POVa - Road segmentation project

Tomáš Dubský <xdubsk08@stud.fit.vutbr.cz>
Marek Mudroň <xmudro04@stud.fit.vutbr.cz>
Filip Osvald <xosval04@stud.fit.vutbr.cz>

December 22, 2023

1 Introduction

The goal of this project is to train, compare and study neural networks used for segmentation of roads in satellite images.

2 Datasets and their augmentation

We have used two datasets:

- DeepGlobe Road Extraction Dataset[1] contains 7334 pairs of images - a satellite photo and a pixel-wise mask of roads. All images are in resolution of 1024×1024 pixels.
- Massachusetts Roads Dataset[2] contains 1108 of such pairs in resolution 1500×1500 pixels. To be able to combine the datasets, we have cropped the images to 1024×1024 pixels.

For those experiments that did use augmentation, we used these augmentations:

- Random flip of the image (horizontal, vertical, or both).

[1] Available at: `https://www.kaggle.com/datasets/balraj98/deepglobe-road-extraction-dataset`
[2] Available at: `https://www.cs.toronto.edu/~vmnih/data`

- Gaussian blur of the image with 5×5 kernel.
- Random change of hue, saturation and value of the image.

To suppress dependency on illumination, all satellite images, even those without the augmentation, were color normalized.

3 Neural network models and experiments with them

We have examined 5 models. The models use UNet as backbone and ResNet18 encoder. All models were trained in 5 epochs using Adam optimizer with learning rate 0.0003. Data for all models was divided in proportion: 75 % train set, 10 % validation set, and 15 % test set.

We have conducted the following experiments with the models:

- First, we tried to determine whether the number of encoder layers improves accuracy of the model. Thus, we have trained model 1 and 2 with different number of encoder layers.
- Then, we have assessed importance of transfer learning (usage of a pre-trained model). We have done this by comparing model 2 and 3 which differ in usage on pre-trained encoder.
- Finally, we have assessed importance of data augmentation by comparing models 4 and 5 which differ only in the fact that 5 was trained on augmented data while the other not.

The following table summarizes the differences among the models:

Model #	# encoder layers	Train dataset	Pretrained encoder	Data augmentation
1	3	combined	no	no
2	5	combined	no	no
3	5	combined	yes	no
4	5	DeepGlobe	yes	no
5	5	DeepGlobe	yes	yes

Figure 1 show a sample prediction of model 3.

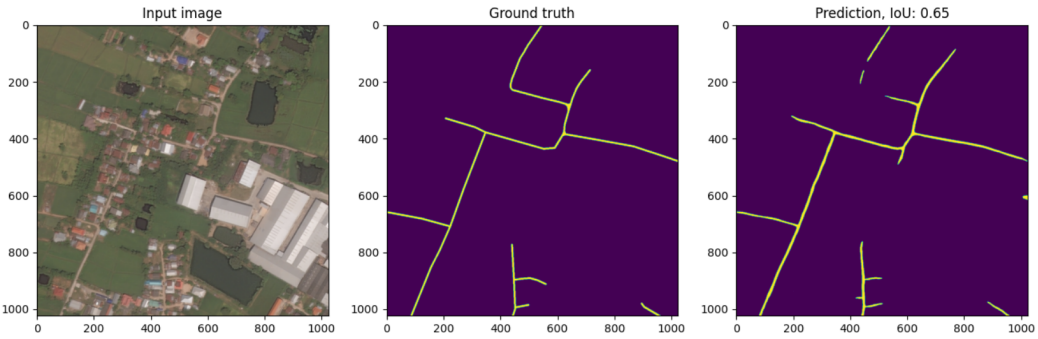

Figure 1: A sample input satellite image, ground truth mask and a prediction of model 3. It can be seen that the model misses some segments or roads that are covered by nearby trees.

4 Results of the experiments

We have compared the models by comparing their accuracy. Accuracy was measured by intersection over union metric. These are the results:

- Increasing the number of encoder layers has slightly improved the accuracy.
- Using a pre-trained model also slightly improved the accuracy.
- However, model trained on augmented data showed very little improvement over model trained on the original data.

Experiment	Models	IoU	IoU Δ
More encoder layers	$1 \rightarrow 2$	$0.449 \rightarrow 0.484$	0.035
Pre-trained model	$2 \rightarrow 3$	$0.484 \rightarrow 0.526$	0.042
Data augmentation	$4 \rightarrow 5$	$0.430 \rightarrow 0.439$	0.009

5 Conclusion

We have trained and compared several models of neural networks for road segmentation in satellite images. Usage of pre-trained model and more encoder layers has improved the model.

All source codes are publicly available at GitHub[3]

[3] Available at:`https://github.com/MarekMudron/road-segmentation`