Methods for improved road segmentation performance

Marc Salvadó Benasco, Olivier Lepel, Ke Li ,
Group: Olivier+Ke+Marc,
Department of Computer Science, ETH Zurich, Switzerland

July 2023

1 Introduction

Our goal is to build a model able to segment roads in satellite images of land. We are given a training set of 144 images and the 144 expected segmentation results (which are binary images), a test set of 144 images and a baseline F1-score to beat. The images are 400 px by 400 px. We start by introducing two baseline algorithms that beat the Kaggle problem baseline score. We then look at some optimisations of their implementations. Finally, we introduce a problem-specific post-processing technique that leverages the particular geometry of this particular setting.

2 Two baselines : Segnet and U-Net

Because the information that will help us decide whether to classify a pixel as road or not is mostly local to this pixel's neighbourhood, and because we want to put in place the same detection of patterns everywhere in the image, it is natural to use Convolutional Neural Networks for this Segmentation Task.

A question to decide on was at what resolution the feature extraction would be carried out. Applying CNNs at high resolution images is very computationally expensive, and thus these models need to be shallow. Alternative deep models downsample the image to a small size in order to grasp global features, and during the later upsampling, they add a previous residual connection that provides local information. We have taken the structure and hyperparameters of two known models that apply such schemes: Segnet[1] and U-Net [2].

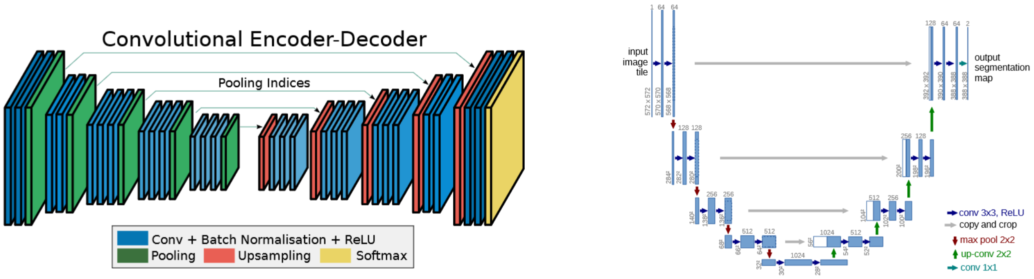


Figure 1: The SegNet and Unet structures as introduced in [1] and [2], respectively.

For experiment I and II —as indexed in Table 1—, we have taken the structures of the SegNet and U-Net models, respectively, except for some details such as the image size and the padding. No weights have been imported, but instead we have trained our models from scratch.

These baseline models have acheived a F1-score of 0.87137 (SegNet) and 0.87897 (U-Net), already beating the required F1-score of 0.86380.

3 Further optimization of U-Net

We now compare the aforementioned baseline models to the U-Net model after some improvements.

In experiment III, the addition to U-Net of batch normalization before each convolution layer has yielded a significant improvement in F1-score to 0.89897.

Keeping the batch normalization layers, the addition of dropout as regularization method (experiment IV) or the construction of an augmented dataset (experiment V, later explained in Section 4) have not improved the score further, with respective F1-scores of 0.89213 and 0.89626.

The use of an external unlabelled dataset of satellite road images to unsupervisedly initialize the model weights, later explained in Section 4, has improved the best F1-score so far to 0.90511, which is the best score we have achieved before any post-processing.

4 Pre-processing

We have performed two types of pre-processing.

One the one hand, in experiment IV, we have built an augmented dataset by modifying the training images via translation, rotation, flip, scale and noise injection. This dataset expansion has not yielded a significant improvement in F1-score.

On the other hand, in experiment VI, we have used external unlabelled satellite road images from the *DeepGlobe Road Extraction Dataset*[3][1] to initialize our model weights. In more detail, the model has been inputted these unlabelled images and has been asked to reconstruct them in order to learn to extract image features, specially regarding the first, downsampling layers. This autoencoder structure, setting the loss as the Mean Squared Error function, has been used as an (unsupervised) initialization/pre-training of the model weights for our (supervised) main task [2] .

[1] Downloaded from
`https://www.kaggle.com/datasets/balraj98/deepglobe-road-extraction-dataset`.

[2] The dataset contains masks for the satellite images, but we have ignored them in order to keep the pre-training unsupervised. The dataset contains 8 571 satellite road images, of which only 3 000 of them have been used (the first ones in alphabetical order, from either the training or the test set) due to storage limit.

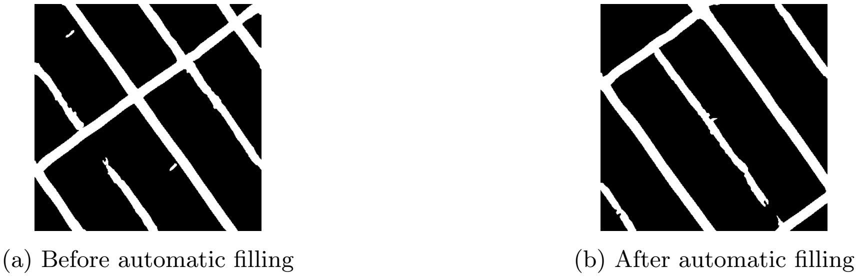
5 Post-processing

5.1 Small island deletion

Our first basic observation is that a road must be able to lead a car from a point to another and that there therefore shouldn't be small patches of road to nowhere. We can detect the small connected components of the predicted road pattern and remove the ones with pixel count lower that a certain threshold.

With a threshold of 200 pixels, we observe an improvement on the SegNet prediction data from 0.87137 to 0.87431. However, this method does not always help when applied to the U-Net and its improvements —see Table 1—.

5.2 Avenue detection and filling



(a) Before automatic filling

(b) After automatic filling

Figure 2: Avenue detection and filling.

Both networks suffer from some discontinuities in the segmentation of some roads (see Figure 2): this is most often caused by the top of trees hiding the road below and or by some shadows. In this section we have developed a new post-processing routine that detects large straight roads and fills the small holes that sometimes appear in them by accident. It runs exclusively on prediction data, and does not take into account any information from the initial satellite image.

First, we notice that there are a lot of parallel roads in the dataset, and that, for each image, the general directions of most roads can be described using one or two angles. The first step of our routine is detecting the two mains directions of the image (Figure 3). We proceed as follows, starting from the prediction image:

- Compute the gradients along x and y (using simple convolutions)
- Compute the gradient vectors at each pixel (using arctan and norm)
- Filter out the pixels with low gradient norm
- Build a histogram of the gradients and extract the two peaks

We then apply the filling procedure along each of the two main axis. It goes as follows:

- We rotate the image so that the direction of interest is horizontal
- For every line of the image, we look for the leftmost and the rightmost pixel labeled as road; if they are sufficiently further apart and if the share of road-labeled pixels between the two is high enough (more than 0.7 for the main direction or 0.8 for the second), we label every pixel in between as a road pixel
- We rotate and crop back the image to its original format

This procedure brings —minor, but noticeable— improvements to the F1-scores, as seen below.

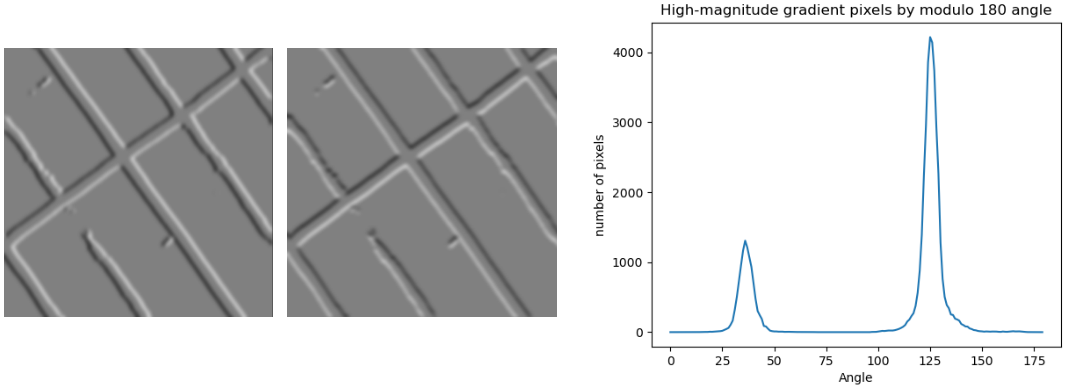


Figure 3: The gradients can be used to extract the main directions of the image.

6 Scores by method and Conclusion

Id	Experiment	No post-processing	SID	ADF	SID+ADF
I	SegNet	0.87137	0.87431	0.87435	0.87744
II	U-Net	0.87897	0.88024	0.88073	0.88202
III	U-Net + BN	0.89897	0.89815	0.90048	0.89964
IV	U-Net + BN + Dropout	0.89213	0.89260	0.89264	0.89313
V	U-Net + BN + DataAug	0.89626	0.89624	0.89688	0.89695
VI	U-Net + BN + UWI	0.90511	0.90515	0.90508	**0.90526**

Table 1: AF: avenue detection and filling, SID: small island deletion, BN: batch normalization, UWI: unsupervised weights initialization.

In Table 1 there are the F1-scores for each model, method and post-processing. One can observe that the *Avenue detection and filling* post-process generally improves the F1-score, whereas it is not clear that the *Small island deletion* post-process has a positive impact. When both methods are applied, the F1-score always increases.

Another conclusion is that U-Net performs better than SegNet regardless of the post-processes, and that batch normalization increases significantly the U-Net performance. On the other hand, dropout and data augmentation via rotation, translation, etc., do not seem to affect much to the U-Net score. Finally, the use of an autoencoder model to unsupervisedly initialize the U-Net weights has yielded the best results, regardless of the post-processing method.

To conclude, the best score has been achieved by the U-Net with batch normalization after the aforementioned autoencoder pre-training, and after applying both the *Avenue detection and filling* and the *Small island deletion* post-processing methods.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2016.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[3] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.