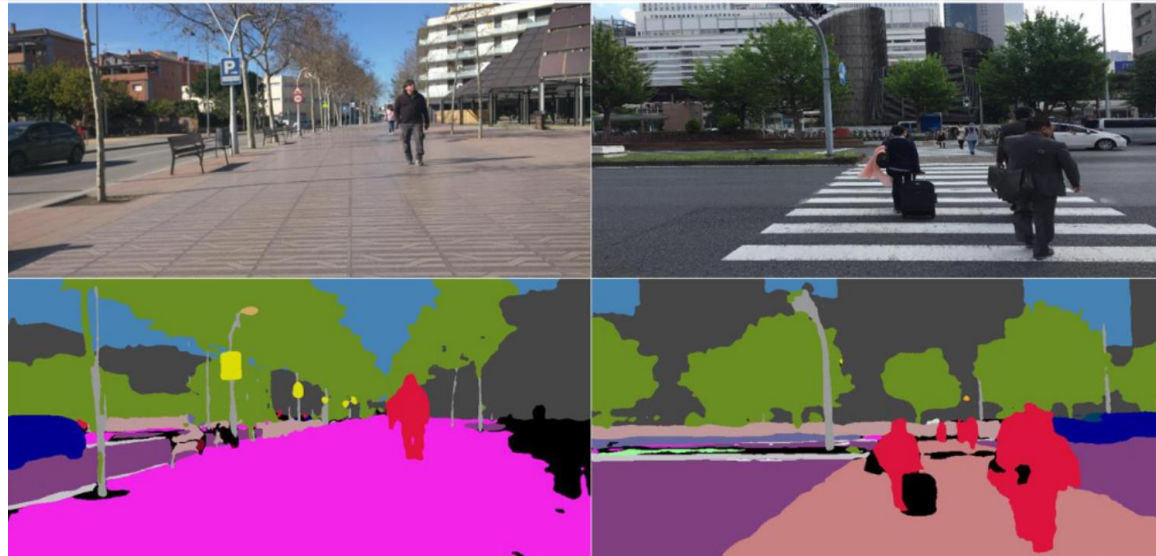# Practical Course:
Computer Vision for Human-Computer Interaction

WS 20/21
Constantin Seibold/Vivek Sharma

COMPUTER VISION FOR HUMAN COMPUTER INTERACTION LAB
INSTITUTE FOR ANTHROPOMATICS AND ROBOTICS

www.kit.edu

What will you learn?

- Apply algorithms from lectures and papers
- Hands on experience
- Get comfortable with machine learning tools
- Learn about common problems and applications in machine learning & vision
- Find solutions to difficult problems

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

General Information

Weekly meeting

- Compulsory attendance
- Talk about intermediate results & problems
- Ask for help and guidance
- Weekly goal: stay „on track"

3 Students per Team

- Use version control (e.g. git)
- Internal git repos provided via the SCC's GitLab (https://git.scc.kit.edu/)
- Divide work into separate tasks and distribute within group

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

At the end of the Practical Course…

- Final presentation of each group
  - 15 Minute talk
  - Each member talks about their contribution
  - The presentation should be about:
    - Goals and usefulness of your chosen topic
    - Your proposed approach
    - Results

- Written report describing the topic/approach/results
  - 4-pages in standard paper format
    - Abstract/Introduction/Method/Results/Conclusion
    - References can be fit on an extra page
  - Written in a conference paper template
    - http://cvpr2020.thecvf.com/sites/default/files/2019-09/cvpr2020AuthorKit.zip

- Final Code Submission
  - Working implementations of Algorithms
  - A Readme-file describing how the code can be used to reproduce the results
    - If the team agrees -> make code publicly available to the community

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Topics

- Human-Drone-Interaction

- Augmented Reality for Users with Low Vision

- Material Classification in Construction Sites

- Object Detection in Construction Sites

- Edge Detection in Construction Sites

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Human-Drone-Interaction

- Dynamically integrate drone usage into daily life

  - Assist for visually impaired

  - Drone-assisted filming

- Three Major Criteria:

  - Person registration and tracking
  - Drone positioning, movement and control
 - Drone-User Interaction
    - Gesture, Speech, … ?

Avila, Mauro, Markus Funk, and Niels Henze. "Dronenavigator: Using drones for navigating visually impaired persons." *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 2015.

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Human-Drone-Interaction

- Potential datasets
  - Person Representations
    - Market-1501
    - CUHK03
  - Detection
    - MS-Coco
    - Eurocity persons
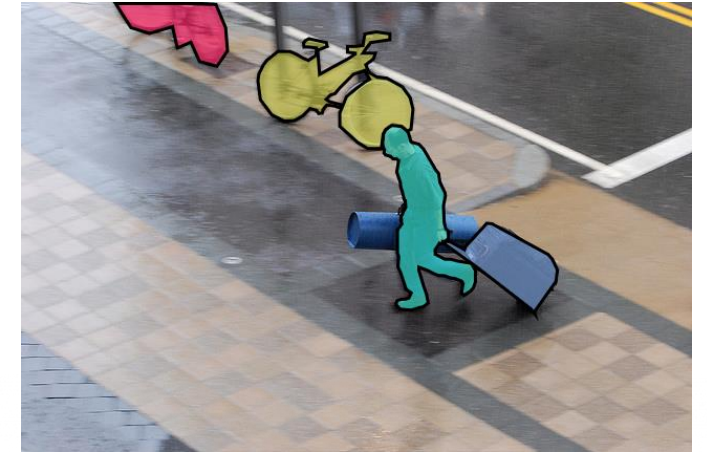
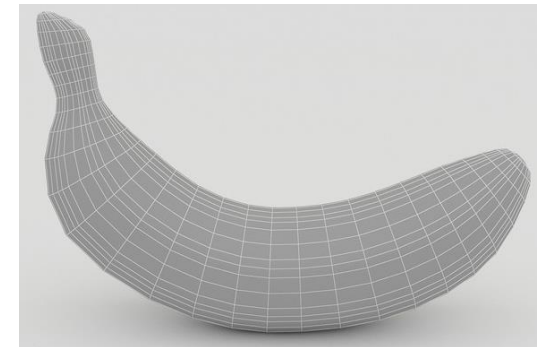- You will use the Parrot ANAFI or Dji Tello
  - Camera, WiFi-Connector, Remote Controller
  - Light-weight
  - Development using **dronekit-python** (https://github.com/dronekit/dronekit-python)
  - Working within drone regulations
    - https://www.bmvi.de/SharedDocs/DE/Publikationen/LF/flyer-die-neue-drohnen-verordnung.pdf?__blob=publicationFile

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Augmented Reality for Users with Low Vision

- Let users investigate virtually investigate objects in a scene

  - Segment Objects in Scene

  - Create 3D-Models from user's selected detection

  - Let user augment 3D-Models in AR

- Three Major Criteria:
  - Instance Segmentation/Object Detection
  - Shape estimation from images
  - Selection of Objects and Manipulation of shapes via user input in AR

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Augmented Reality for Users with Low Vision

- Usable Datasets:
    - Instance Segmentation
        - MS-COCO/COCO-Stuff (https://github.com/nightrome/cocostuff)
        - Open Images (https://storage.googleapis.com/openimages/web/index.html)
    - Shape Estimation
        - ShapeNet (https://www.shapenet.org/)
        - ABC (https://deep-geometry.github.io/abc-dataset/)

- You will use a Microsoft Hololens2 (https://www.microsoft.com/de-DE/hololens/hardware)
- Development using Unity
    - https://docs.microsoft.com/de-de/windows/mixed-reality/develop/install-the-tools?tabs=unity

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Material Classification in Construction Sites

- Given an image of the construction site scene

  - Classify the material type of each region in the scene (metal, glass, wood, etc.)

  - Perform semantic segmentation method to identify different parts of the room, such as floor, ceiling, wall, window, door.

  - Further classify the identified room parts' material type

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
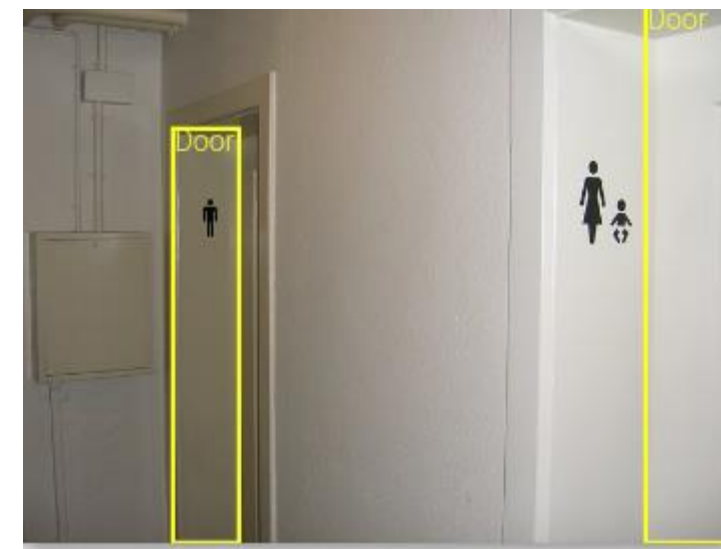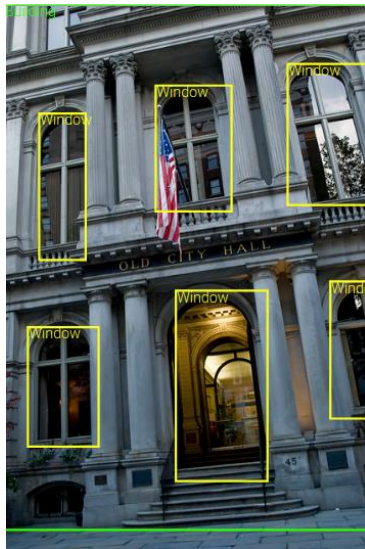Institute for Program Structures and Data Organization

Material Classification in Construction Sites

- Usage existing subsets of datasets applicable for this problem

  - ImageNet (C)

  - Places2 (C)

  - ADE20K (S) (https://groups.csail.mit.edu/vision/datasets/ADE20K/)

  - COCO-Stuff (S,D) (https://github.com/nightrome/cocostuff)

  - OpenSurfaces (C) (http://opensurfaces.cs.cornell.edu/)

  - OpenImages V6+ (C) (https://storage.googleapis.com/openimages/web/index.html)

  - Web (C)

- Evaluation using real life construction site images

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Object Detection in Construction Sites

- Given an image of the construction site scene

    - detect the object classes (scratches, stains, defects, doors, window, stairs, ceiling floor, wall).

- Use case will be based on construction site images, dataset will be provided for training the models.

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Object Detection in Construction Sites

- Usage existing subsets of datasets applicable for this problem

    - ImageNet (C)

    - ADE20K (S) (https://groups.csail.mit.edu/vision/datasets/ADE20K/)

    - COCO-Stuff (S,D) (https://github.com/nightrome/cocostuff)

    - OpenImages V6+ (C) (https://storage.googleapis.com/openimages/web/index.html)\

- Evaluation using real life construction site images

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Edge Detection in Construction Sites (1-2 People)

- Given an image of the construction site scene

  - Analyse current state-of-the-art edge detection methods for object classes doors, window, stairs, ceiling floor and wall.

- We expect the group to train the models on publicly available datasets with only these object classes.

- Use case will be based on construction site images, dataset will be provided for testing the models.

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Edge Detection in Construction Sites

- Usage existing subsets of datasets applicable for this problem

    - ADE20K (S) (https://groups.csail.mit.edu/vision/datasets/ADE20K/)

    - COCO-Stuff (S,D) (https://github.com/nightrome/cocostuff)

- Evaluation using real life construction site images

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Topic Selection

- Find a team of three people

- Each team sends us
  - A ranking of their preferred three topics
  - until 7th 23:59 of November
  - per Email at {constantin.seibold, vivek.sharma}@kit.edu

- Scenarios
  - Re-Implement not publicly available model from paper
  - Change publicly available model by trying out parameters/losses
  - Fuse two different architectures in some way
  - Use existing model for a novel task

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Create a Plan

- Check related work
  - What has been done in this topic specifically?
  - What has been done for the overarching task?
    - What pretrained models do exist?
    - What datasets do exist?

- How can you use related work for your task?

- Who focuses on what?

- Create a rough schedule for how you approach your task

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Implementation

- Choose Framework
  - **TensorFlow**, see tutorial https://www.tensorflow.org/tutorials
  - **PyTorch**, see tutorial https://pytorch.org/tutorials/
  - Torch, Theano, Caffe
  - …

- Each team can use an 8GB GPU on our servers
- Plot learning curves/results and show at weekly meeting
- Split work equally between team members

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Evaluation

- Split Dataset into distinct training, validation and test set
- Use training and validation to tune your model
- The test set is used at the very end

- Check out metrics in related work
  - Classification: Accuracy
  - Detection: mAP
  - Segmentation: mIOU

- If your project requires a user study, check in related work how similar studies are performed

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Presentation

- 15min Presentation per Team

- Explain your topic, approach and results
  - If applicable, a demo/video would be appreciated

- Allocate enough time for each team member to talk about their contributions

- What if your solution does not work?
  - No problem at all.
  - A presentation showing your development and why it didn't work is perfectly acceptable as well

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization

Organization

- Select teams of 3 students each
- Meeting schedule

- Week 0: Introduction
- Week 1: Present ideas on how to approach the problem
- Week 2: Read related work and consider how to incorporate these into your task
- Week 3: Implementation
- …
- Week 13: Presentations

- Weekly meeting for discussion and status updates with corresponding supervisor
  - Set a consistent date for weekly meetings

- Register Projektpraktikum with KIT's Studienbüro…(Modulhandbuch Nummer #102966, Teilprüfung #105943)
- For these slides, other information, announcements and updates→check website [coursemember/321meins]

Constantin Seibold/Vivek Sharma –
Practical Course Computer Vision for Human-Computer Interaction

Architecture-driven Requirements Engineering Group
Institute for Program Structures and Data Organization