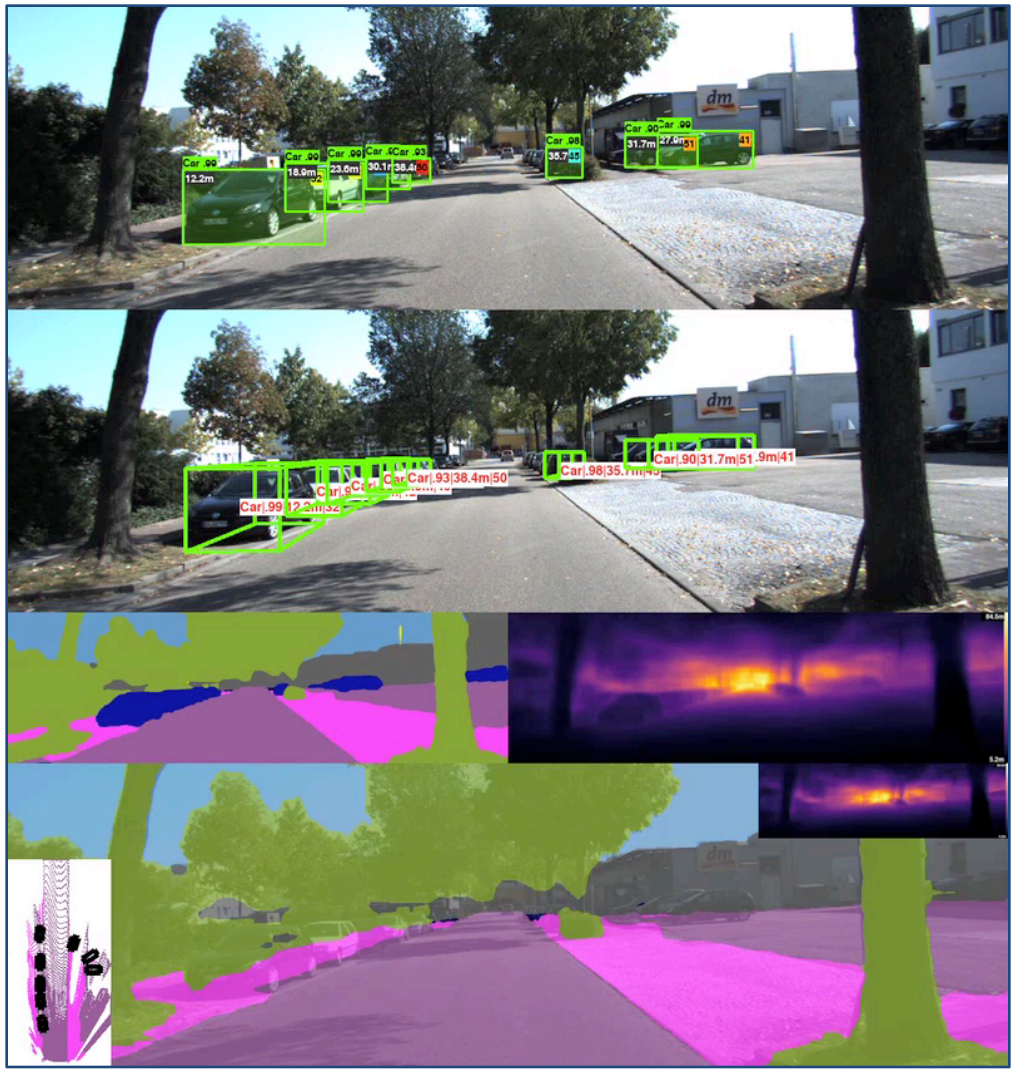
[Kanazawa et al., CVPR 2019]

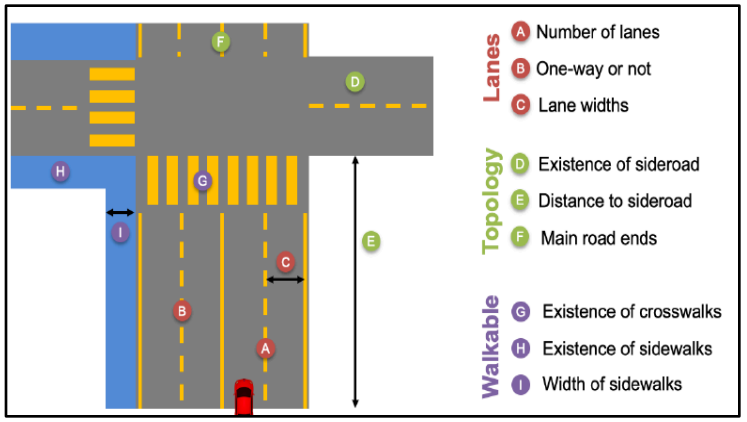
Objects and Stuff



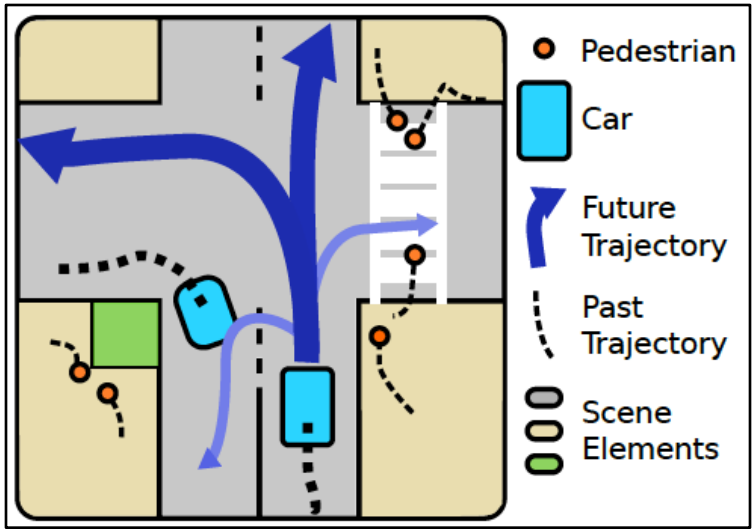
[Choi, ICCV 2015]

Objects and Stuff





[Wang et al., CVPR 2019]

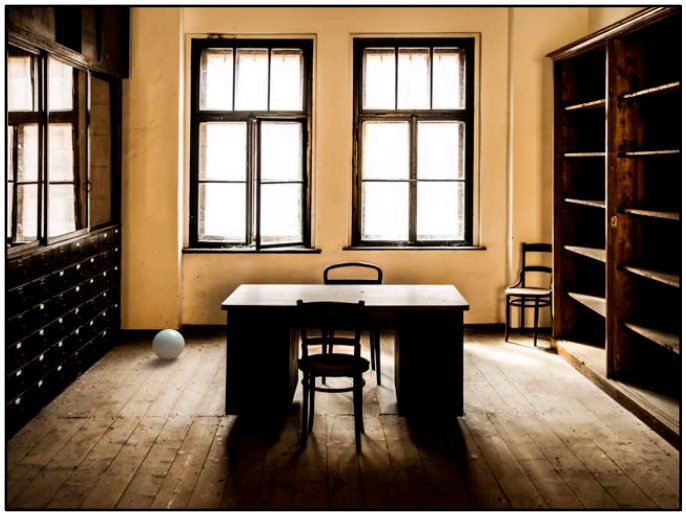


[Lee et al., ICCV 2017]

Material and Lighting

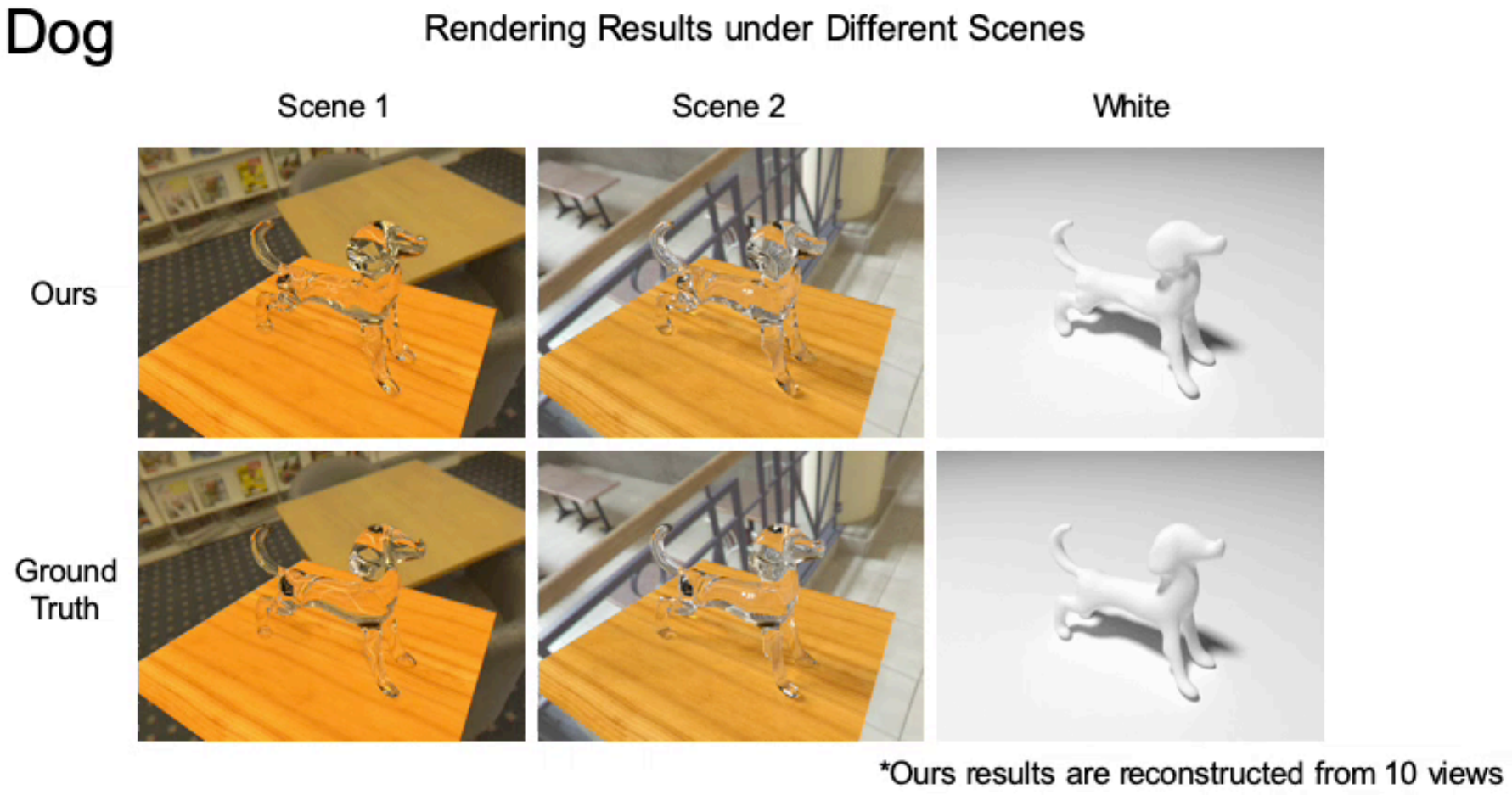
Input image

Augmented image



[Li et al., CVPR 2020]

Material and Lighting



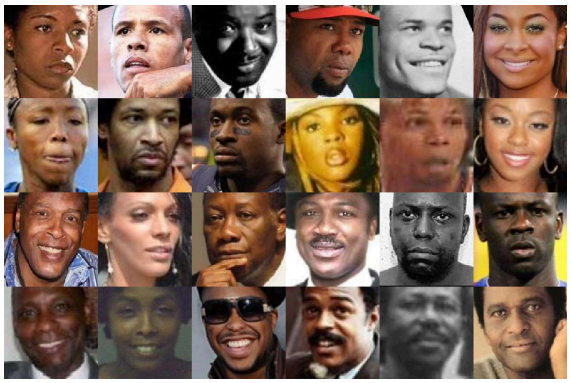
[Li, Yeh, et al., CVPR 2020]

Bias and Privacy

Proportion in datasets: 80%

Caucasian

Proportion in datasets: 10%



African-American

Proportion in datasets: 5%

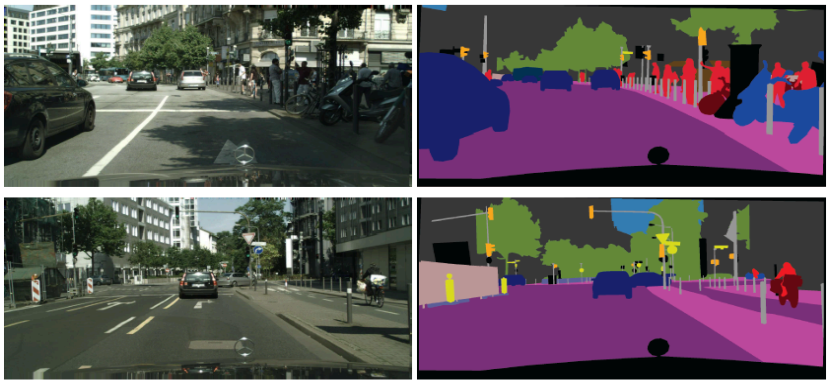
East-Asian

Training on biased data without domain adaptation		
High accuracy	Low accuracy 😕	Low accuracy 😕

Using domain adaptation to address dataset bias		
High accuracy	High accuracy 🙂	High accuracy 🙂

Bias and Privacy

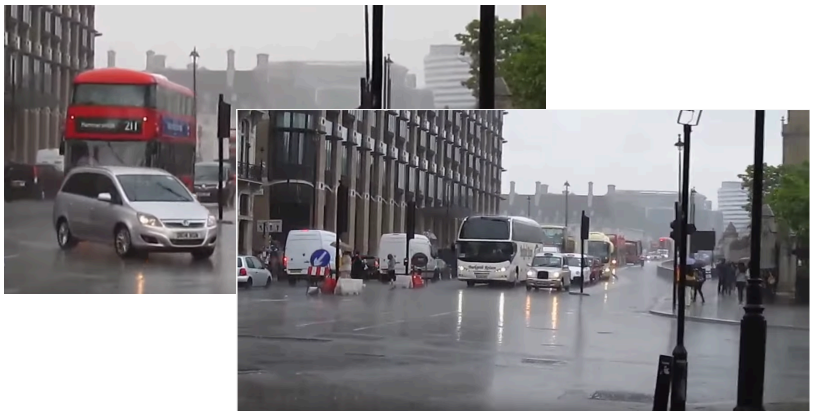
Source domain: good weather, **with** labels



Labels require **1.5 hours** per image!

Train on source, **apply** on target

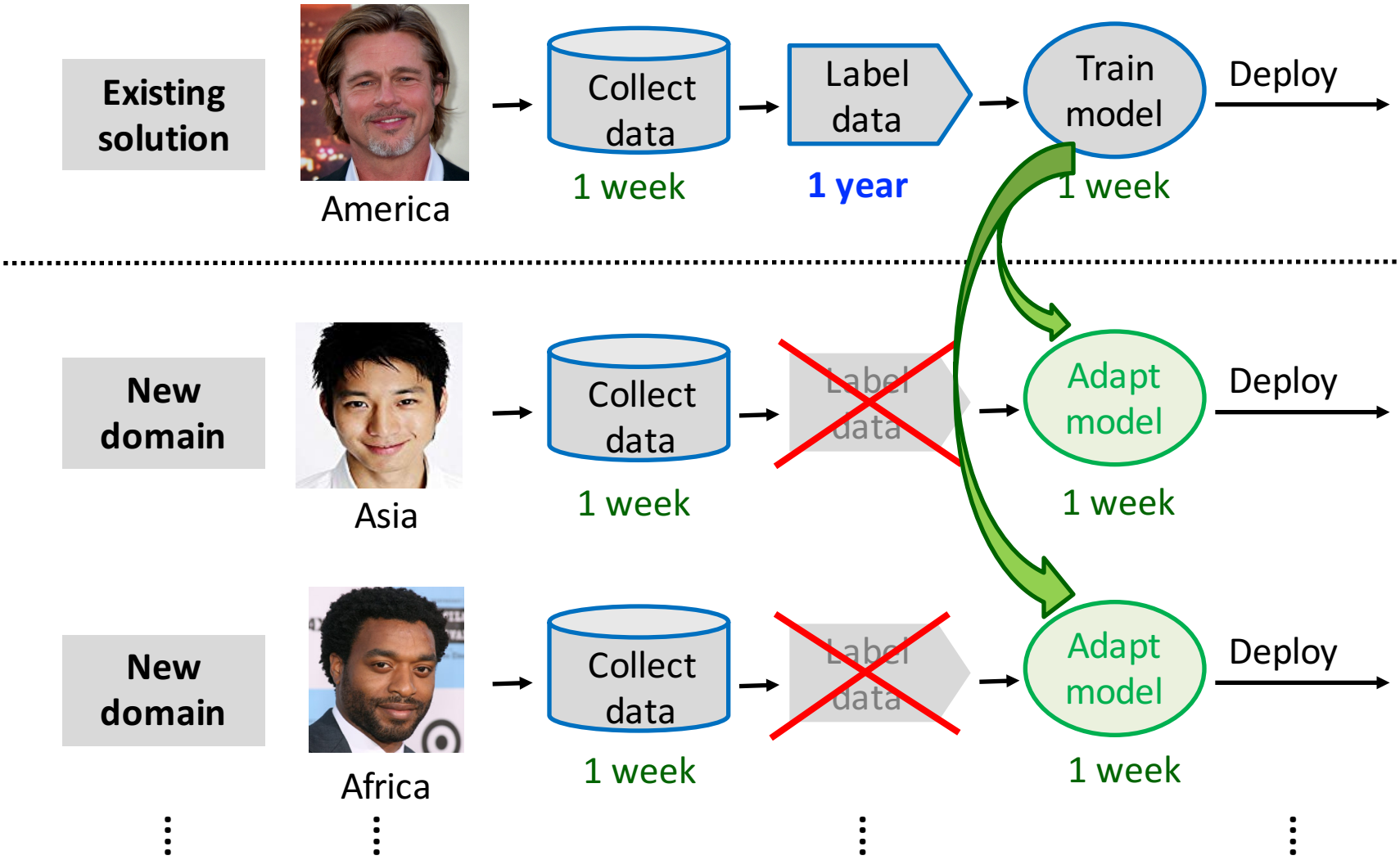
Target domain: rainy weather, **no** labels



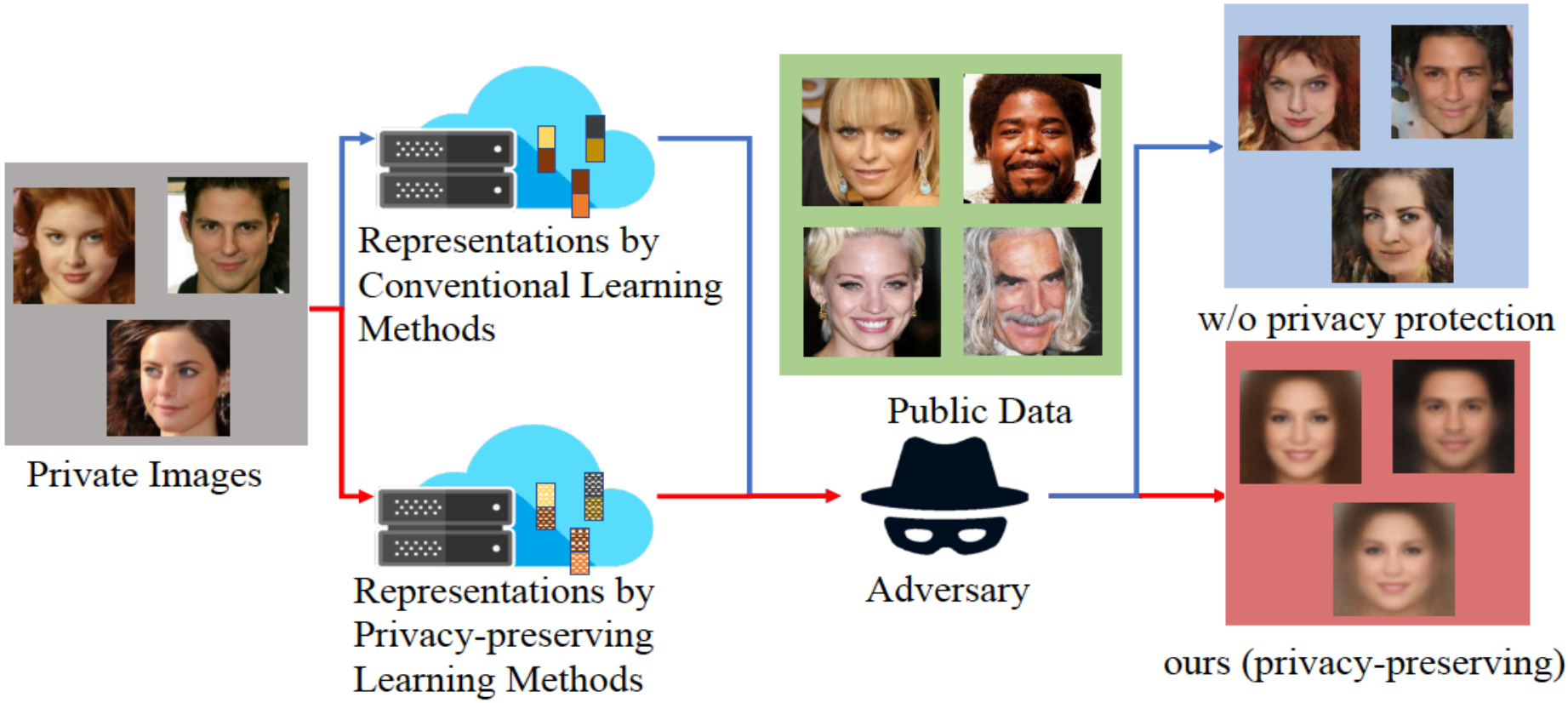
Train on source, **adapt** to target

[Tsai et al., ICCV 2019]

Bias and Privacy



Bias and Privacy

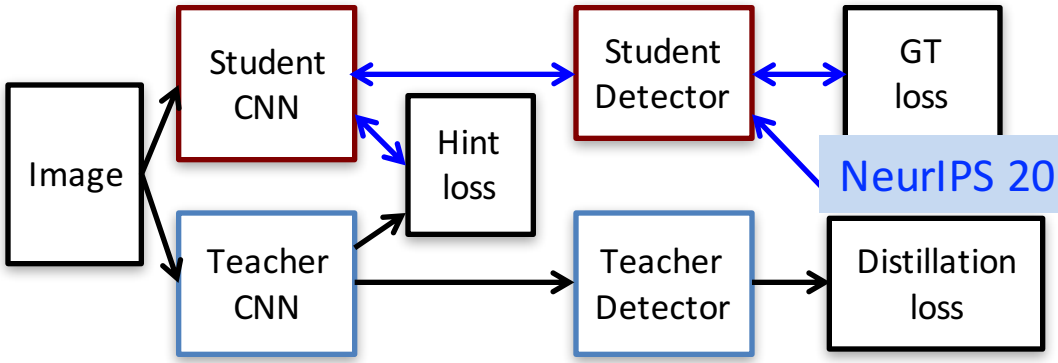


[Xiao et al., AAAI 2020]

My Interests

Scene understanding for self-driving

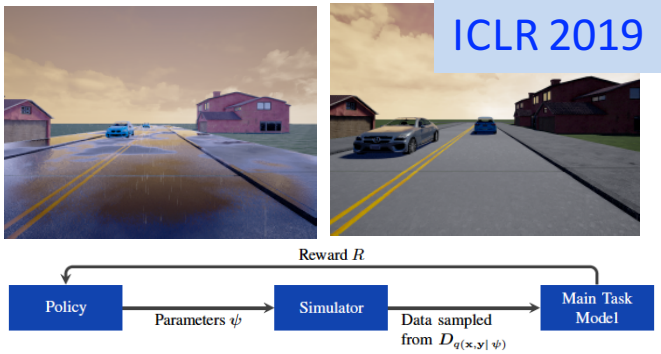
Distillation networks for fast and accurate object detection



NeurIPS 2017

Distillation for compressed CNN (student) to mimic uncompressed CNN (teacher), to achieve greater accuracy at the same speed.

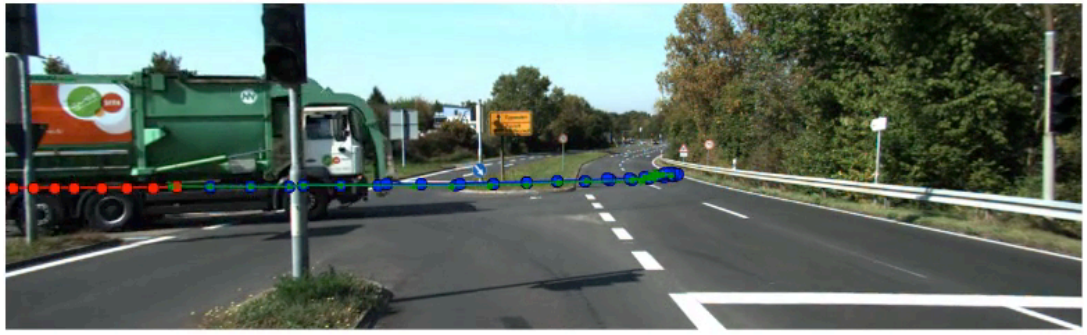
Learning to simulate



ICLR 2019

Reinforcement learning for simulations. Ensuring diversity and coverage.

Learning to predict uncertain future behavior

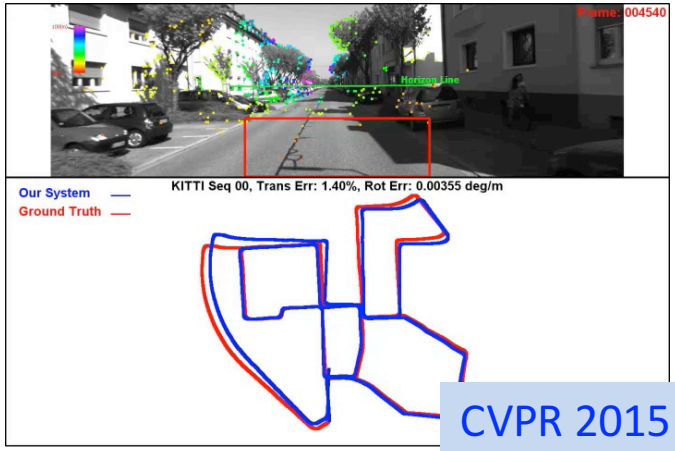


DESIRE: Deep Stochastic IOC RNN Encoder-Decoder

CVPR 2017

- Deep CVAE (autoencoder) to generate **diverse** hypotheses.
- RNN to rank predictions based on **motion**, **scene** and **interactions**.
- Deep inverse reinforcement learning for **long-term future rewards**.

Monocular SFM

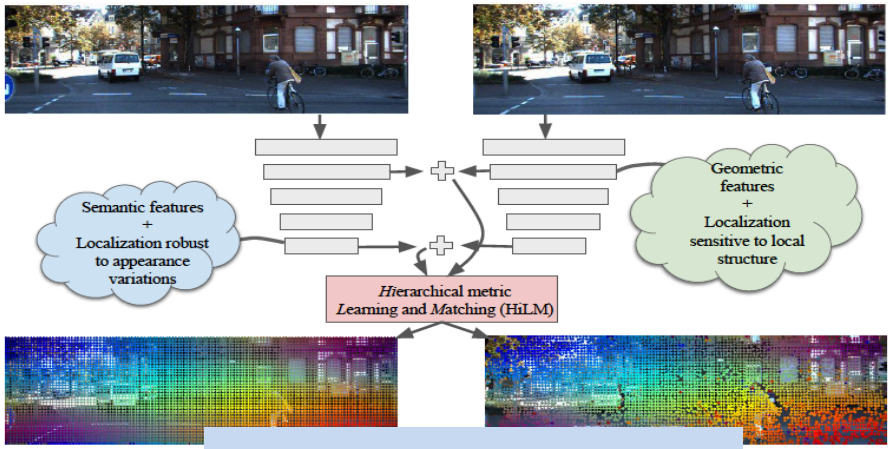


CVPR 2015

Large-scale, real-time, monocular SFM. Accuracy comparable to stereo systems.

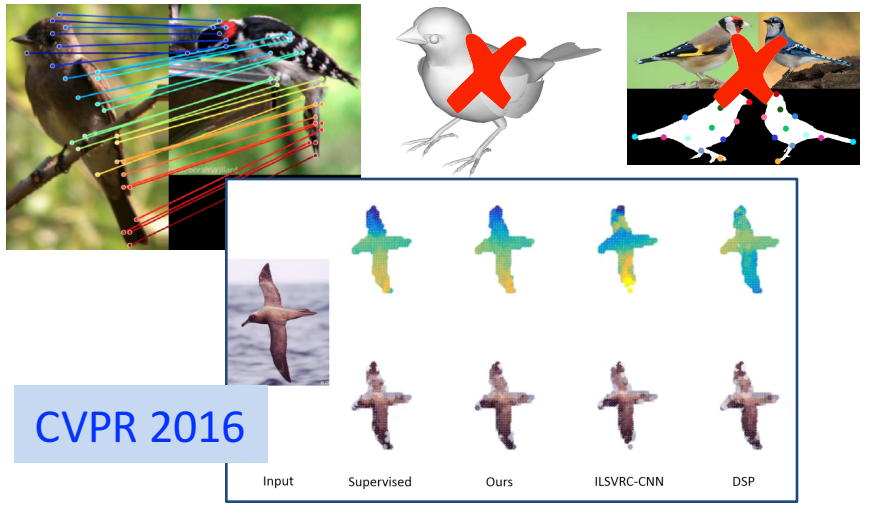
Geometric and semantic 3D reconstruction

Metric learning for correspondence



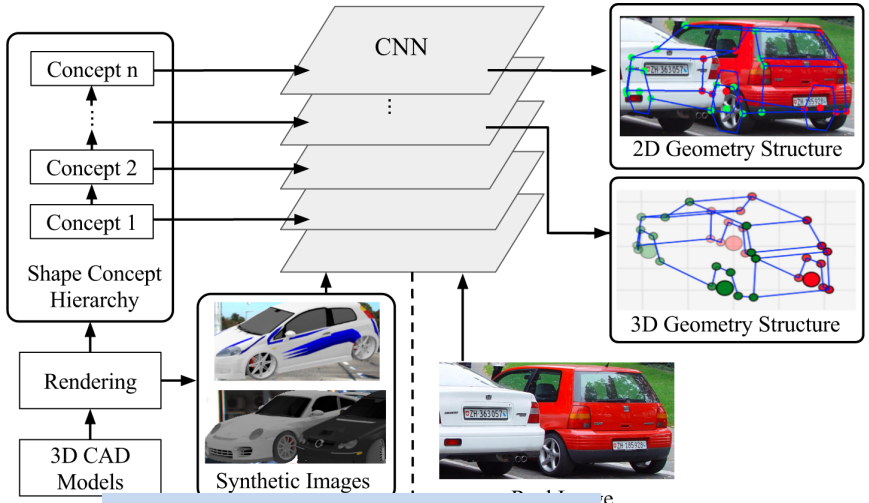
NeurIPS 2016, ECCV 2018

Weakly supervised semantic reconstruction



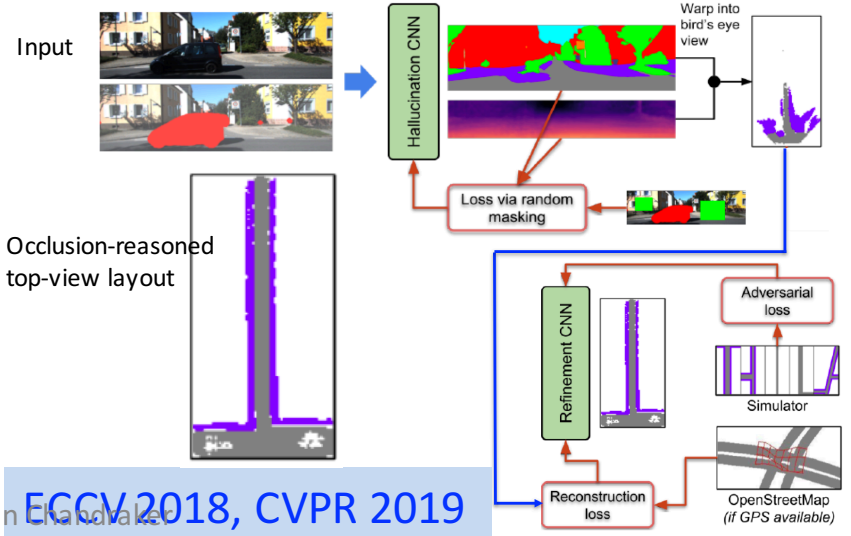
CVPR 2016

Deep supervision for occlusion-reasoned parts



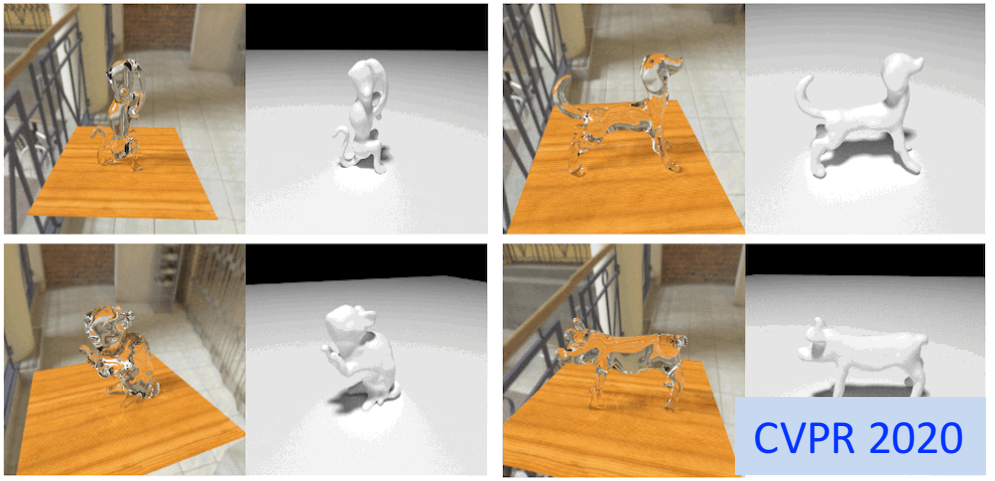
CVPR 2017, PAMI 2018

Occlusion reasoning and large transformations



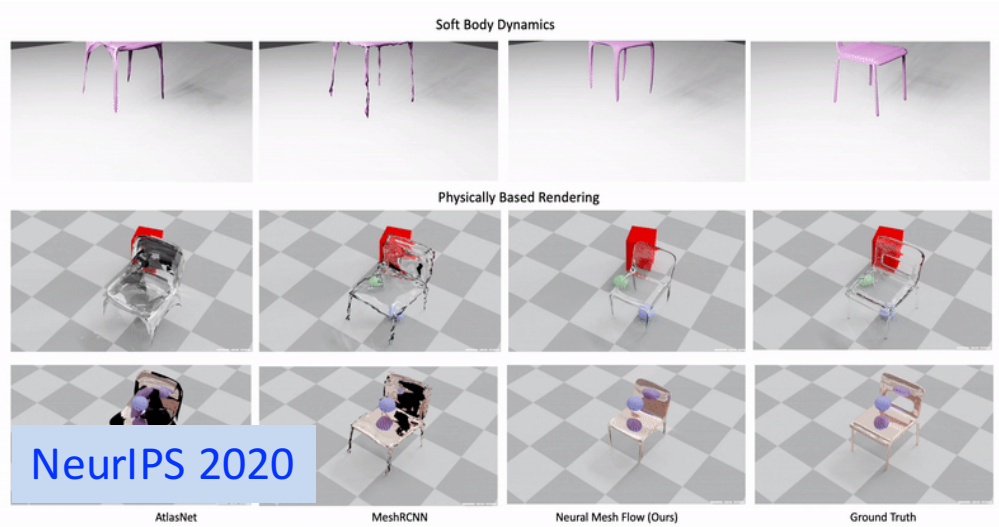
ECCV 2018, CVPR 2019

Physically-based learning for shape and material



Complex refractions and reflections

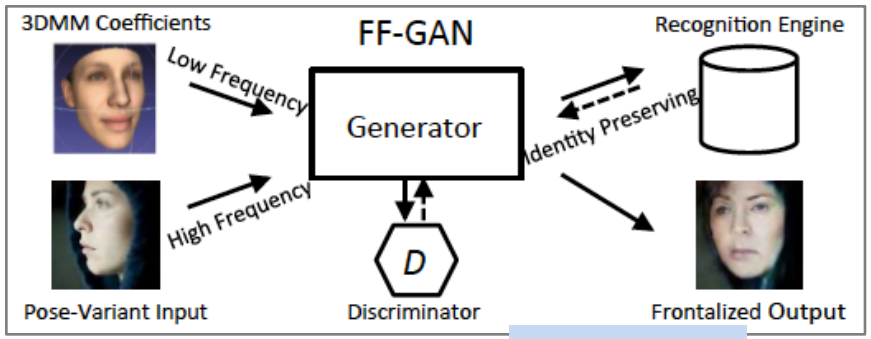
Complex materials and lighting

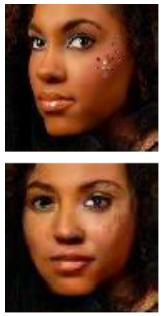


Physically realizable representations

Unsupervised adaptation to new domains

Face recognition for profile inputs



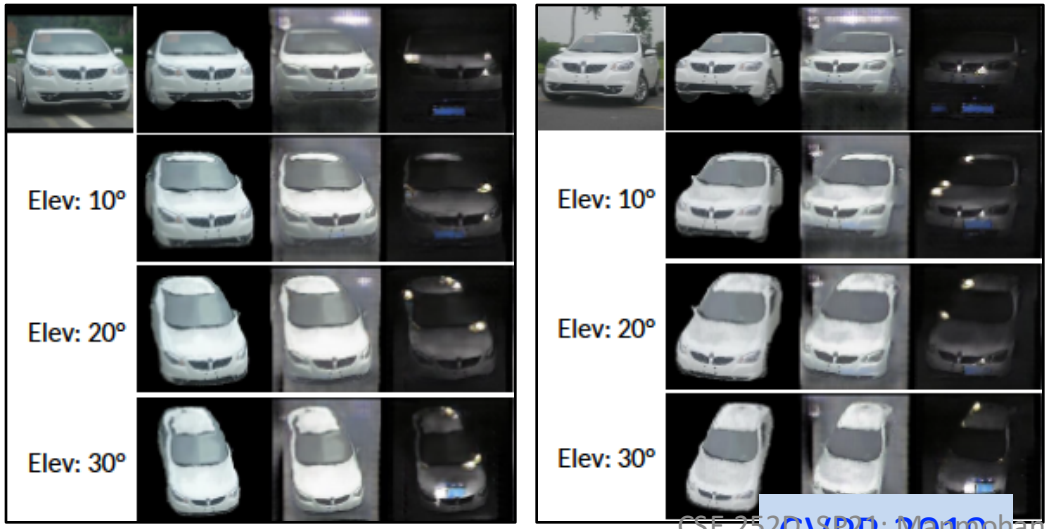


ICCV 2017

Segmentation across geographies

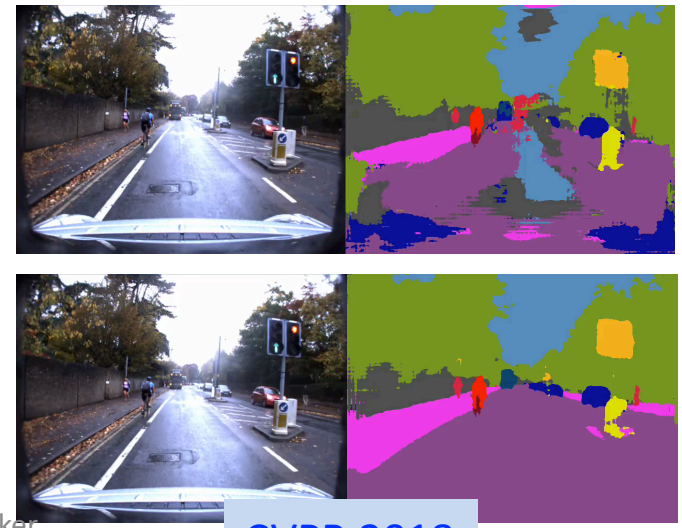
ICCV 2019

Car recognition across camera and lighting conditions



CVPR 2019

From rainy to good weather



CVPR 2018

Overall goals for the course

- Introduce the state-of-the-art in computer vision
- Study principles that make them possible
- Get understanding of tools that drive computer vision
- Enable one or all of several such outcomes
  - Pursue higher studies in computer vision
  - Join industry to do cutting-edge work in computer vision
  - Gain appreciation of modern computer vision technologies

Course Details

Course details

- Each class will cover topics in computer vision
- Examples of topics
    - Correspondence
    - Optical flow
    - Structure from motion
    - Face recognition
    - Human pose estimation
    - Material and lighting
    - Semantic segmentation
    - Object detection
    - Action recognition
    - Domain adaptation

Course details

- Topics structured into a few modules
  - Background
  - Structure and Motion
  - Faces and Humans
  - Objects and Stuff
  - Material and Lighting
  - Bias and Adaptation

Course details

- "Lightning" presentations
  - Provide a broad view of the field
  - An important skill to digest and present literature
  - Four students to present in one class
  - Papers to be assigned by instructor
  - Order of presentation: alphabetic (Googledoc will be posted)
- Send recorded presentation video 3 days before class
  - Will share PPT template
  - Well-practiced and fluent presentation
  - Incorporate feedback from instructor or TA
  - Include question to class
  - Ask and answer questions after presentation

Course details

- Presentation format:
    1. Motivation and problem description
    2. Prior work
    3. Method overview
    4. Method analysis
    5. Experiments
    6. Future work and discussion

Course details

- Class webpage:
  - http://cseweb.ucsd.edu/~mkchandraker/classes/CSE252D/Spring2021/
- Instructor email: mkchandraker@eng.ucsd.edu
- TA: Yu-Ying Yeh (Email: yuyeh@eng.ucsd.edu)
- Grading
  - 10% presentation
  - 60% assignments
  - 30% final exam
  - Ungraded quizzes
- Aim is to learn together, discuss and have fun!

Take-home message

- Computer vision is a key branch of AI
- Enables several modern applications around us
- A lot of highly visible and high-impact activity
- Huge industry interest
- This is a great time to study computer vision!