Experimenting with the architecture of neural networks on image segmentation and depth prediction problems

MSc. thesis outline

Zsófia Klára Ujfalusi

AI and neural networks are in the forefront of research currently. People and companies begin to use them more and more for work and for everyday life purposes. For example, web shops offer 24h help with an AI robot assistant, Youtube, Facebook and web shops offer personalized music, advertisements, products by building recommendation systems.

One of the recent applications of AI is to automatize transportation. With the development of AI and deep learning self-driving cars has come within reach. Self-driving cars can potentially eliminate human mistakes leading to less accidents. At the same time people can use their time spent with driving for other purposes. There are multiple companies developing self-driving cars and promising level 5 self-driving cars (without any human control) to be available in the next decade. One of them is AImotive, a rapidly evolving Hungarian startup. They use the camera images and process them with the AI system for example to find objects surrounding the vehicle. This thesis was written in cooperation with this company.

I examined two problems: segmentation and depth prediction. I observed that with connecting information in lower level layers to layers representing high level features I can improve the results (Accuracy, Jaccard index, Depth loss) substantially. In the case of segmentation this overall improvement was achieved by improvement in classes of smaller objects (pole, traffic light, traffic sign). I also saw that the exact way of connecting the lower level information is not crucial. In both cases I found that the structure called Recombinator Network led to the best results.

I also examined how different upsampling methods work. For the models without branching I found that in case of depth prediction unpooling gave the best results. Deconvolution in 2 steps was the second best method for this task, while for segmentation it was the best. On the other hand for the more complex models (containing branching) when I used deconvolution instead of bilinear upsampling I did not get better results, on top of that I had to use more parameters.

I found that using weights of a network trained on ImageNet as a starting point improved the results substantially for segmentation. However it barely improved the results for depth prediction, but by visualization I have seen that the networks learned entirely different features.