Deep Learning for Road Segmentation

Stanford University, CS230: Deep Learning

Authors: Chandler Watson (watsonc) , Justin Dieter (jdieter), Anjali Roychowdhury (aroyc) **Mentor:** Patrick Cho

Predicting:

Road segmentation for autonomous vehicle applications is difficult, as images can include partial occlusions, shadows, etc. Worse, such applications typically require **real-time performance**. Our team modified Marvin Teichmann's KittiSeg framework to, using a genetic algorithm, **optimally drop components** of a FCN8-VGG16 model to optimize for MaxF1 while running in less time than the original model.

Data:

All models were patch-trained on the **KITTI Road Detection** dataset with a 241/48 (289 total) image train/validation split. Inputs are color road images in various locations/lighting conditions with ground truth color maps (Figure 1).

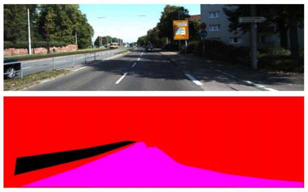

Figure 1: *Training Example with Ground Truth from the Kitti Road Detection dataset*

Features:

During training, the FCN8 model accepts a single random 256x256 image patch, encodes it, then upsamples it into a block with one softmax output per original pixel. **No other features** (i.e. LIDAR, stereo imaging, etc.) are used.

Model:

The standard KittiSeg model uses an FCN8-VGG16 model as an encoder, consisting of five pooling layers and thirteen standard convolutional layers:

$$\mathbf{y}_{ij} = f_{ks}\left(\{\mathbf{x}_{si+\delta i, sj+\delta j}\}_{0 \le \delta i, \delta j \le k}\right)$$

It then uses a series of transpose convolutions for upsampling, following the definition below:

$$C_1(y^i) = \frac{\lambda}{2}\sum_{c=1}^{K_0}\|\sum_{k=1}^{K_1} z_k^i \oplus f_{k,c} - y_c^i\|_2^2 + \sum_{k=1}^{K_1}|z_k^i|^p$$

We optimized runtime using a **genetic algorithm** with mutation (random bit toggling) and crossover (random selection with union) on binary vectors representing layers to drop, splicing in `tf.Tile` ops wherever shape changes occur. Evaluation was performed after 250 steps as a benchmark.

Results:

Despite the simplicity of tiling relative to convolution, runtime was **not strongly affected** by any layer drops. As a follow up, a FCN8 with `conv2_2`, `conv3_2`, and `conv4_2` bypassed was trained for 1.5k steps alongside a vanilla FCN8. All results are in Figure 2.

	Max F1	Avg. Precision	Runtime (ms)
FCN8	**90.2379**	**90.6890**	358.1518
Learned FCN8	17.6806	30.0485	**356.5779**
FCN8 (1.5k steps)	**93.1652**	**91.5854**	**361.4553**
Trimmed FCN8, (1.5k steps)	33.3214	21.6512	372.0243

Figure 2: *Summary of Performance Metrics for All Models*

Trimmed FCN8

FCN8

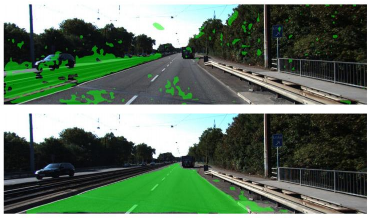

Figure 3a and 3b: *Visualization of Predictions of FCN8 and Trimmed FCN8 after 1.5k steps*

Discussion:

Although layer splicing appears to hurt FCN8's ability to classify images, trimmed FCN8 sometimes finds road-like features (Figure 3). Loss **appears to stagnate** when training with several learning rates (w/ Adam), indicating the new loss function may be stuck in a local minimum. It was later found that statistics were recorded on dropped layers, suggesting further inspection of the TF graph **could lead to decreased run times**.

Future:

Future work could include a second attempt with longer training and extensive hyperparameter tuning, as well as **exploring alternatives** to our method: optimizing the network for parallelization, following Molchanov et al. on **network pruning**, or performing **full image training** like Oliveira et al. Alternatives to the genetic algorithm (i.e. cross-entropy method) could also be explored.

References:

M. Teichmann. "KittiSeg." Internet: https://github.com/MarvinTeichmann/KittiSeg, Oct 16, 2017 [Mar 1, 2018].

G. Oliveira et al. "Efficient and robust deep networks for semantic segmentation." *The International Journal of Robotics Research*. Jun 2017.

P. Molchanov et al. "Pruning Convolutional Neural Networks for Resource Efficient Inference." https://arxiv.org/abs/1611.06440. Jun 8, 2017 [Mar 1, 2018].

J. Fritsch, T. Kuhnl, and A. Geiger. "A new performance measure and evaluation benchmark for road detection algorithms." in ITSC, 2013.