

Class Project 2 Report: Road Segmentation Challenge

Angerand Grégoire, Goullet Boris, Grondier Julien

Abstract—In this project, we were given a dataset consisting of aerial photographs of suburban areas. Our task was to apply different machine learning methods to this dataset to train a model that would predict where roads were present in these photographs.

summarize results

I. MODELS APPLIED

A. Logistic regression

At first, we tried using a logistic regression. We did this by cutting up the images into small patches (either 4x4, 8x8, or 16x16) and training a logist model on 3 features for that patch: mean Red pixels values, mean Green pixel values, and mean Blue pixel values.

However, results were less than satisfactory, even with more/different features or patch sizes, so we quickly dropped that idea.

B. Convolutional Neural Network

We thus turned our focus to CNNs. We started off by using the provided code from `tf_aerial_images.py`.

We played around with its parameters and noticed that we had much better results with a patch size of 4x4, instead of the default 16x16. We also reduced the size of its convolution filters to 4x4 instead of 5x5, to match the patch size.

Custom CNN classifier: We then tried building our own Convolution Neural Network rather than using the provided one, for better modularity. We were aiming for the architecture in Figure 1.

However, after a lot of debugging, we failed to achieve anything and had to fallback on the provided file. The code can be checked in `custom_classifier.py`

II. FEATURE PRE-PROCESSING

Our CNN works on sub-patches of 4x4x3 (RGB) of the training images with labels from the corresponding “groundtruth” patch.

We train the CNN by cutting up 500 different source images: the 100 given as a dataset, plus variations of

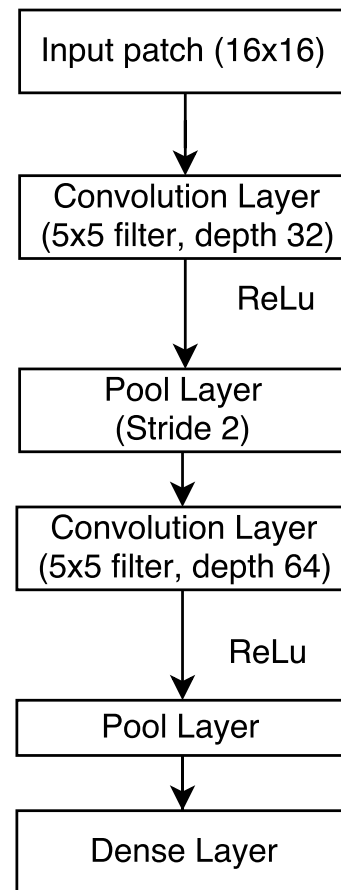


Figure 1. Architecture of our attempted custom CNN

those (transposed, flipped horizontally, flipped vertically, and flipped both vertically and horizontally).

Furthermore, we tried adding a new label as groundtruth, manually annotated, corresponding to the position of roofs. However, this did not improve prediction rate compared to having only ‘background’ and ‘road’ as labels.

III. RESULTS POST-PROCESSING

After obtaining the raw prediction from the convolutional neural network, we do a post-processing step in order to remove most of the noise.

The post processing consists of a series of two convolutions.

The first one has been designed to fill single patch holes in the prediction. The final one uses a cross shaped kernel 2. This shape has been chosen to match the many right angle intersections present in our dataset. In some cases the roads are not aligned to the image axes, to handle those we extract the most significant straight lines from the image using Hough transform and rotate our kernel to match their orientation.

These convolutions reduce the noise significantly. We tried to improve the result further by discarding small unconnected patches from the final mask. While this reduced the false positive amount by removing a few roofs, it did also discard small roads leading to a slightly worst prediction overall.

Our result composed of 4x4 blocks is converted to the required 16x16 blocks by `mask_to_submission.py` which simply takes all 16x16 blocks with a mean value greater than 0.5 (i.e. where more than half of its sub-patches are labeled ‘road’).

effect, private roads — in our opinion, car parks should be considered roads, despite what the training set states.

Our model also misses roads when they are covered by a large tree — although technically one can’t be sure there is a road under it, we just infer its probable presence from the presence of two road parts on either side. Furthermore, when a road is very thin and very close to the buildings surrounding it (e.g. European cities’ historical centers) it is often not detected.

V. SUMMARY

While we had high hopes for using a custom CNN architecture, we are disappointed that we weren’t able to build a functional one. However, using the basic, provided CNN, we achieved results that we would deem satisfactory.

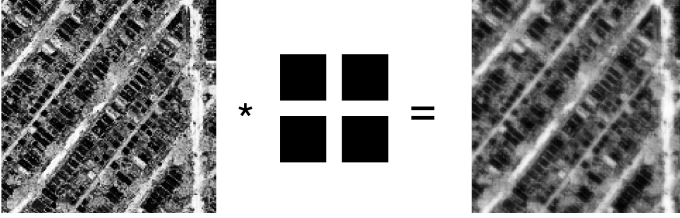


Figure 2. Prediction convolution



Figure 3. Prediction before post processing, prediction after post processing, final prediction overlaid on the original image

IV. DISCUSSION

We remark that our model has a hard time differentiating flat, grey roofs from the road, probably because those are expansive grey surfaces, exactly like roads. Car parks are also often marked as roads, as they are, in