

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



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".

Computer Vision

Give eyes to a computer



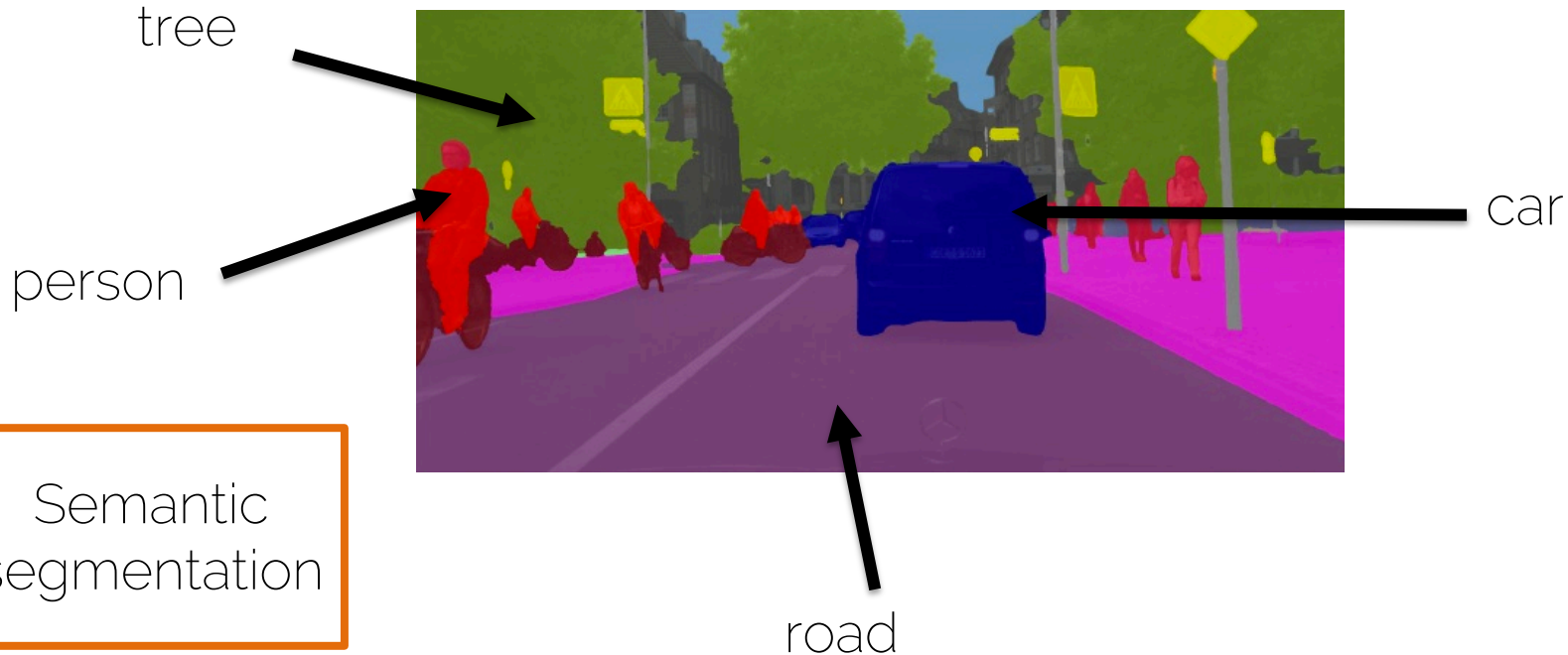
Computer Vision

Understand every pixel of an image



Computer Vision

Understand every pixel of an image



Computer Vision

Understand every pixel of an image



tree

person 2

Instance-
based
segmentation

car

Semantic
segmentation



person 3

road

person 1

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



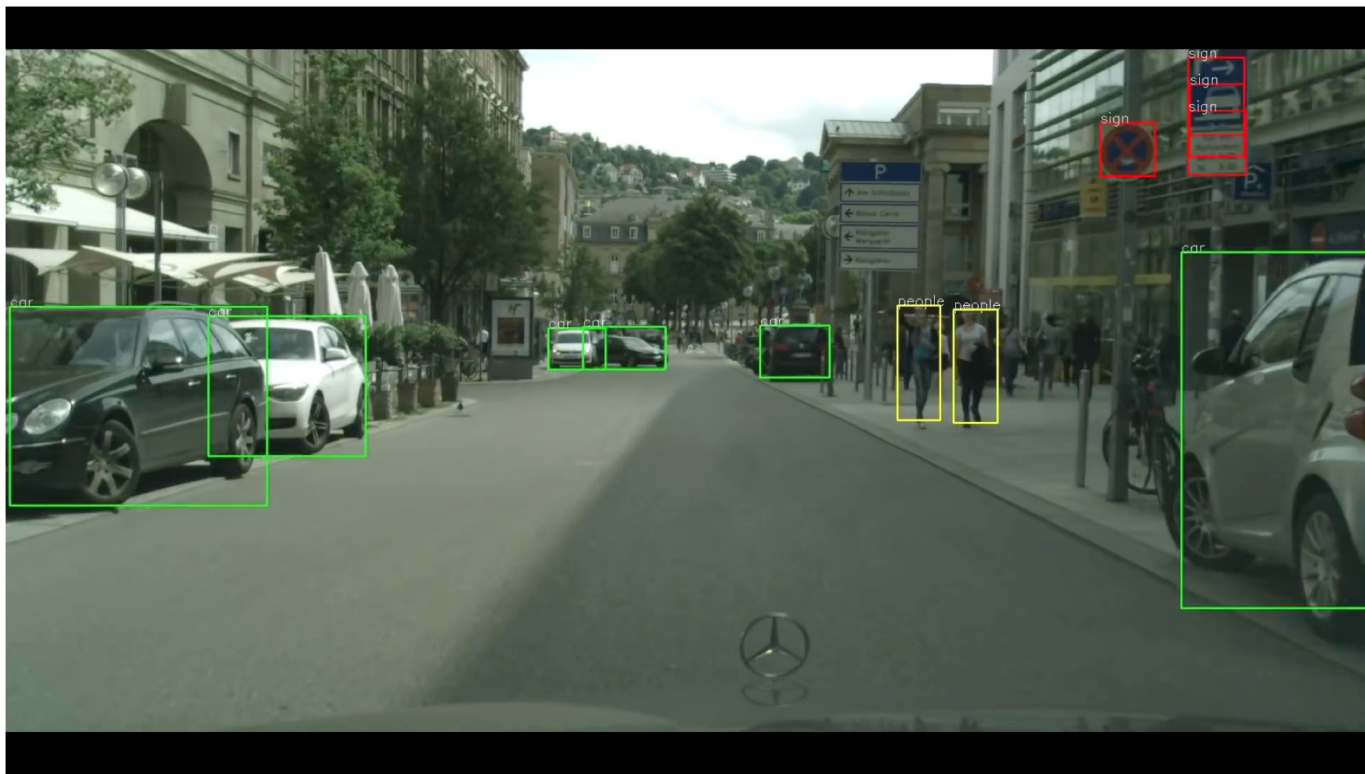
Multiple
object
tracking

Instance-
based
segmentation

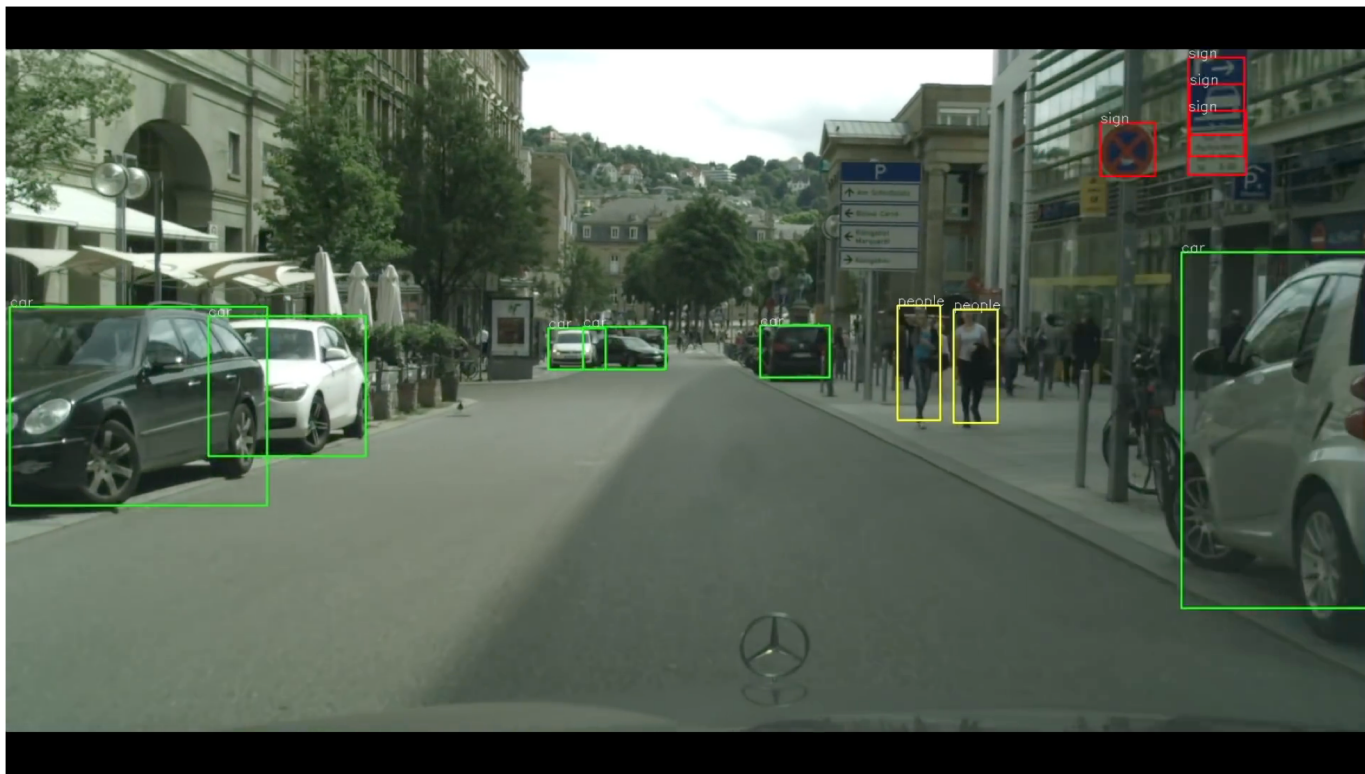
Semantic
segmentation



Autonomous driving



Autonomous driving



Understanding an image

Classification



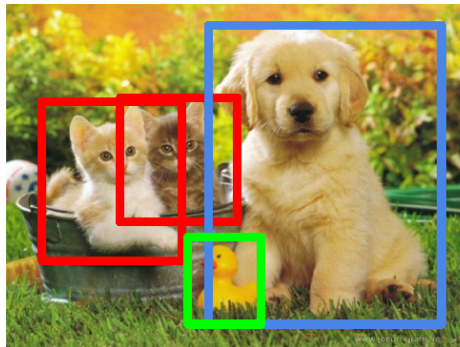
CAT

**Classification
+ Localization**



CAT

Object Detection



CAT, DOG, DUCK

**Instance
Segmentation**

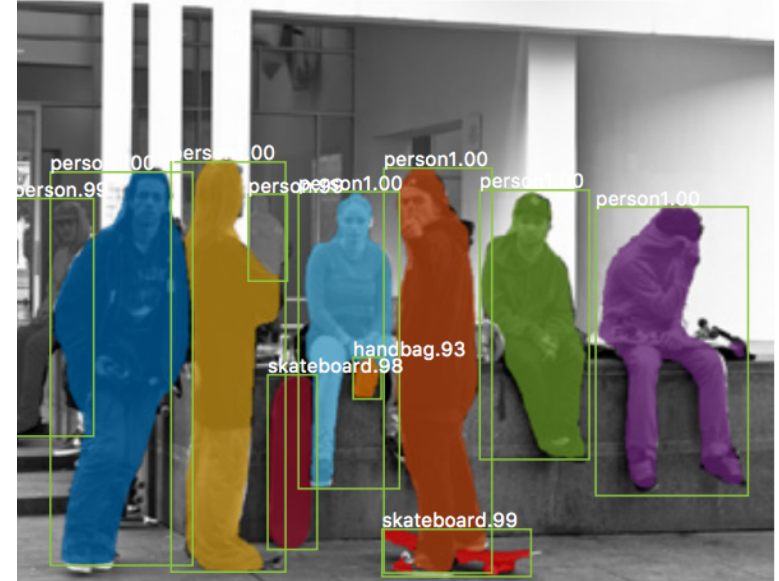
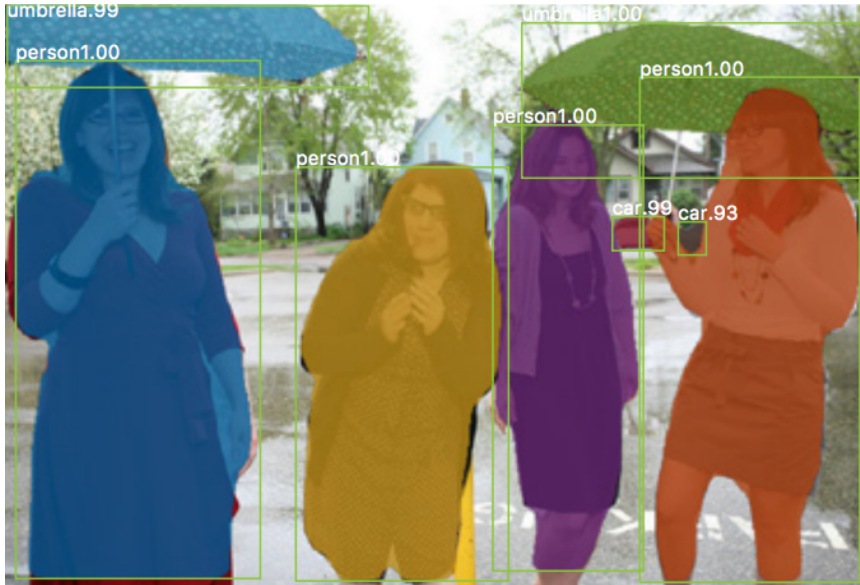


CAT, DOG, DUCK

Single object

Multiple objects

Understanding an image



Understanding an image



Understanding an image



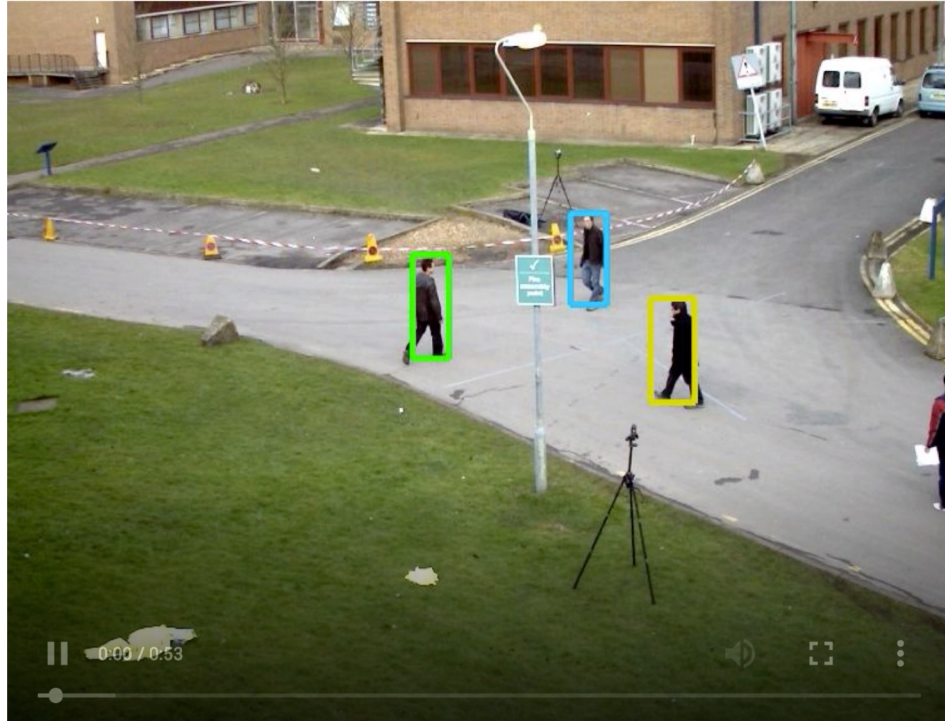
Understanding an image

- 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

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



Rough schedule/content

- 1. Introduction
- 2. Object Detection 1
- 3. Object Detection 2
- 4. Single/Multiple object tracking
- 5. Multiple object tracking
- 6. Trajectory prediction
- 7. Semantic segmentation
- 8. Instance Segmentation
- 9. Video object segmentation
- 10. 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/>

About the lecture

- Theory: 10-11 lectures
 - Every Wednesday ~~16:00-18:00 (MIHS 2)~~
- Lectures will be recorded this year!
 - Due to the virus situation, we will make all slides and video recordings available every Wednesday

<https://dvl.in.tum.de/teaching/cv3dst-ss20/>

Grading system

- Exam: tbd
- There will be a retake exam as this course will be moved permanently to Summer Semester
- Completing the practical part successfully gives a bonus of 0.3

Practical part

- Internal Kaggle competition
- We will have a tracking challenge.
- You will all start from the same point (code)
- After that, it will be an open competition.
- Check out the video of the presentation of the challenge!

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.

Computer Vision 3: Detection, Segmentation and Tracking