

Data

I used only Inria AIL dataset for training. The dataset was split so that the validation set consists of the first 5 images of a city, i.e. chicago1.tif, chicago2.tif etc.

Model

For training I used a UNet-like architecture with ResNet34 encoder pretrained on ImageNet and skip-connections between encoder and decoder.

Augmentations

The images were randomly (by 50% of chance) augmented by the following operations:

- horizontal and vertical flip
- affine transforms
- piecewise affine transforms
- perspective transforms
- brightness perturbation
- contrast perturbation
- hue and saturation perturbation
- additive gaussian noise
- sharpen

Training

The model was fed with random 384x384 color image crops with batch size 12. The learning rate changed from $5e-5$ to 0 as a cosine every 20k iterations. The entire training took 200k iterations.

Loss

The loss is $BCE + (1 - dice)$, where BCE is binary cross entropy and dice is Dice coefficient.

Inference

A 5000x5000 image was cut by 2048x2048 patches with 25% overlap. Each patch was augmented by flipping it around its horizontal and vertical axes and fed to the network, then the predictions were averaged.