

# Computer Vision and Deep Learning for Autonomous Driving (CVDLAD)

**Introductory meeting - Summer semester 2019**

PD Federico Tombari, Prof. Nassir Navab,  
Tutors: Fabian Manhardt, Helisa Dharmo, Shun-Cheng Wu, Manuel  
Nickel

# Seminar contents

- The seminar includes a selection of the most recent and relevant papers in the field of computer vision and deep learning aimed at autonomous driving
- Papers are selected to cover different aspects of the topic:
  - Semantic Segmentation
  - Scene Understanding
  - Multi-modal sensor fusion
  - SLAM and 3D reconstruction
  - Synthetic dataset creation
  - Car Simulators
  - Semi-supervised large-scale dataset annotations
  - Trajectory planning and lane change prediction

# Goals

- You are going to learn:
  - about relevant works in the field of Computer Vision and Deep Learning
  - what is Autonomous Driving and why it is a relevant technology for future applications
  - how to read and understand a scientific article
  - how to write a scientific report
  - how to give a talk to an audience, and deal with related questions afterwards

# Seminar Schedule

- 5 sessions, 1 every Monday, 4-6pm
- 3 presentations per session + 1 external talk from an industry expert
- Seminarraum 03.13.010
  
- Paper assignments:
  - selected students can express up to 3 preferences
  - We will then match them to a paper and tutor trying to maximize global happiness

# Presentation

- Each presentation is 15/20 minutes + 10 minutes for Q&A
- Slides templates (Powerpoints, Latex, ..) provided on website
- The presentation should cover all relevant aspects of the paper
  - Introduction and state of the art
  - Main contribution(s)
  - Experimental results
  - Discussion, summary and future work
- The presentation should be self-contained
- All students are expected to attend all presentations and interact during Q&A (this will influence your final mark)

# Report

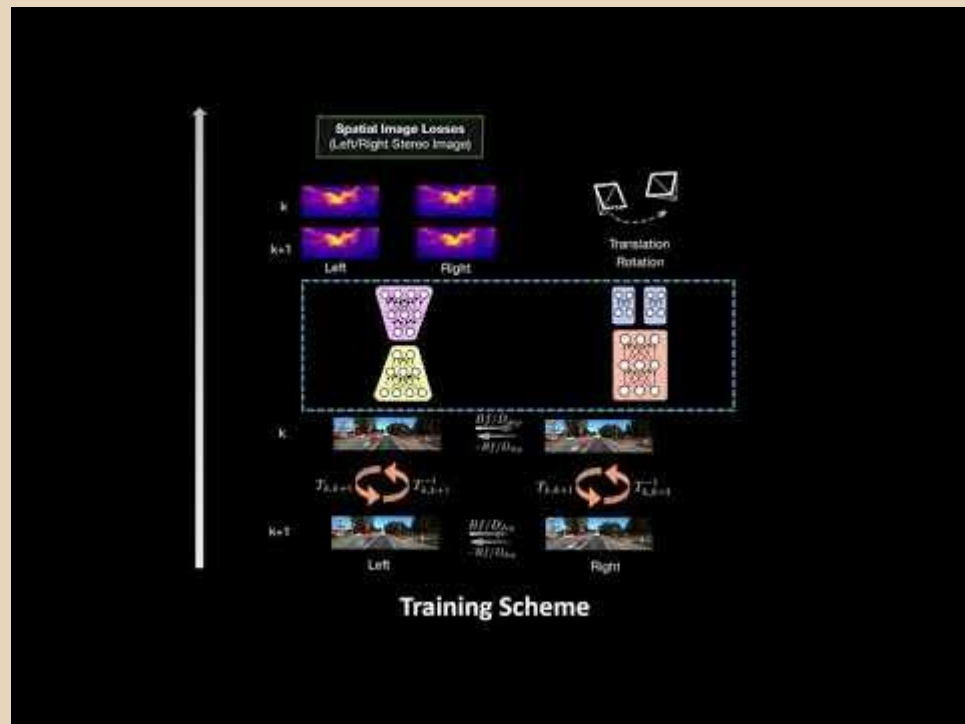
- The report should summarize the paper in the way it has been presented during the talk, and provide the student's opinion concerning the main contributions and impact
- Language: English
- Max 8 pages
- Template on course website
- Once ready, send the report to supervisor, within **two weeks** from the day of the presentation

# Evaluation criteria

- Quality of presentation (both regarding slides and speech)
- Quality of the report
- Comprehension of the scientific contents of the presented work
- Interaction and participation during the other talks

# R. Li et al, “UnDeepVO: Monocular Visual Odometry through Unsupervised Learning”, 2018

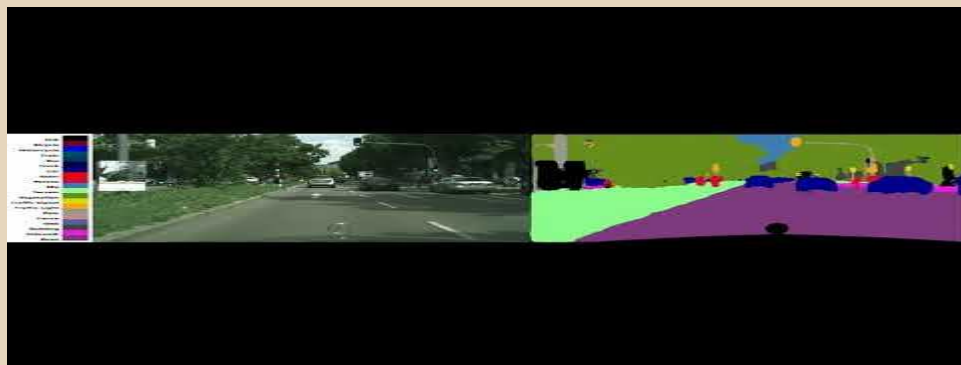
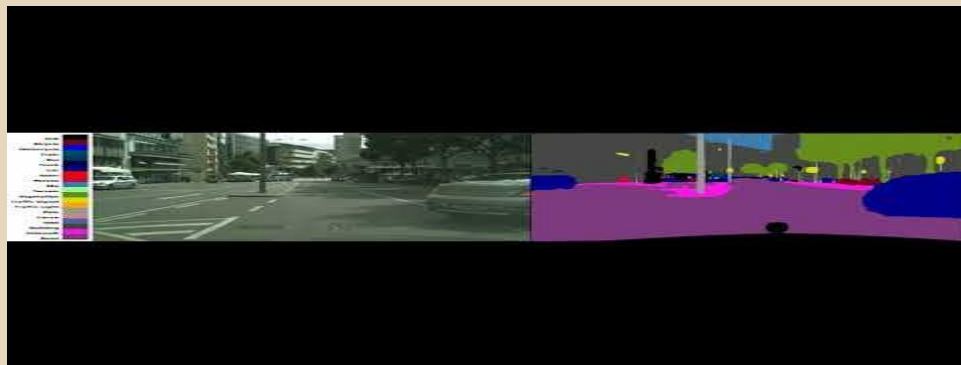
- Unsupervised Monocular 6-DoF camera pose and depth estimation
- Absolute scale recovery through training on stereo image pairs
- During inference only monocular images are employed





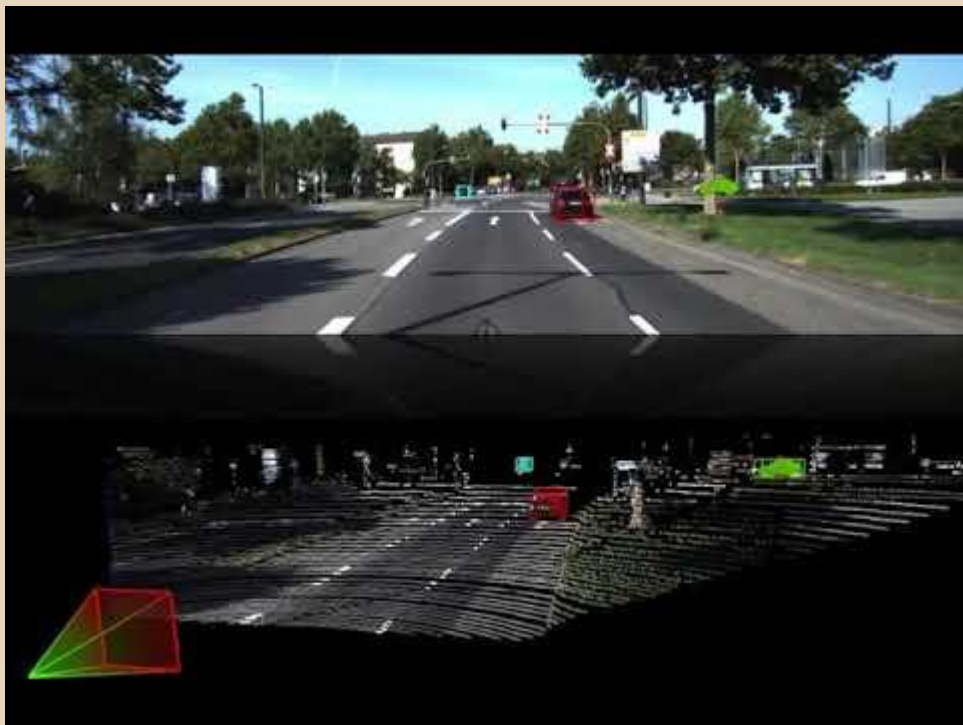
# S. Mehta et al, “ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation”, 2018

- Semantic segmentation of high resolution images under resource constraints
- Based on a new convolutional module, efficient spatial pyramid (ESP)
- Efficient in terms of computation, memory, and power.



## F. Manhardt et al, “Roi-10D: Monocular Lifting of 2D Detection to 6D Pose and Metric Shape”, 2018

- Monocular 6-DoF car pose and shape estimation
- Lifts 2D detections from RetinaNet to 6D pose through a novel loss that operates directly in 3D, without the need of multi-loss balancing
- Shape parameter estimation and decoding through a Fully-Convolutional Auto-Encoder



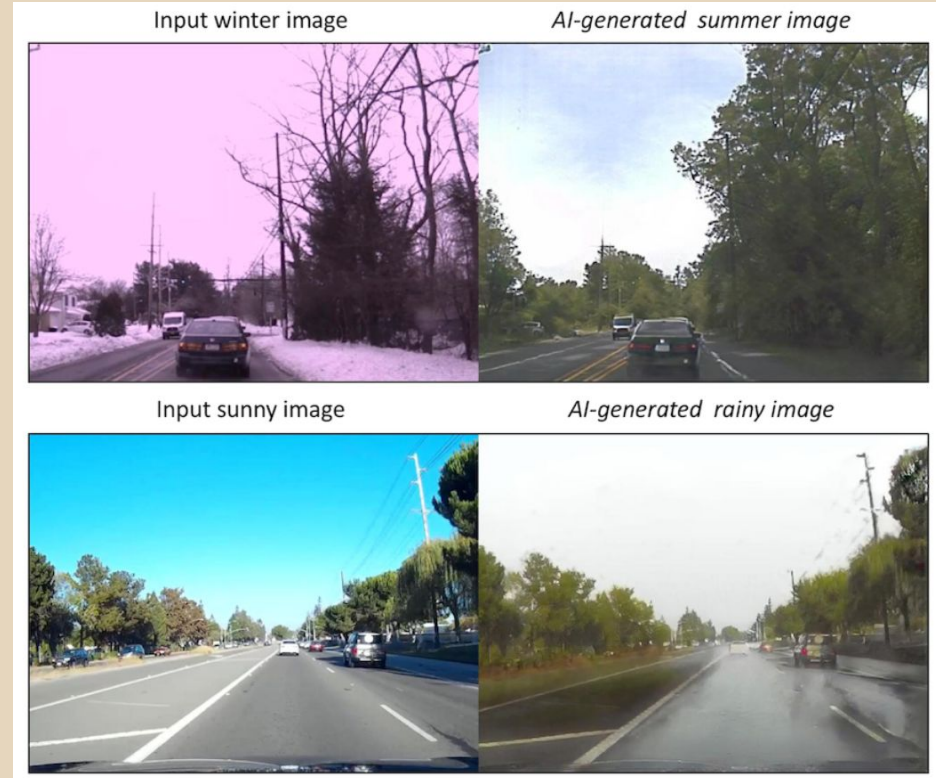
## T.-C. Wang et al., “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs”, 2018

- New method for synthesizing high-resolution photo-realistic images from semantic label maps using conditional GANs
- Employing novel multi-scale generator and discriminator architectures to generate images of size 2048x1024
- Extension of the framework to enable interactive visual manipulation



## M. Liu et al., “Unsupervised Image-to-Image Translation Networks”, 2017

- Use of Generative Adversarial Networks (GANs) to generate novel illumination/weather conditions and image modalities
- Works in an unsupervised manner
- Applied to street scenes for autonomous driving applications



# A. Maqueda et al, “Event-based Vision meets Deep Learning on Steering Prediction for Self-Driving Cars”, 2018

- Steering angle prediction is not robust to many scenes characterized by high dynamic range (HDR) due to e.g. motion blur or low light
- Steering angle prediction by means of an event camera (naturally respond to motion in the scene)
- Advantage over standard cameras: Low latency, high temporal resolution and HDR

The system produces a prediction of the steering angle of the vehicle at 1/3 seconds in the future.

