Semantic Segmentation

Problem statement

In this analysis, two popular architectures for performing semantic segmentation on remotely sensed aerial imageries were tested with the Vaihingen datasets – the Seg-Net and the U-Net. The aim was to reconsider if the Seg-Net architecture outperforms the well-know U-Net architecture for segmenting image classes. These two networks were trained to segment roads, buildings, cars, vegetation and

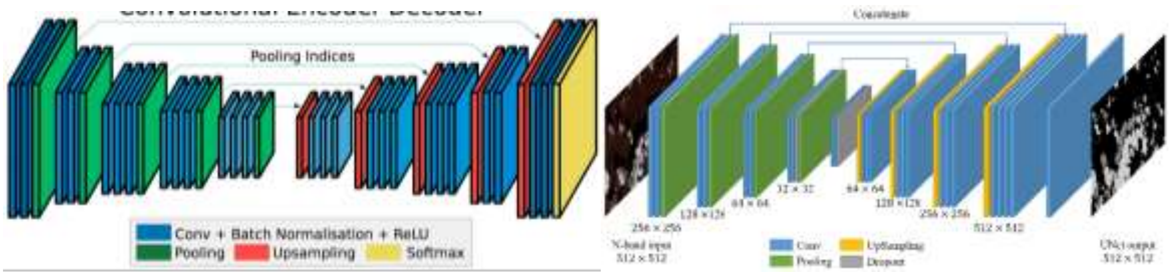

Figure 1: Model architectures of the Seg-Net (left) and the U-Net (right).

Datasets

In this analysis, the popular Vaihingen dataset used for urban semantic segmentation was utilized as the base data for comparing the two popular model architectures. The Vaihingen data is 3-channel RGB images and masks collection with unique RGB values representing various classes. The classes represented in the data include: 0. Clutter/background, 1. Impervious surfaces, 2. Building, 3. Low vegetation, 4. Tree, and 5. Cars. To speed up model training, the datasets were split into smaller patches of size 256 by 256.

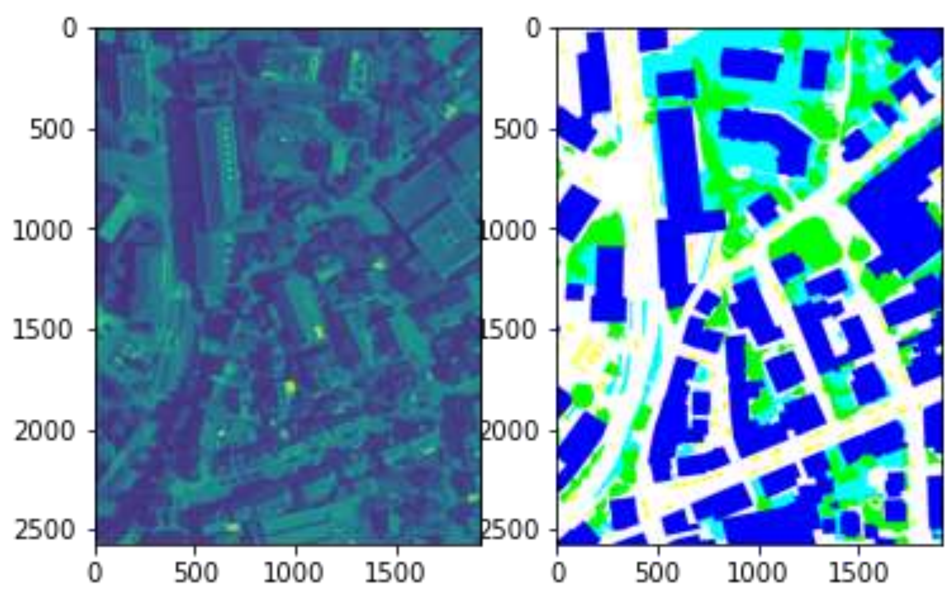

Figure 2: Vaihingen datasets – sample image and corresponding mask.

Model Design

As a basis for comparisons, the Seg-Net and U-Net architectures were tested using the VGG-16 batch norm architecture pretrained on ImageNet as their backbone for extracting features into the bottleneck. Both architectures also utilized the Stochastic Gradient Descent Optimizer for their respective training loops using the same class weights for each respective class.

Outcomes

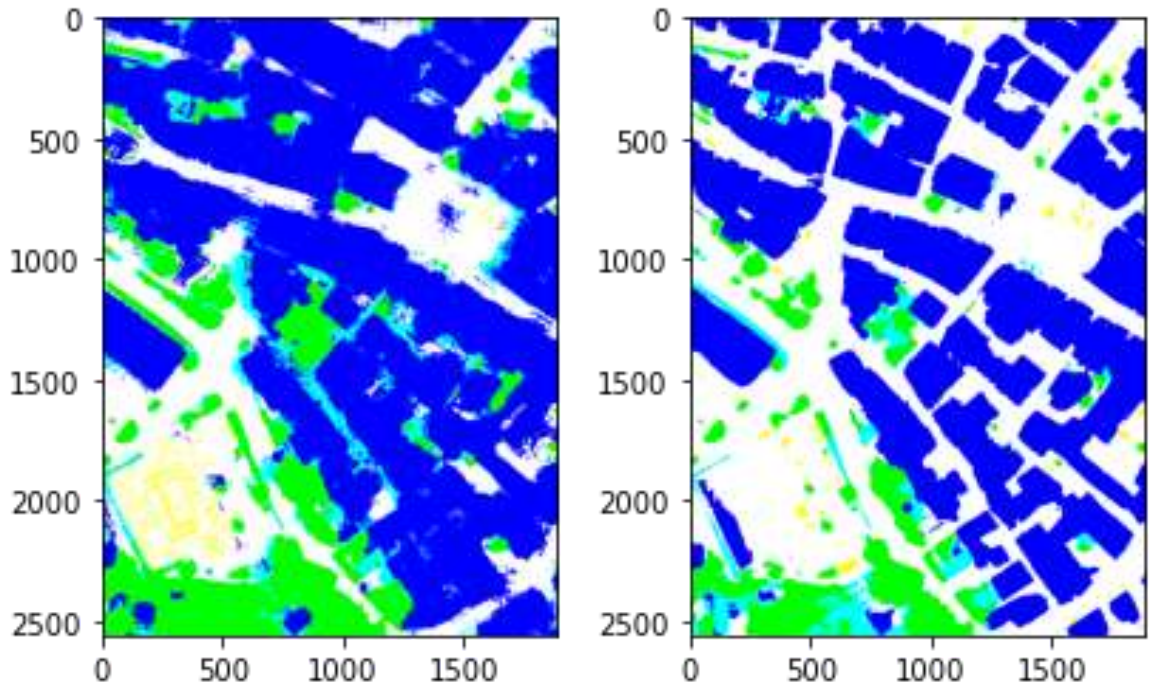

Figure 2: The outcome of the urban segmentation of the Seg-Net Model on the left, and the U-Net Model class segmentation on the right

	F1-Score	
Classes	Seg-Net	U-Net
Roads	0.7804	0.8696
Buildings	0.8547	0.9165
Low vegetation	0.5959	0.6280
Trees	0.7807	0.8340
Cars	0.0368	0.6124

Table 1: The comparative evaluation of the F1-Score of the semantic classes for both the Seg-Net model architecture and the U-Net architecture

Overall, the U-Net model architecture yielded a better model performance compared to the Seg-Net architecture with Seg-Net leading to a global accuracy of 80.21% and the U-Net a global accuracy of

87.66%. The balance between the recall and precision shown in Table 1 also alludes to the model performance corroborating the high performance of the U-Net model in discriminating the classes and with lesser False Positives and Negatives.

Other Experiments:

Apart from the experiments outlined above, other experiments were tested for urban semantic segmentation using the Seg-Net model architecture with different pretrained backbones. Amongst these, the most important ones and the corresponding challenges encountered were:

I. Urban semantic segmentation with the LandCover AI datasets provided in the Torchgeo Dataset library. Here, the Seg-Net architecture, implemented through the PyTorch Lightning framework, was developed, and tested on the 512 by 512 images and masks of the 4-classes dataset. The process ran into a CUDA Memory (which the author hasn't figured out the best way to debug this kind of error). The assumptions of the possible causes of these errors could be: 1. The chip sizes were too large to be passed into the memory at a go, 2. Due to the large number of datasets, there is a need for a higher performance hardware accelerator.

II. Urban semantic segmentation with the RIT-18 datasets. Similar issue as faced in the first failed experiments was also faced here.

References

Semantic segmentation of aerial images with deep networks

Torchgeo datasets

https://paperswithcode.com/dataset/rit-18

https://www2.isprs.org/commissions/comm2/wg4/results/