Convolutional Neural Networks for driving condition classification

Modern computer vision models using deep convolutional neural networks are very good at classifying and detecting objects from images. By utilizing transfer learning and pre-trained models that have been trained on massive data sets, it is possible to train robust classifiers on different problems with relatively limited computational resources and data.

The goal of this thesis would be to apply these techniques to get a reliable classifier for detecting weather conditions on roads around Finland. The classification problem to be solved would consist of at least two different parts. First, to identify the adverse weather conditions that might affect drivers on the roads, i.e., whether it is raining, snowing, foggy, or clear. We could add more classes, such as sunny, overcast, and so on, but most likely, these are not interesting for the task at hand. The second task would be classifying the road conditions on the surface of the road. The possible classes for the road conditions could be icy, slushy, snowy, wet, and clear. It is possible that combining some of these classes will be necessary, but these would be the likely beneficial classes that we can detect for providing useful information.

Various approaches have been tried for creating models that can predict some of the classes mentioned above. Generally, they focus on the weather classification task. Recently nearly all approaches use convolutional neural networks, and most of them use pre-trained models as the basis for the models. We will use similar approaches and try to extend previous attempts to get better results. Most of the previous approaches try to classify the weather conditions in a general picture of the outdoors. Here we will use traffic camera footage, which means that using image data of the weather phenomena in a general sense may not be that useful for improving performance due to the uniqueness of traffic cameras in terms of their positions and angles. Thus most of the data used will be from our own data set, and possibly the effectiveness of some available weather image could be evaluated and possibly used to augment the data set.

As the data set to train these models, a collection of videos from existing traffic cameras from roads around Finland will be provided. As this data will be raw and unlabeled, some manual labeling will have to be done in order to get a sufficient amount of labeled data for the models that we are going to train. Besides just creating manual labels for the videos to train supervised models, we will also explore the possibility of other ways of creating labels. We will attempt to use semi-supervised learning to automate labeling and active learning to find data points that might be useful targets of manual labeling. Then the effectiveness of these techniques will be evaluated to see how beneficial they are for improving our metrics on the classification tasks.

Since our data consists of videos, we will try to find the best models by training both image and video classifiers and then compare their performance. The baseline models that we will be attempting to improve on later will be image classifiers that use various pre-trained networks that have been trained on other massive datasets. We will test models that have been trained on ImageNet and possibly on the Places365 or other data sets and comparing them to find the best base model. The interesting network architectures that we will be comparing for using as the base model will include at least ResNet and EfficientNet and possibly others if they seem like they could provide a performance boost. To improve the base classifier, we will first introduce the concept of attention over images and try adding an attention layer to the network to see how it would impact the performance of the classifier.

The first video classifier that we would train would be a basic recurrent neural network over embeddings of video frames created with a convolutional neural network. Here we might also be able to compare the different pre-trained architectures for the CNN portion of the network. With this network as a baseline, we can look into some other improvements to create a better video classifier by augmenting the model with an attention layer or trying out different kinds of video classifiers.

Finally, we can compare the two distinct ways of solving the two described classification problems. In the first case, we would create separate classifiers for the weather conditions and the road conditions. The other option would be to use multi-task learning and to create a single network that would produce both outputs. Then we can compare the performance of both of these approaches and see if one is superior to the other.