From On-Road to Off: Transfer Learning within a Deep Convolutional Neural Network for Segmentation and Classification of Off-Road Scenes

Christopher J. Holder\textsuperscript{1,2}, Toby P. Breckon\textsuperscript{1}, Xiong Wei\textsuperscript{2}

\textsuperscript{1} School of Engineering & Computing Sciences, Durham University, UK
\textsuperscript{2} Institute for Infocomm Research, ASTAR, Singapore

Results

Classification accuracy as the CNN is trained: on full off-road data set after different amounts of on-road pre-training (top); on different sized subsets of off-road data after no pre-training (bottom left); on different sized subsets of off-road data after 30,000 iterations of on-road pre-training (bottom right)

Example images from our off-road test dataset (top row), their respective manually labelled ground truth images (middle row), and results output by our best performing CNN (bottom row)

Methodology

- We use the CNN architecture of Segnet [1]: 13 encoder layers (convolution and pooling), followed by 13 decoder layers (convolution and upsampling) to give full pixelwise classification.
- CNN is pre-trained on the Camvid [2] dataset of 367 manually labelled urban road scene images. (63,000,000 labelled pixels)
- Our pre-trained CNN is then trained for 10,000 iterations on our dataset of 295 manually labelled off-road images. (35,000,000 labelled pixels)
- Each pixel is assigned one of eight class labels: sky, water, dirt, paved road, grass, foliage, tree, or man-made obstacle.
- We evaluate performance on 37 unseen test images, measuring accuracy as the proportion of manually labelled pixels that are correctly classified by the CNN.

Conclusions

- Headline figure of 0.92 accuracy for the CNN trained on the full off-road dataset after 30,000 iterations of on-road pre-training.
- A CNN that has undergone more pre-training will reach optimum performance sooner, however after ~5000 iterations performance gains are marginal.
- A CNN that has trained on a smaller dataset will invariably perform worse.
- The benefits of pre-training are more evident when a small dataset is used: The 8 image dataset only achieved an accuracy of 0.51 without pre-training, but reached 0.81 with pre-training.
- Overall, pre-training is no substitute for dataset size, however in circumstances where only a small dataset is available, pre-training with a larger dataset can help to improve performance.

\begin{itemize}
\end{itemize}