Computer Vision 3: Detection, Segmentation and Tracking

The Team

Lecturers

Prof. Dr. Laura Leal-Taixé

Dr. Aljosa Osep

What this course is:

- A course on Computer Vision
  - Object detection
  - Instance and semantic segmentation
  - Multiple object tracking in 2D and 3D

- Other CV courses:
  - Computer Vision 2: Multiple View Geometry (WS)

What this course is NOT:

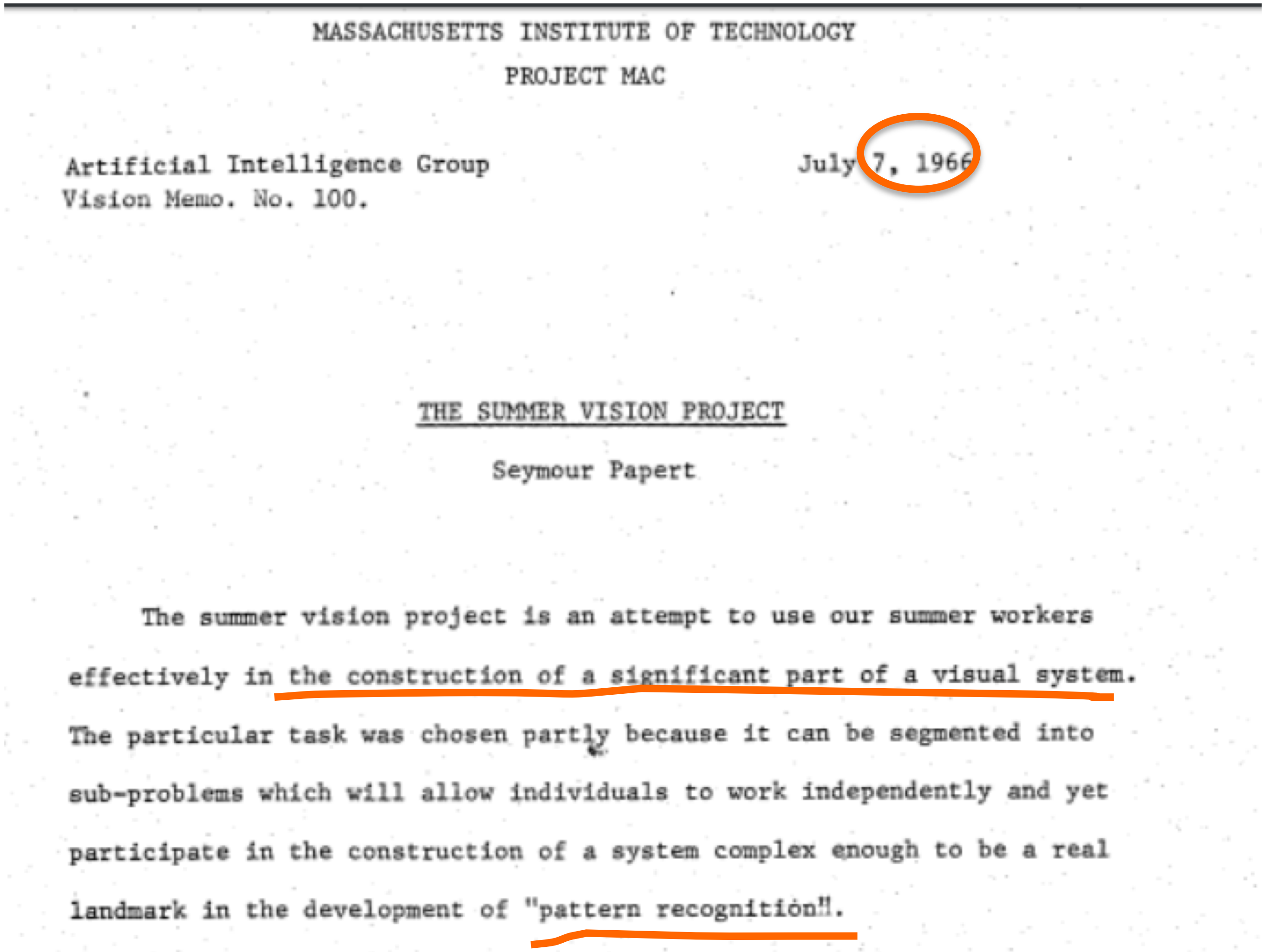
- An Introduction to Deep Learning
  - Take "Introduction to Deep Learning" if you are not familiar with basic DL concepts

- A practical project course
  - Take "Advanced Deep Learning for Computer Vision"

- A theoretical introduction into 3D Vision
  - Take "Computer Vision 2: Multiple View Geometry (WS)"

What is Computer Vision?

- First defined in the 60s in artificial intelligence groups
- “Mimic the human visual system”
- Center block of robotic intelligence



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

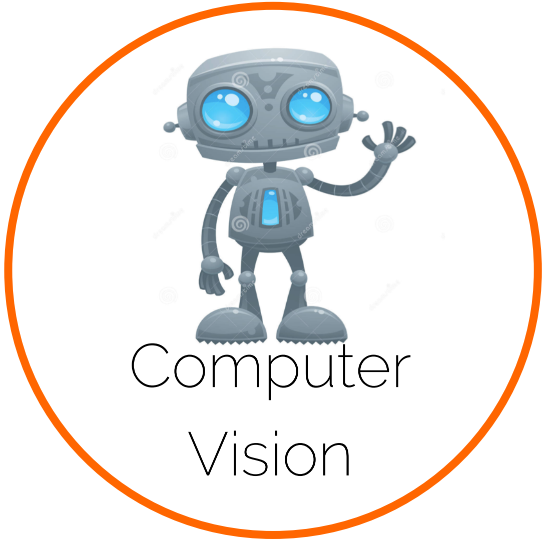
July 7, 1966

THE SUMMER VISION PROJECT

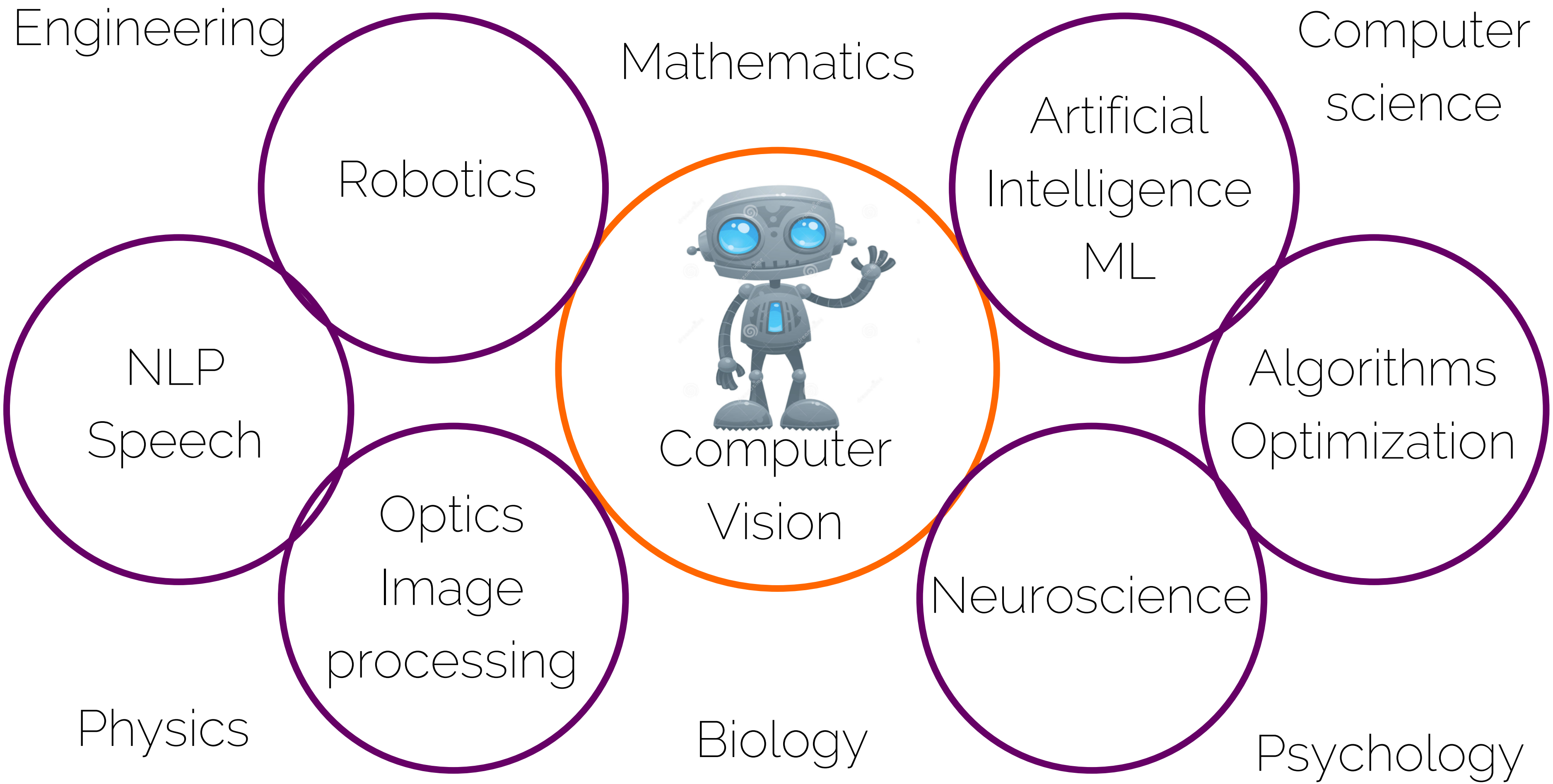
Seymour Papert

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".

Some decades later…



Computer Vision



Engineering

Mathematics

Computer science

Robotics

Artificial Intelligence ML

NLP Speech

Computer Vision

Algorithms Optimization

Optics Image processing

Neuroscience

Physics

Biology

Psychology

Computer Vision

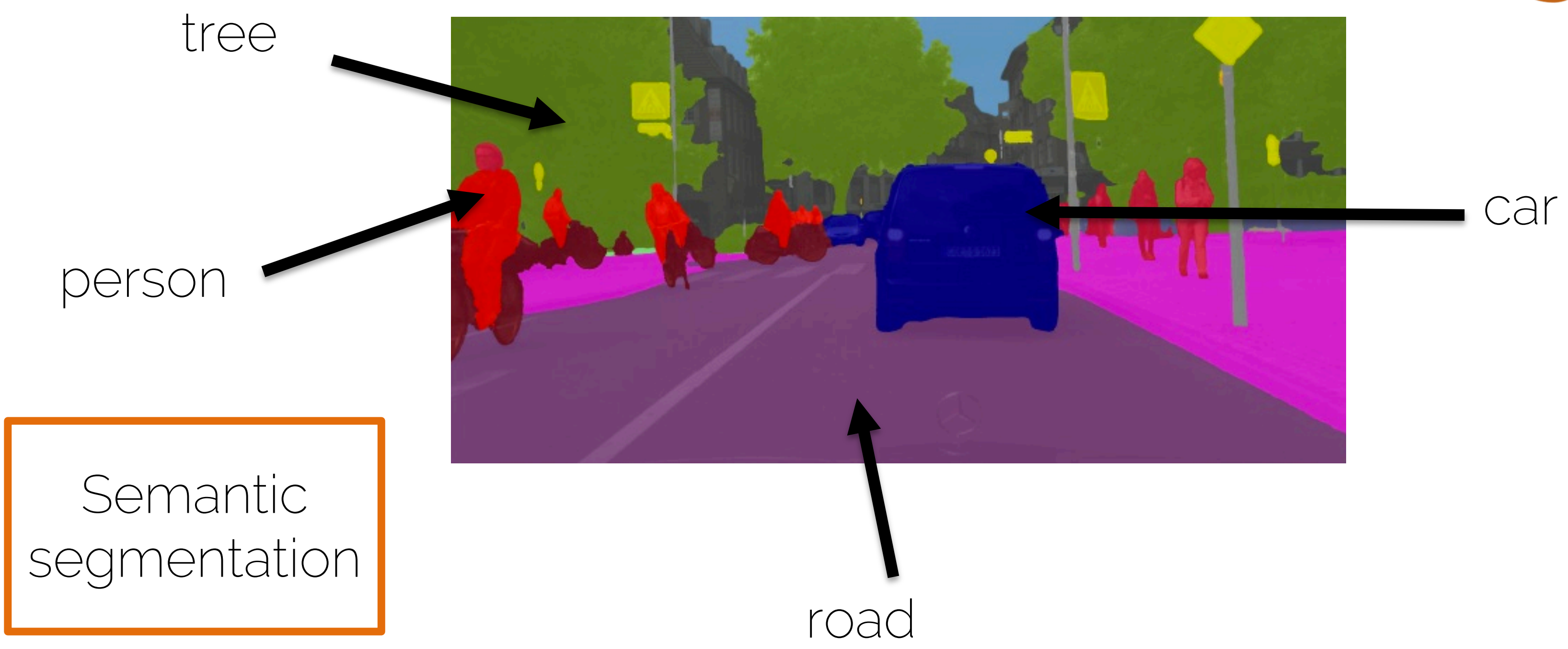
Give eyes to a computer

Computer Vision

Understand every pixel of an image

Computer Vision

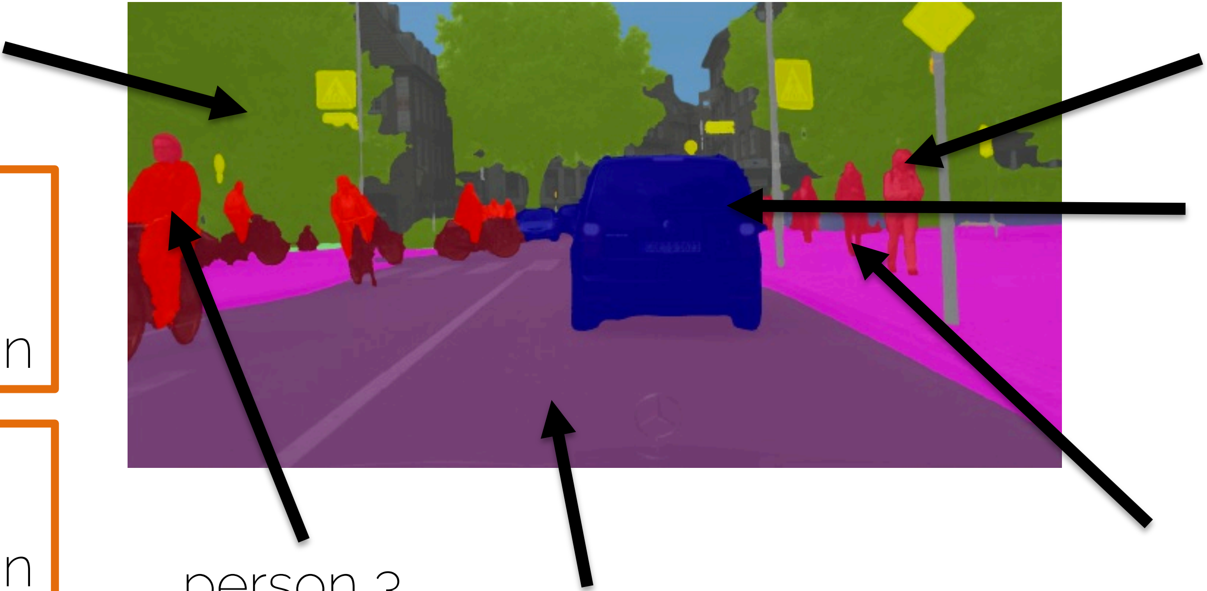
Understand every pixel of an image



Semantic segmentation

Computer Vision

Understand every pixel of an image



tree

person 2

car

person 3

road

person 1

Instance-based segmentation

Semantic segmentation

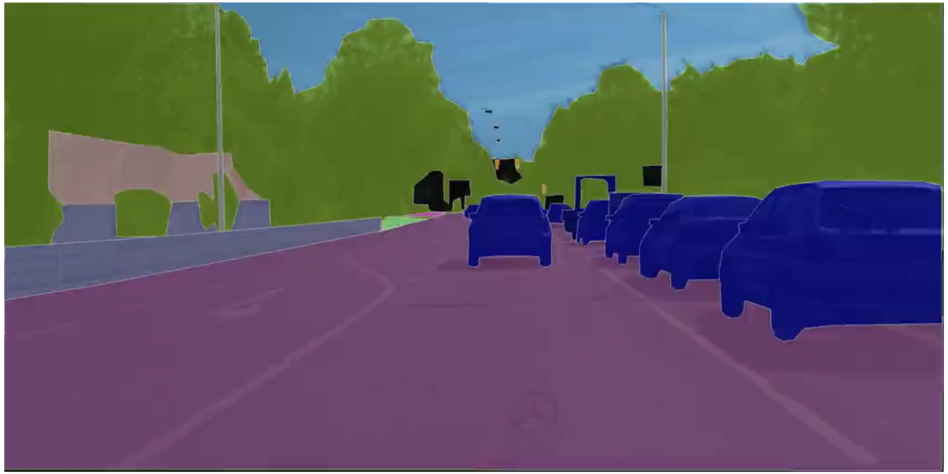
Computer Vision

Understand every pixel of a video

Multiple object tracking

Instance-based segmentation

Semantic segmentation



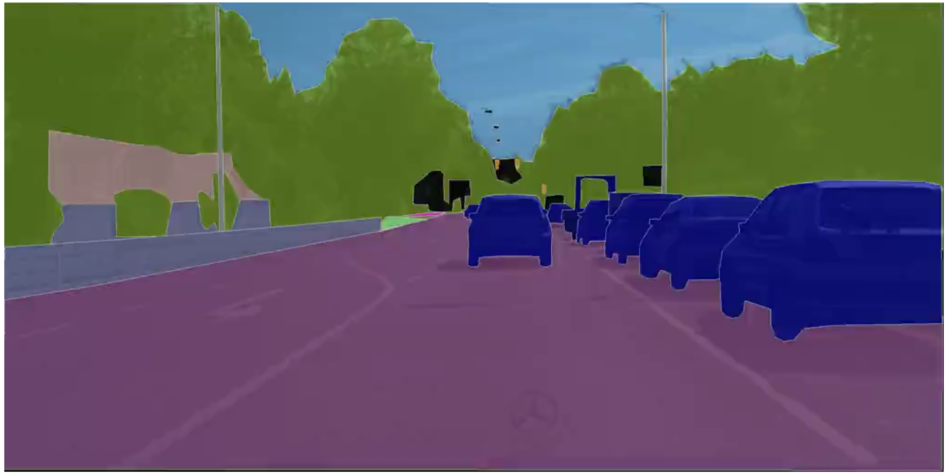
Dynamic Scene Understanding

Understand every pixel of a video

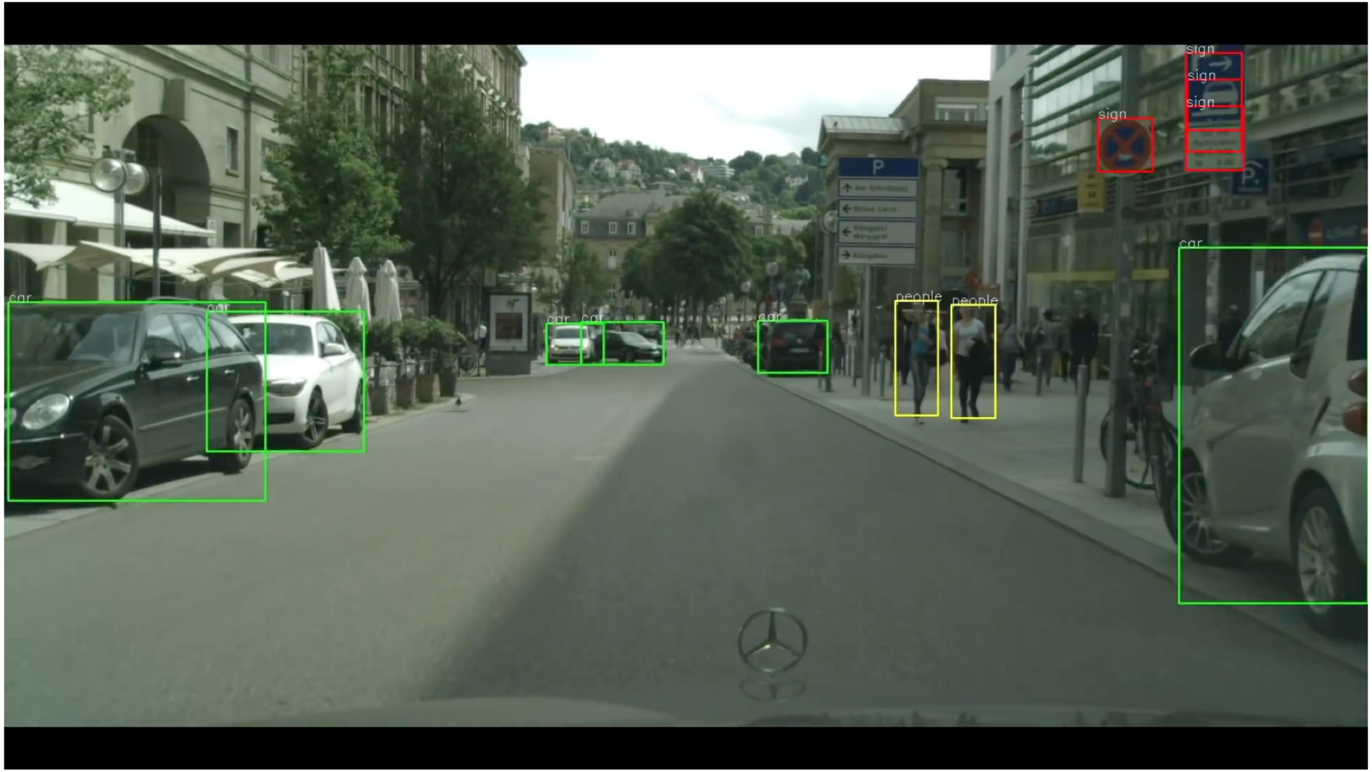
Multiple object tracking

Instance-based segmentation

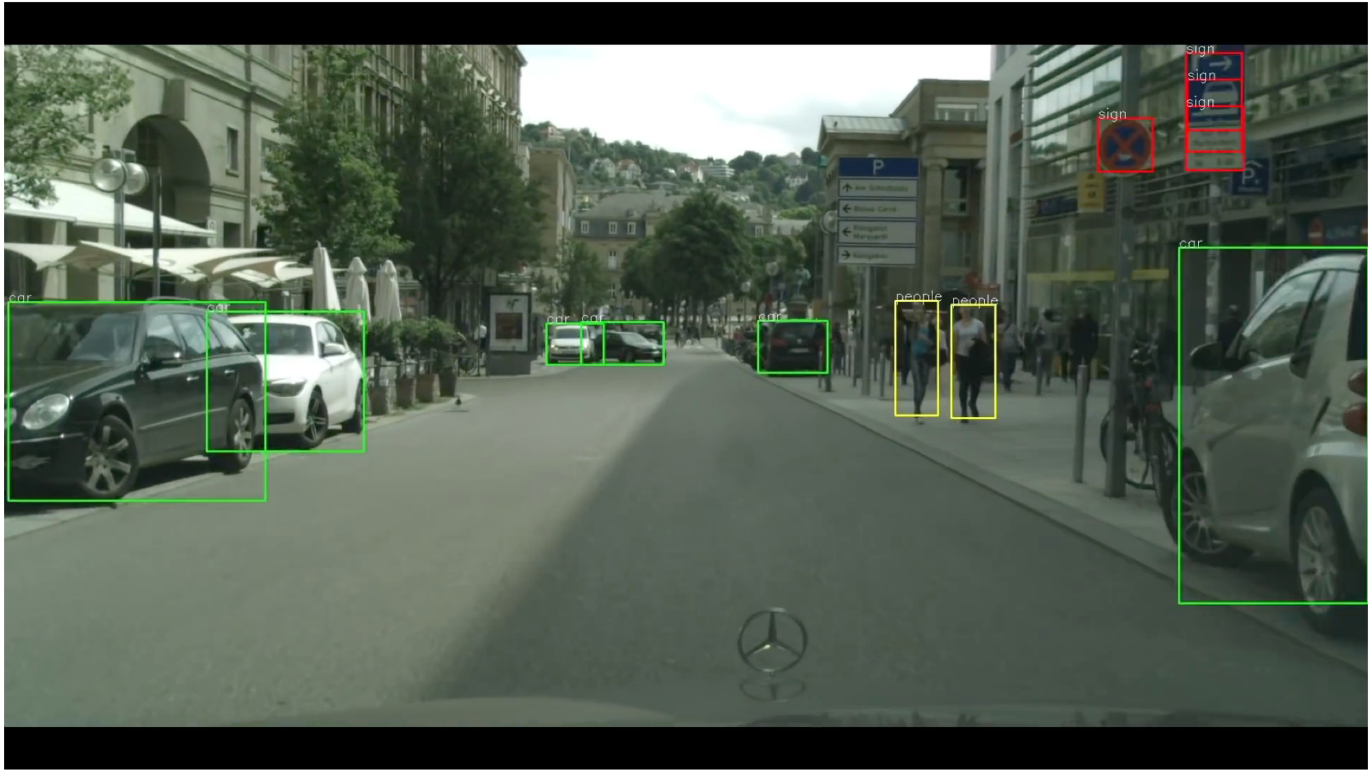
Semantic segmentation



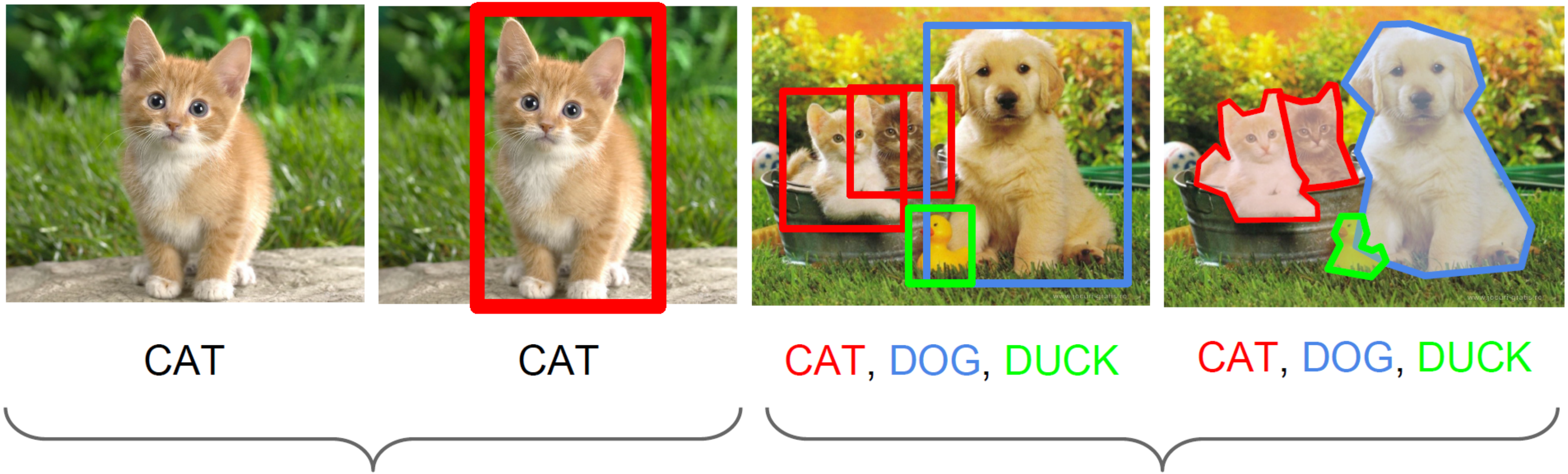
Autonomous driving



Autonomous driving



Understanding an image



Classification

Classification + Localization

Object Detection

Instance Segmentation

CAT

CAT

CAT, DOG, DUCK

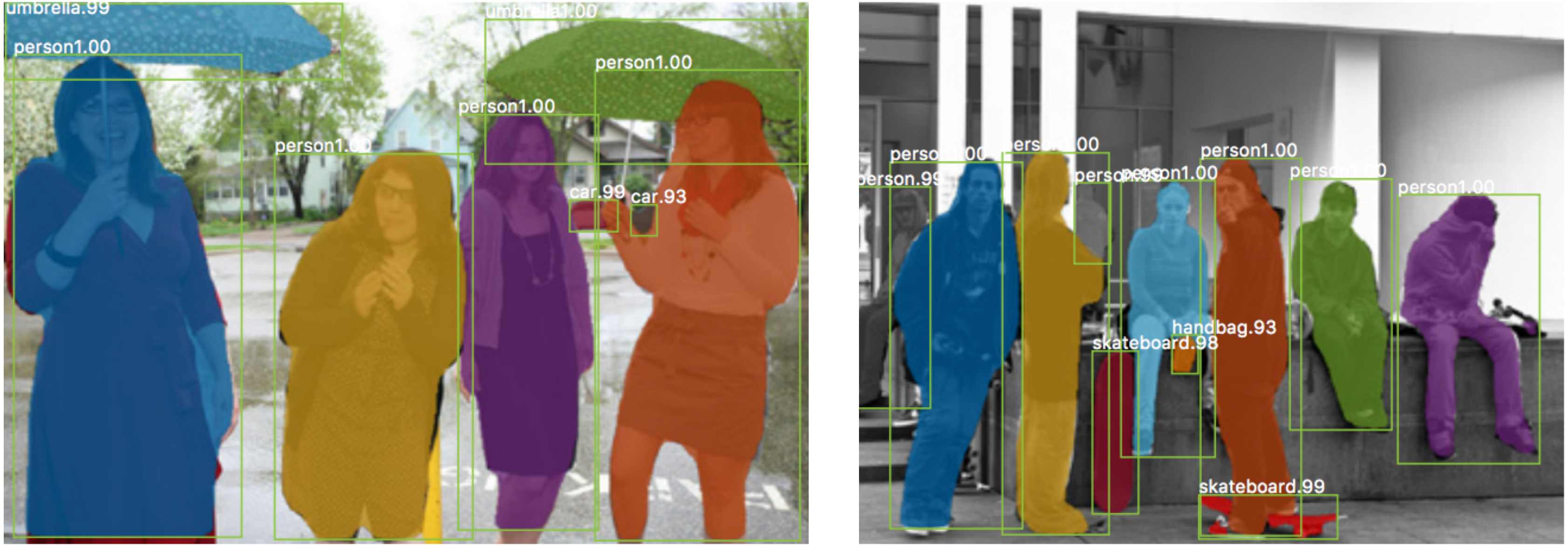
CAT, DOG, DUCK

Single object

Multiple objects

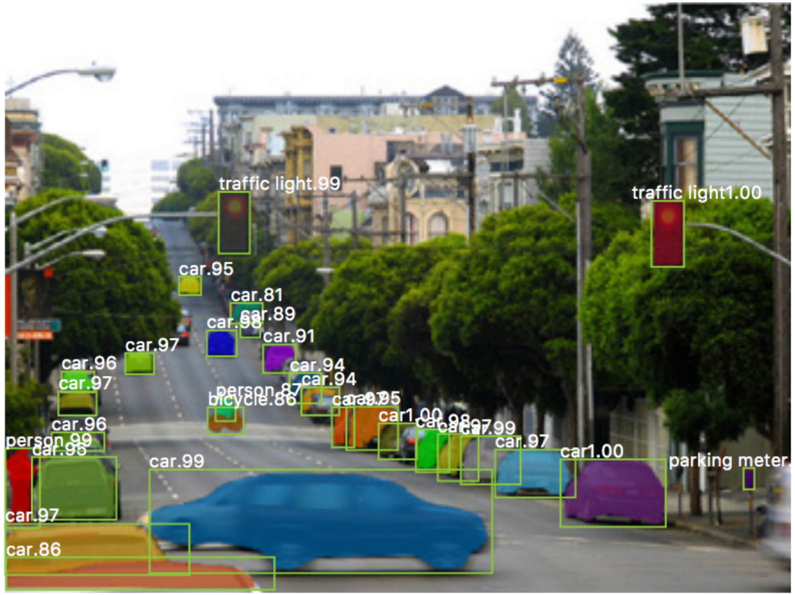
Credit: Li/Karpathy/Johnson

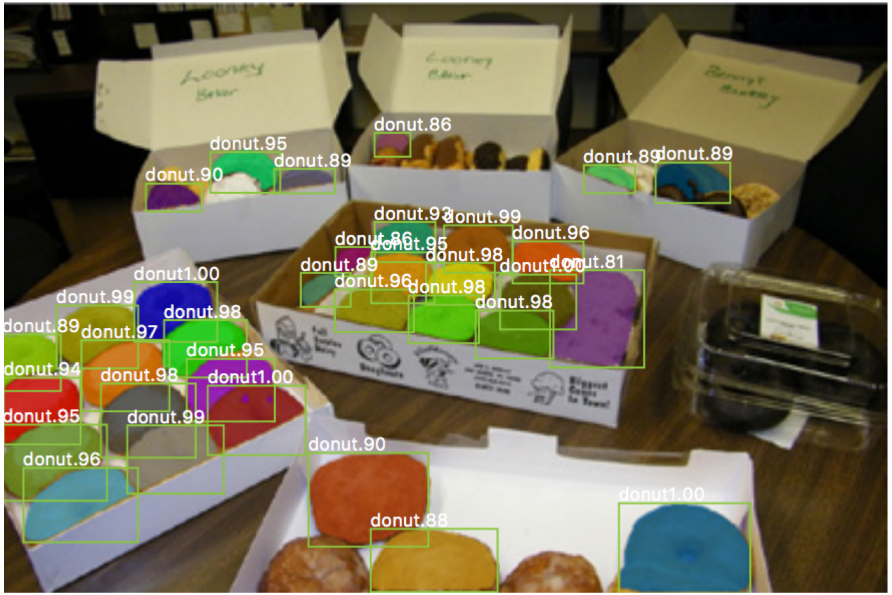
Understanding an image



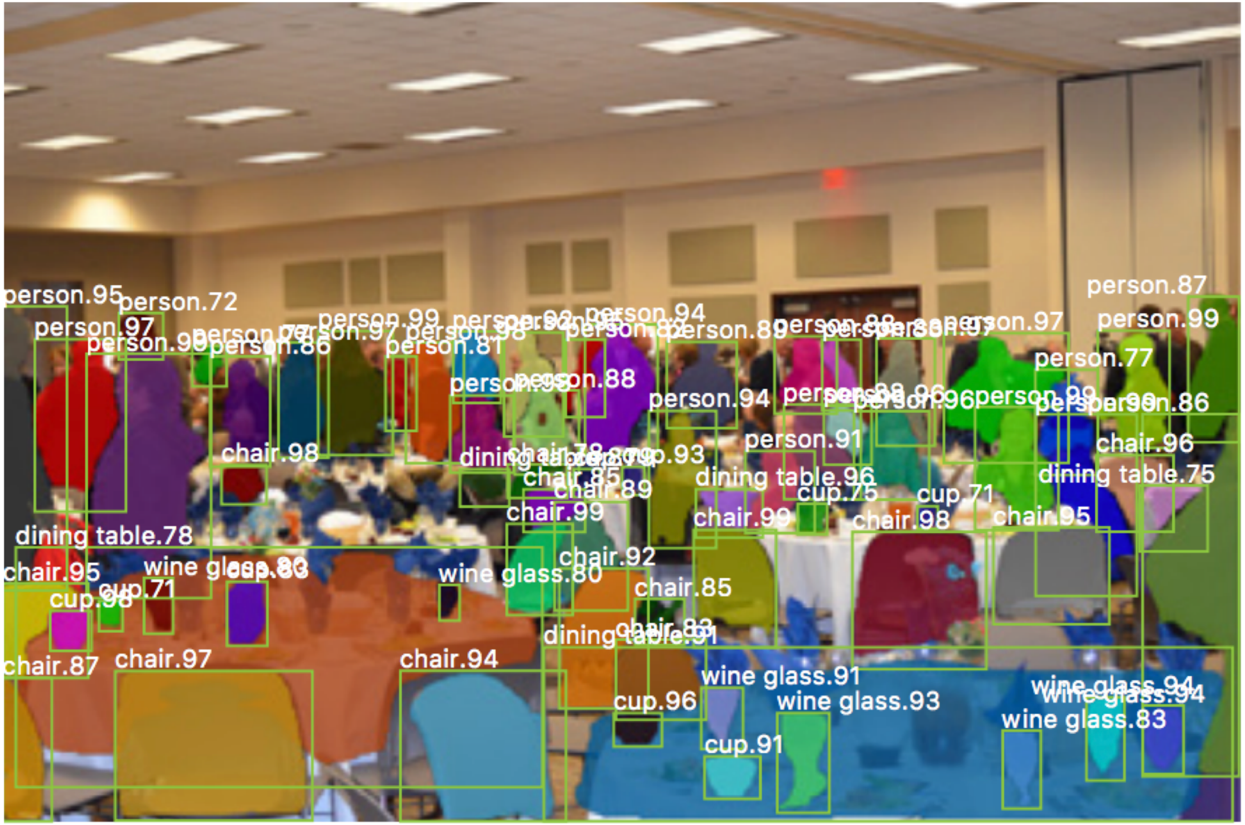
K. He, G. Gkioxari, P. Dollar, R. Girshick. Mask R-CNN. ICCV 2017.

Understanding an image

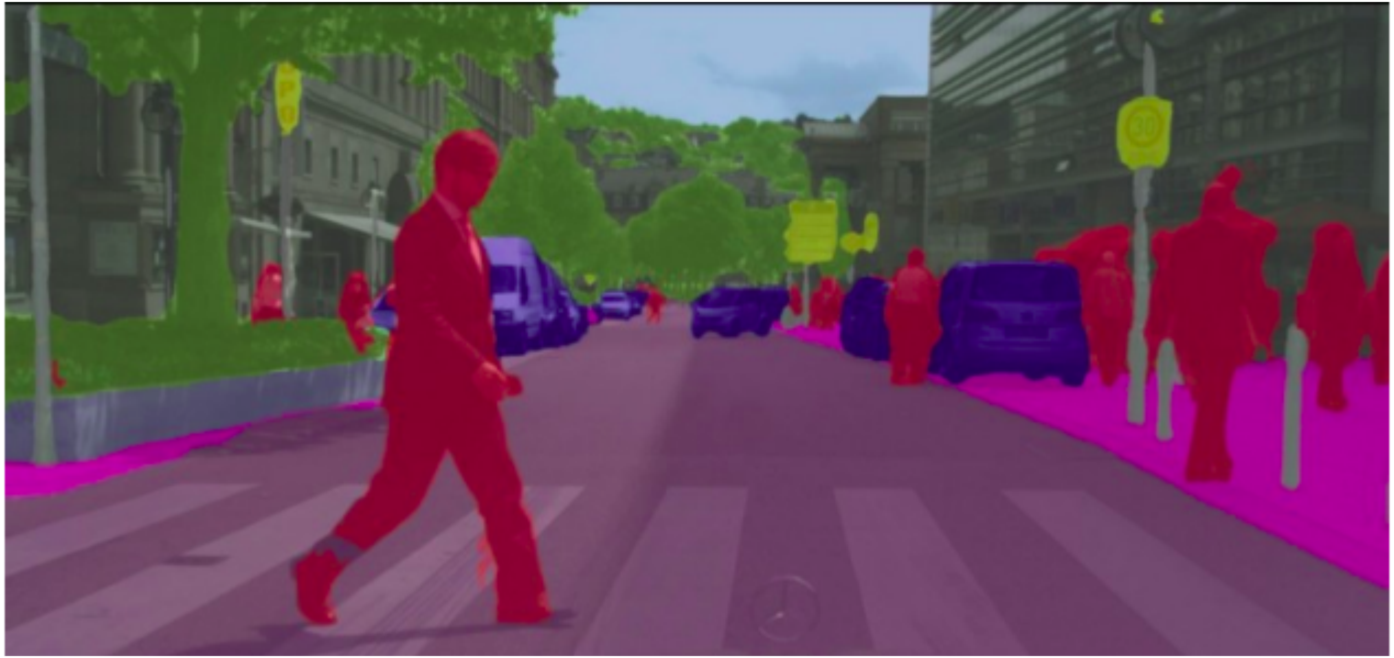




Understanding an image



Understanding an image



Understanding an image

- Different representations depending on the granularity
  - Detections (coarse)
  - Segmentations (precise)
  - Semantic with/without instances (person 1, person 2)

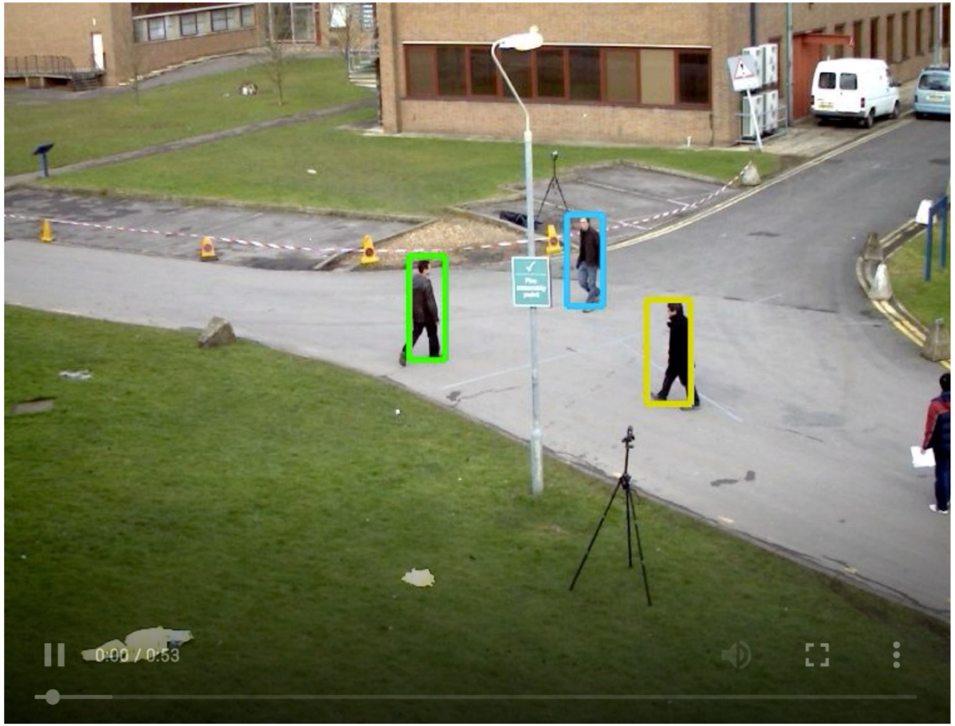
- Goes well with Deep Learning

Understanding an video

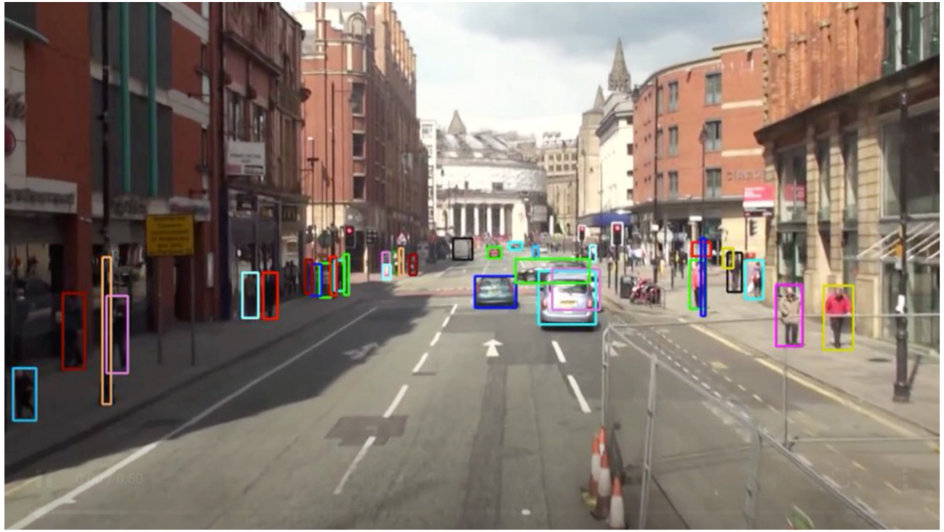
- Temporal domain which brings us advantages
  - A lot of redundancy
  - A smoothness assumption: things do not change much from one frame to another

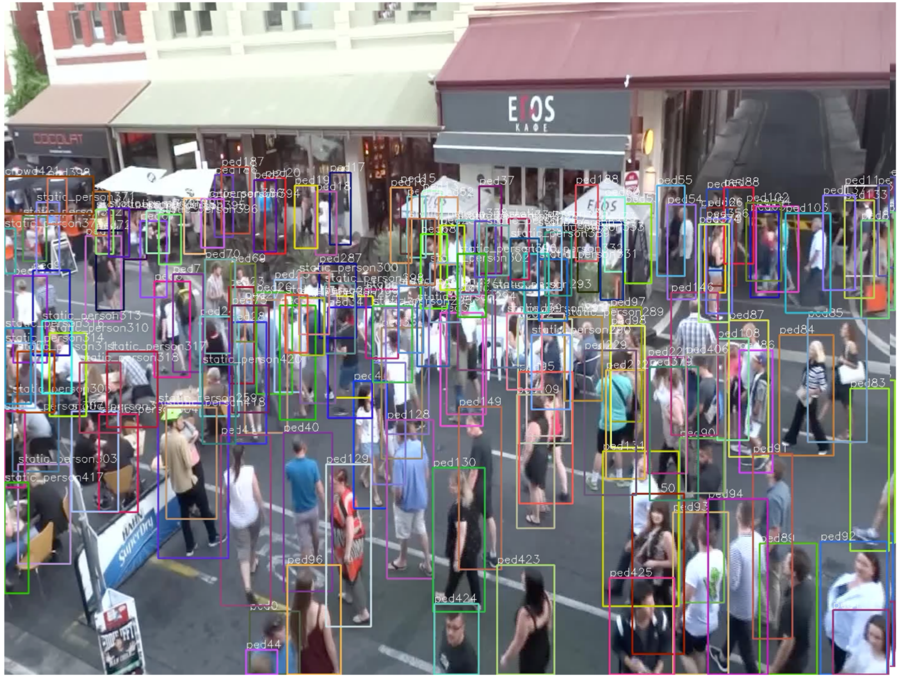
- … but also disadvantages
  - At 30 FPS, image the computation one has to do to process a video….
  - Occlusions, multiple objects moving and interacting…

Understanding an video: then



Understanding an video: now





Understanding an video

- Where is every object going?
- How are objects interacting?
- Get consistent results in the temporal dimension

Rough schedule/content

- 18.10: Introduction
- 25.10: No lecture
- 1.11: Holidays
- 08.11: Detection 1
- 15.11: No lecture
- 22.11: No lecture
- 29.11: Detection 2
- 06.12: Single/Multiple object tracking
- 13:12: Multiple object tracking
- 20.12: tbd
- Christmas Break
- 10.01: Trajectory prediction
- 17.01: Video object segmentation
- 24.01: : Semantic/Instance Segmentation 1
- 31.01: Semantic/Instance Segmentation 2
- 07.02: Going towards 3D tracking and segmentation

Rough schedule/content

- RCNN, Fast RCNN and Faster RCNN
- YOLO, SSD, RetinaNet
- Siamese networks – Person Re-Identification
- Message Passing Networks
- Network (non-neural) flow for tracking
- Generative Adversarial Networks – trajectory prediction
- Mask-RCNN, UPSNet (panoptic segmentation)
- Deformable/atrous convolutions
- 3D – data, algorithms.

Our Research Lab

Dynamic Vision and Learning Group

https://dvl.in.tum.de/

Course organization

About the lecture

- Theory: 10-11 lectures
    - Every Friday 16:00-18:00 (MI HS 2)
    - Exceptions will be communicated.
    - Practice: no physical sessions
- Lecture will NOT be recorded

https://dvl.in.tum.de/teaching/cv3dst-ws1920/

Grading system

- Exam: 11th February, 16:00-17:30
- No retake exam as the course will be moved to the Summer Semester

- Completing the practical part successfully gives a bonus of 0.3

Moodle

- Announcements via Moodle - IMPORTANT!
  - Sign up in TUM online for access: https://www.moodle.tum.de/
  - We will share common information (e.g., regarding exam)
  - Ask content questions online so others benefit
  - Don't post solutions

Emails & Slides

- All material will be uploaded on Moodle and the web
- Questions regarding the syllabus, exercises or contents of the lecture, use Moodle!
- Questions regarding organization of the course:

dst@dvl.in.tum.de

- Emails to the individual addresses will not be answered.

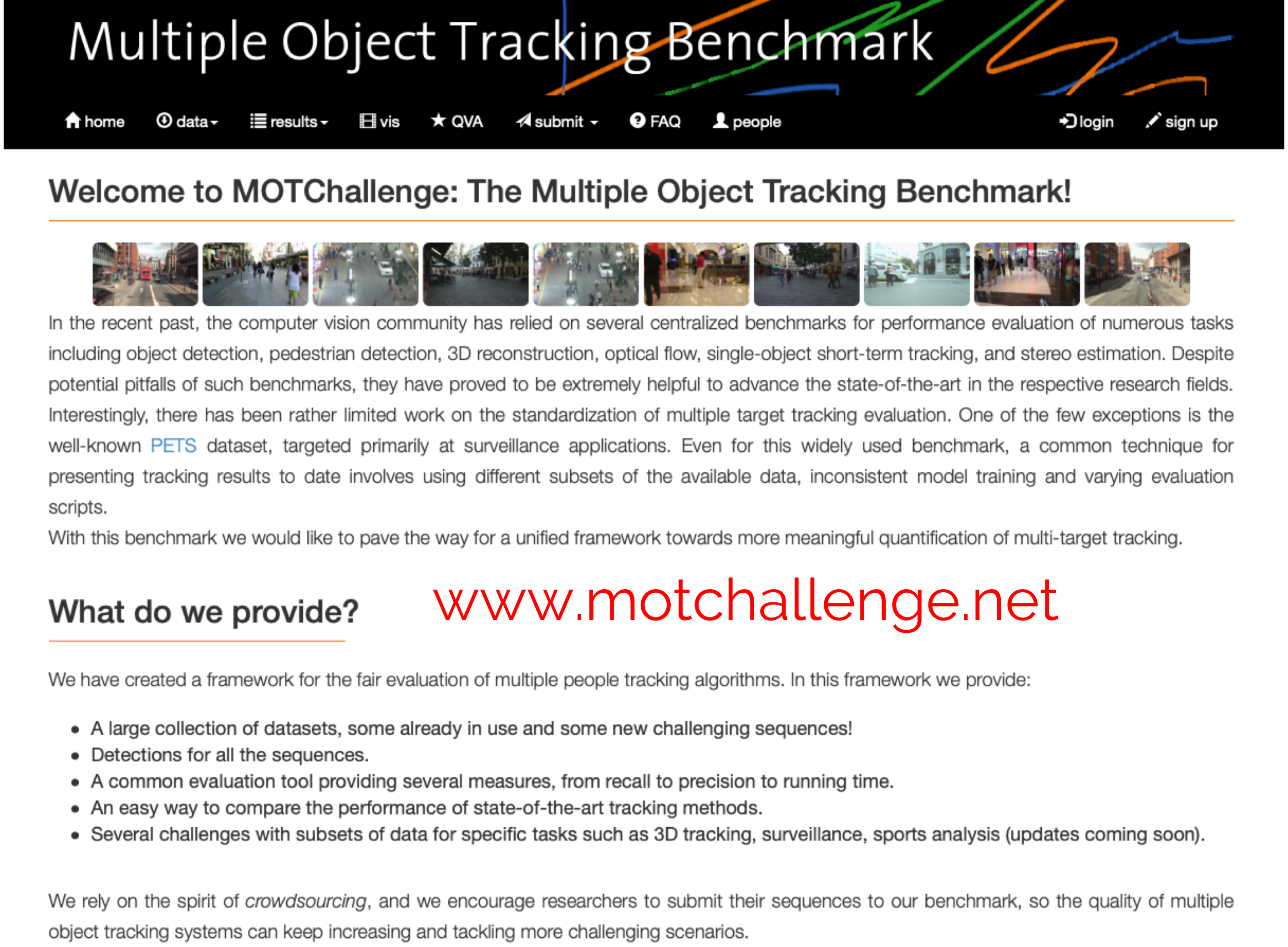
Practical part

- Internal Kaggle competition
- We will have a detection and/or tracking challenge.
- You will all start from the same point (code)

- After that, it will be an open competition.

Practical part: rules

- Individual competition, but you can share tips and tricks
- Note: the grade will be based on the performance of the whole class, so do not share too many tips ☺
- You can copy pieces of code from online resources, but you have to understand everything that is in your code and be able to explain how it works.
- There will be presentations.
- We will not provide help from the coding side.
- All questions will go through Moodle (no office hours).



Welcome to MOTChallenge: The Multiple Object Tracking Benchmark!

In the recent past, the computer vision community has relied on several centralized benchmarks for performance evaluation of numerous tasks including object detection, pedestrian detection, 3D reconstruction, optical flow, single-object short-term tracking, and stereo estimation. Despite potential pitfalls of such benchmarks, they have proved to be extremely helpful to advance the state-of-the-art in the respective research fields. Interestingly, there has been rather limited work on the standardization of multiple target tracking evaluation. One of the few exceptions is the well-known PETS dataset, targeted primarily at surveillance applications. Even for this widely used benchmark, a common technique for presenting tracking results to date involves using different subsets of the available data, inconsistent model training and varying evaluation scripts.

With this benchmark we would like to pave the way for a unified framework towards more meaningful quantification of multi-target tracking.

What do we provide?

www.motchallenge.net

We have created a framework for the fair evaluation of multiple people tracking algorithms. In this framework we provide:

- A large collection of datasets, some already in use and some new challenging sequences!
- Detections for all the sequences.
- A common evaluation tool providing several measures, from recall to precision to running time.
- An easy way to compare the performance of state-of-the-art tracking methods.
- Several challenges with subsets of data for specific tasks such as 3D tracking, surveillance, sports analysis (updates coming soon).

We rely on the spirit of *crowdsourcing*, and we encourage researchers to submit their sequences to our benchmark, so the quality of multiple object tracking systems can keep increasing and tackling more challenging scenarios.

Practical part

- The idea is to implement improvements to a baseline detector
- You can borrow ideas from the lectures!

- All training and testing will be done on the MOTChallenge benchmark.

https://motchallenge.net/data/MOT16/

Practical part

- The idea is to implement improvements to a baseline detector
- You can borrow ideas from the lectures!

- All training and testing will be done on the MOTChallenge benchmark.

https://motchallenge.net/data/MOT16/

Next lectures

- Detection – 8th November

https://dvl.in.tum.de/