Active Learning on video data for Deep Neural Networks (Computer Vision models for Autonomous Driving)

Motivation and background

Deep neural networks constitute the state-of-art for many problems in perception and will be an integral part of perception systems in autonomous vehicles. These machine learning models require a large amount of diverse training data which are largely annotated by humans. We would like to reduce the amount of data needed to train these models, while not compromising on performance. This thesis aims to investigate ways of selecting data to annotate, given large scale unlabelled video data and models trained on existing but limited amounts of data.

Problem description

Data to train perception models are typically recorded in sequences. Each sequence comprises several hundreds of frames recorded in a short time interval (~10 seconds). In order to train a computer vision model, say an object detector, we typically sample frames from a large collection of unlabelled sequences, annotate them and use them for training.

However, given a large number of sequences, we would like to select the most useful samples to annotate such that every sample can help improve the model performance. For instance, increasing the diversity of samples (locations, types of objects, weather conditions, etc) may help increase the utility of the overall training dataset.

In this thesis, we are interested in looking specifically at video data (sequences of images) and how perception models perform on them. In general, deep neural networks tend to be unpredictable, and their output can vary dramatically between frames. For example, in the illustration below, an object detection model outputs a false detection for a pedestrian in one frame, but not in the very similar surrounding frames.

Another example below shows a segmentation model for road detection also flickering dramatically across frames (due to the shadow).
The thesis should aim to investigate ways of automatically picking out such interesting frames, so that they can be annotated and added to the training set). A starting point could be to look at intermediate feature representations of the models and visualizing them for different samples to understand why similar inputs still induce different behaviour.

Furthermore, the thesis can also explore additional ways of using unlabelled video data for training through semi-supervised learning.

We expect the students to

- Perform a literature study on promising methods for active learning and data selection.
- Study the neural networks currently developed at Zenuity for road/lane classification and object detection, and if necessary extend them to enable the investigations needed for this thesis project.
- Develop techniques for selecting frames to annotate that can be applied to the different computer vision models.

Applicant profile

2 students interested in deep learning for computer vision. Experience with Python programming, and familiarity with at least one of the deep learning libraries (e.g. Tensorflow, PyTorch or Keras), are requirements.

Contacts

Erik Werner (erik.werner@zenuity.com), Amrit Krishnan (amrit.krishnan@zenuity.com).

Applications should be submitted via the webpage: http://career.zenuity.com/jobs/137573