UAV Navigation above Roads with CNN's

Thomas AYOUL, Toby BUCKLEY, Felix CREVIER
Stanford University

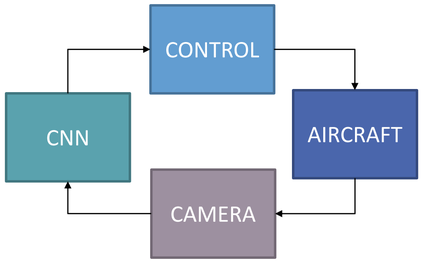
Objective

UAV navigation over roads without relying on a preloaded map

- Train a deep segmentation neural net for road detection
- Develop a tool to create/update a road graph from segmented image
- Implement the feed-forward aircraft controller to follow edges of graph
- Embed system onboard a drone and test in the field

Main challenges

- Obtain good per-class accuracy on landscapes different from training set
- Simplify architectures to enable real-time inference onboard aircraft



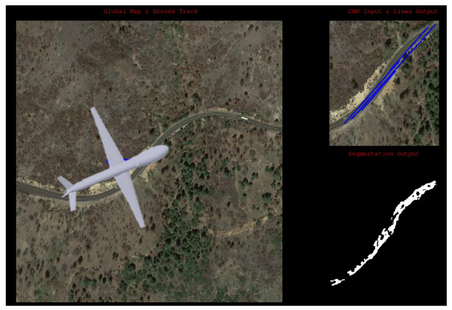
Outline

UAV with GoPro

Simulation

Simulation environment for software testing

- Pygame – OpenGL environment
- Run-time inference using best trained CNN
- Realistic aircraft dynamics and navigation
- Global map is over Yosemite National Park



Simulation interface snapshot

1. V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv:1511.00561v2 [cs.CV], 2015
2. Volodymyr Mnih. Machine Learning for Aerial Image Labeling. PhD thesis, University of Toronto, 2013.
3. F. Lei, et al. "Convolutional Neural Networks for Visual Recognition". http://cs231n.github.io/

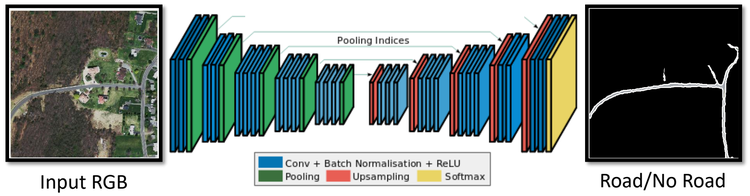
Dataset

University of Toronto Massachusetts Roads & Buildings dataset

- Data collected and annotated by Volodymyr Mnih
- 1000 RGB satellite images split into 16 375x375 input images.
- Networks trained on 4500/16 000 sub-images
- Annotations are pixel-wise road/no-road booleans

Network Architecture

CNN based segmentation Encoder-Decoder



Input RGB — Road/No Road

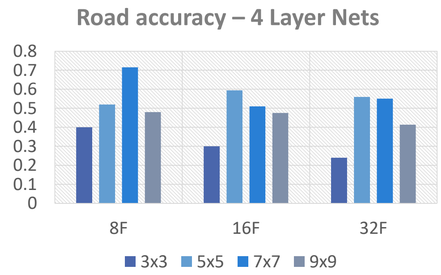
- Architecture based on Alex Kendall's Segnet
- Implementation is done in a modified Caffe framework

Our Best network

- 4 *Conv + Batch + ReLU* layers and 4 *Upsampling + UpConv* layers
- 7x7 Kernel, Stride 1, 8 Filters, 2x2 Max Pooling
- 23,338 learned parameters
- Class priors: Road 0.2 | No Road 0.8
- Performance: **72% Road accuracy**

Hyperparameter Tuning

Sweep on Kernel Size and Number of Filters



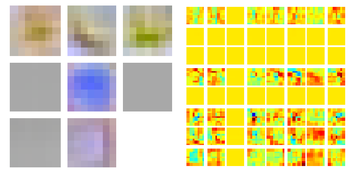
Observations

- Road accuracy is a better performance metric as it's prior is low
- Medium sized kernels seem to perform best : larger scale features
- More filters ≠ better accuracy: low geometric complexity

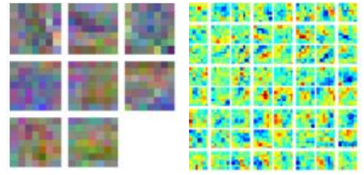
Analysis

Filters – Best Net

- Filters highlight color patterns and some forms of edges
- At higher learning rates, filters are smooth or dead



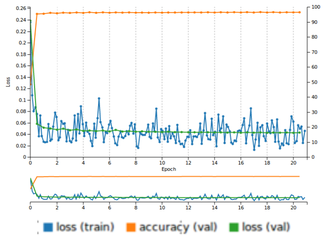
$\eta = 0.01$



$\eta = 0.0005$

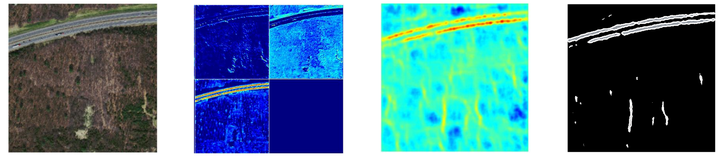
Loss – Best Net

- Loss is based on **overall pixel accuracy**
- Loss exhibits strange oscillatory behavior, even at low learning rates
- A loss based on **road accuracy** could lead to better results



Color vs edge detection

- The layer 1 activation maps illustrate a gray/not gray filter (color detection)
- The layer 4 heat map labels creases in the foliage as roads (edge detection)
- It would seem edges require more abstraction, yet CNN classifiers usually have edge filters in the first layer. Possibly a training issue.



Original image — Activation Map 1 — Heat Map 4 — Class inference

Conclusions

- CNN's with powerful Encoding-Decoding Kendall architectures were trained
- Good performance was obtained the test set, but less so on images from other landscapes
- To improve accuracy, techniques from Mnih, such as the TABN loss and structured prediction post-processing should be investigated