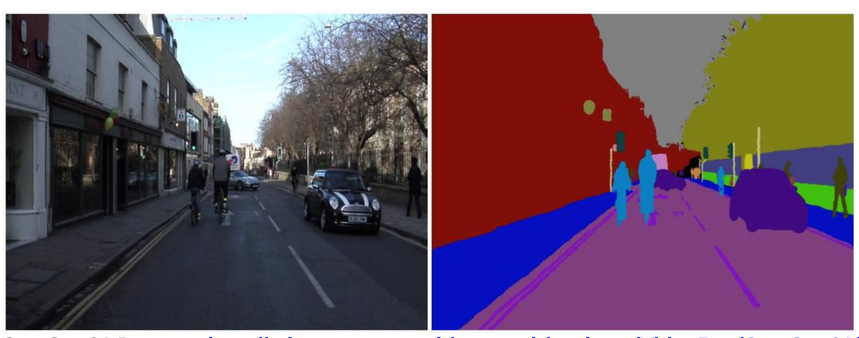
Large Scale Photo-realistic Image Synthesis for Neural Networks Training in Autonomous Driving

Joel Kronander, Magnus Wrenninge, Apostolia Tsirikoglou, Jonas Unger | LINKÖPINGS UNIVERSITET

Autonomous driving is one of the most revolutionary technologies of our time. Computer vision, among others, contributes to autonomous driving development with vision-based tools, like object detection, recognition of traffic signs and road segmentation. The last decade, Deep Learning's Neural Networks (NNs) with numerous layers, have proven to outperform in standard computer vision tasks. Nowadays, the easy access to image data and the enormous available computing power has enabled the development of high level software for building and training large scale networks. However, designing NNs still lacks (1) systematic approaches for specific problems, and (2) suitable ways of detecting and/or generating data for the specific tasks. In our work we build an image synthesis framework for large scale generation of photo-realistic data to train and test deep NNs used in automotive applications.

Background & Motivation



CamSeq01 Dataset, http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamSeq01/

Most training data is *captured images*, labeled and segmented by hand in a fully manual slow, expensive and not completely reliable process, that can affect the learning algorithms.

Related work on *synthetic data*, even though they are simplistic approaches based on game engines rather than photo-realistic imagery, relatively small and in good weather conditions, proves that they perform well alone or along real data in training NNs.

Research Goal & Questions

We aim to develop the required theory and tools to generate training data using image synthesis, for a variety of scenarios in urban environments. The long term goal is to provide with a complete, photo-realistic dataset of *millions of synthetic images of various street scenes and viewpoints*, that can be used for training networks in automotive applications.

Through this project we attempt to address significant research questions, including:

- Given a specific problem, how the world can be divided into a suitable set of concepts/labels?
- Which would be most efficient sampling approach for the dimensions that represent these concepts?
- Which would be the most effective network model, batch size and the real/synthetic data fraction?

Key Concepts & Methods

- *Photo-realism*: Study the impact of realism on the results and the network generalization error.
- Teach the network as many instances of the concepts as possible (small generalization error). In practice, this means to include as many variations of the chosen labels as possible.
- Semantic segmentation: using Fully Convolutional Networks (FCNs)

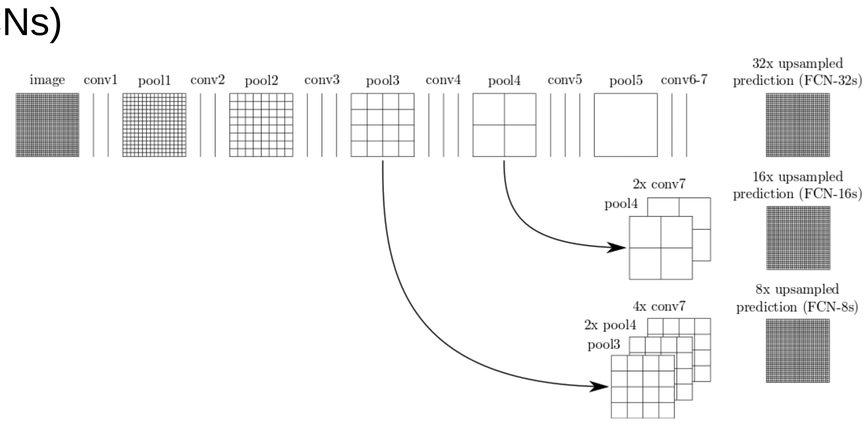
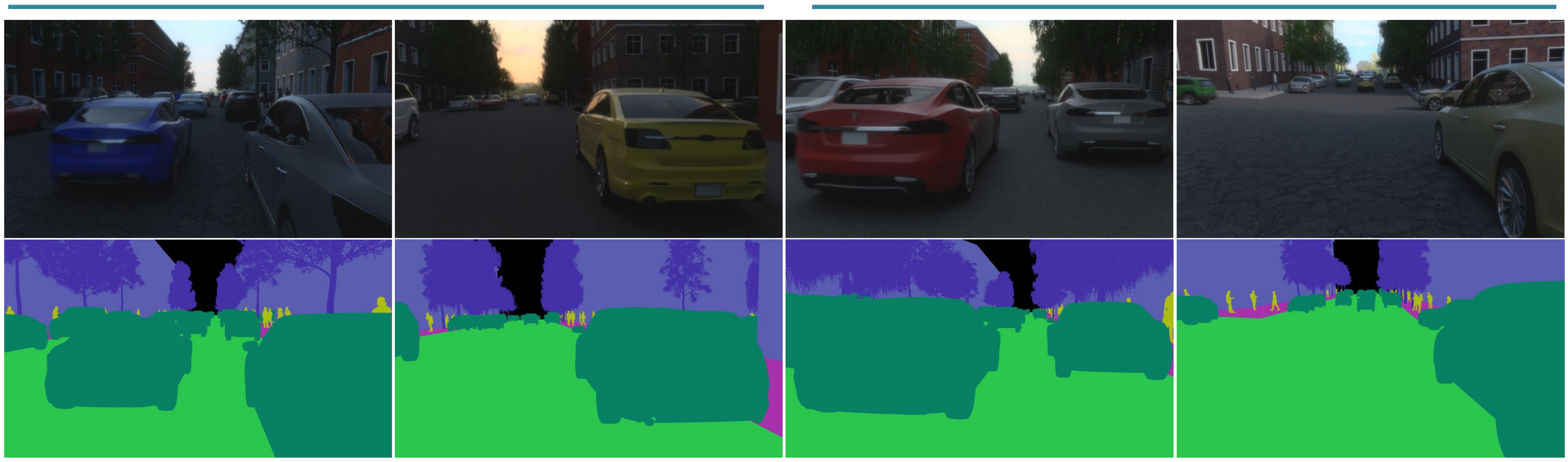


Illustration of FCN connections https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Project Plan & Milestones

1. Generation of millions of variations of virtual scenes and exploration of rendering pipelines for the specific task.
2. Reproduce and *compare* with previous work on synthetic data, like SYNTHIA http://synthia-dataset.net/, with FCNs and test with real data.
3. Investigate efficient sampling strategies and training sets of mixed real and synthetic data.
4. Investigate how controlling the context distribution in the data (vehicle types, colors, viewpoints etc.) affects the training.
5. Explore different network architectures, looking for the maximum information each architecture can capture and trying to formulate one or more approaches for the specific problem.
6. Investigate active learning strategies to generate data that trains the network to perform better at tasks with low performance scores.



WASP WALLENBERG AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM