IMAGE SEGMENTATION USING U-NET FROM SCRATCH AND PRE-TRAINED MODELS

Akash Kumar Singh (2023AIY7582)

*Kshitij Kumar Verma (2023CEZ8116) helped in creating report.

Course Code: AIL861, Submitted To: Prof. Sudipan Saha

ABSTRACT

The study presents the approach to detect the vegetation cover comprising of trees and grass over the different type of land cover over Delhi, India using multispectral images. Image segmentation model U-NET is trained over the High-Resolution multispectral images of the Zurich, Switzerland. The results are compared with existing state of the art pretrained models, and the accuracy assessments are performed and validated using manually annotated ground truth masks. The results show the good accuracy of the trained model compared with the existing models on the training dataset. It is shown that multispectral images of different regions can be used for training the segmentation models and application on the different geographic regions.

Index Terms— Image Segmentation, Land Use Cover, U-NET, Deep Learning

1. INTRODUCTION

Land use and Land cover analysis is a crucial part of geospatial analysis to monitor the changes over a geographic region for various environmental monitoring and planning for future development. Vegetative cover mostly in urban areas is crucial for various regions like air quality improvement, temperature regulation, flood management and many other reason especially in densely populated city like Delhi due to mostly densely packed houses and varying land cover in short distances. Detection of trees and grass is challenging due to mostly same spectral signature of both trees and grass using the multispectral images. Using a high-resolution multispectral dataset to train the model can mitigate some of the challenges.

2. STUDY AREA AND METHODOLOGY

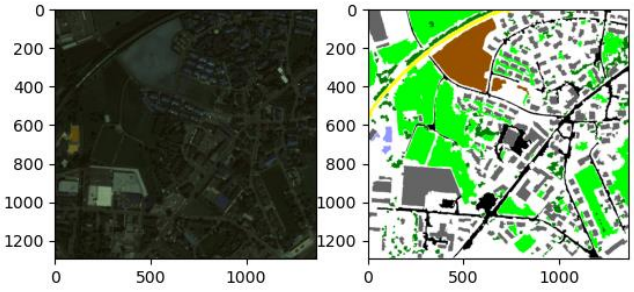
The training dataset consists of Pan Sharpened high-resolution images of QuickBird acquired over city of Zurich, Switzerland in August 2002 which consist of 4 bands including NIR Band along with the ground truth masks consisting of 8 classes including Trees and Grass. Only three classes are considered out of nine classes namely others, trees and grass. The training and test dataset images are loaded in RGB Bands and normalized using the MinMax Scaler in the range 0 to 1.

2.1. Training and Test Dataset

2.1.1. Zurich Dataset

For training we have used image number 1 to 5 and for test set image number 16 to 20.

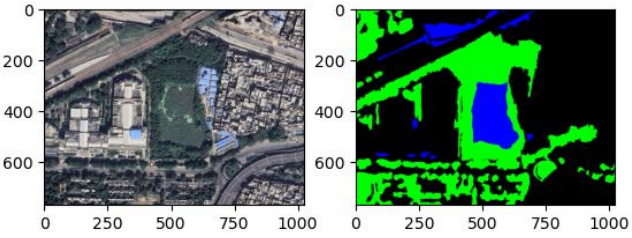
Each of the training images is divided into 224*224*3 overlapping patches to create the final training images set. Later the training set is divided into training and validation with the ratio of 90:10. For the training masks, RGB values are converted into the class label numbers.



Train Set

2.1.2. Delhi Dataset

The test dataset is consisting of google earth images over the city of Delhi, India with the scale of 1 in 100m. The ground truth mask by hand annotation is generated consisting of two classes of Trees [0,255,0] and Grass [0,0,255] for the test datasets.



Test Set

2.2. Model

The weights of two pre-trained models namely Resnet34, Seresnet34 are used as initial weights to train the UNet models instead of random-initialization . Further, the U-NET model is trained from scratch on the training data. For the scratch U-NET, down sampling of 16,32,64,128,256 and BatchNormalization along with dropout is used.

2.3. Loss Functions and Metrics

Class weightage is assigned to each of the classes as 0.1,0.5,0.4 for the subsequent classes. Because of the unbalanced distribution of classes and most focus on the two classes, the total loss is defined as sum of dice loss (with class weights) and focal loss. For the performance metric Mean IOU score (with class weightage) is used and the accuracy but the best model is selected based on the validation IOU metric.

2.4. Training Parameters

For pre-trained models, Adam optimizer with learning rate 0.0001 is used in the study. Model from scratch is trained using the 0.001 as the learning rate.
Each model is trained on 300 Epochs. Batch size of 16 is used throughout the process.

2.4. Test Set Prediction

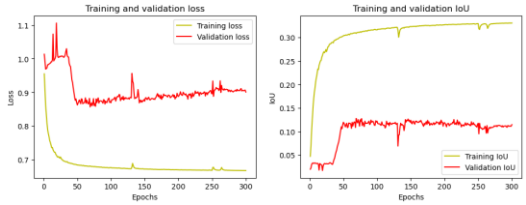
Each non-overlapping patches is extracted from the image (only the last patch is overlapping if image size is not multiples of 224), the predictions are made using the best model and stitch each of the predicted mask to get final image mask. Then finally Mean IOU score is computed using the Ground Truth mask and again converted back the class label to RGB and shown the plot in each of the model. For Delhi data we used the ensemble of these three trained models on Zurich dataset for prediction. However, we got an unsatisfactory score of 15% Mean IOU.

3. RESULTS & DISCUSSIONS

The Mean IOU Score obtained on the test set on Zurich and Delhi DataSet using three models used are tabulated below.

Zurich Data	
Models	**Mean IOU**
Resnet34	0.486
SEResnet34	0.542
UNet (scratch)	**0.532**
Delhi Data	
Ensemble	0.153

The Training and Validation curve using Resnet34 on Zurich Dataset is shown below.



Below are the segmented masks generated by respective models and their corresponding ground truth mask lables.

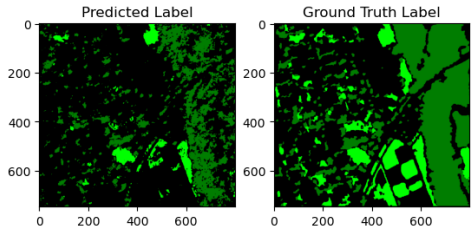


Figure 1: Resnet34 (Zurich Data)

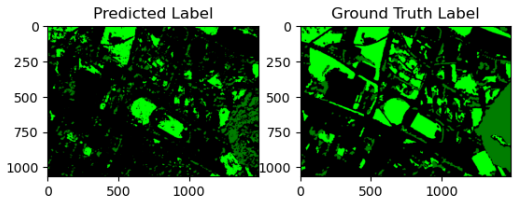


Figure 2: Seresnet34 (Zurich Data)

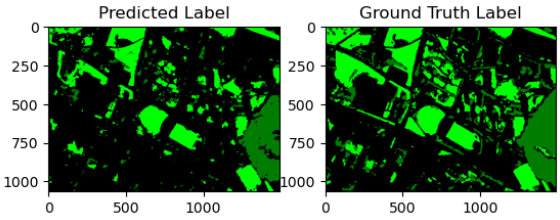


Figure 3: U-Net (Scratch)

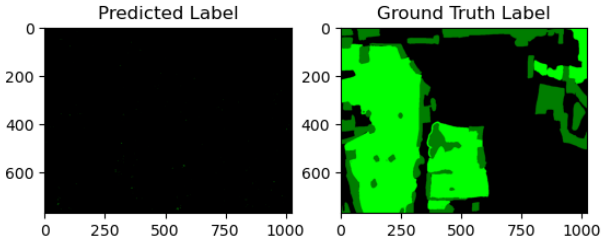


Figure 4: Ensemble Model (Delhi Data)

From these results we can see that model trained from scratch has achieved better performance than most of the models trained using initial weights of the pre-trained networks.