Aerial Image Segmentation using the U-Net Deep Learning Model

Sibusiso Mgidi Thabo Rachidi

University of the Witwatersrand, Johannesburg, School of Computer Science and Applied Mathematics

Introduction

Motivation

- Experiment with a U-Net convolutional network to determine its viability as a method to perform semantic segmentation on aerial images.

Problem

- The automatic interpretation of aerial images as a pixel-wise labelling task.

Goal

- Construct a U-Net convolutional network that can detect buildings in an aerial image.
- Deploy a proof of concept website that uses the best models for prediction.

Dataset and Preprocessing

To conduct the semantic segmentation study, we used the Messachussets building dataset from Kaggle [1].

- Dataset consists of 151 RGB images with corresponding grayscale masks.
- RGB images and masks are down-scaled to 512×512 and 256×256 dimensions.
- Training, validation, and test set split is 131, 10, 10 respectively.
- Applied horizontal and vertical flip data augmentation.

Research Design

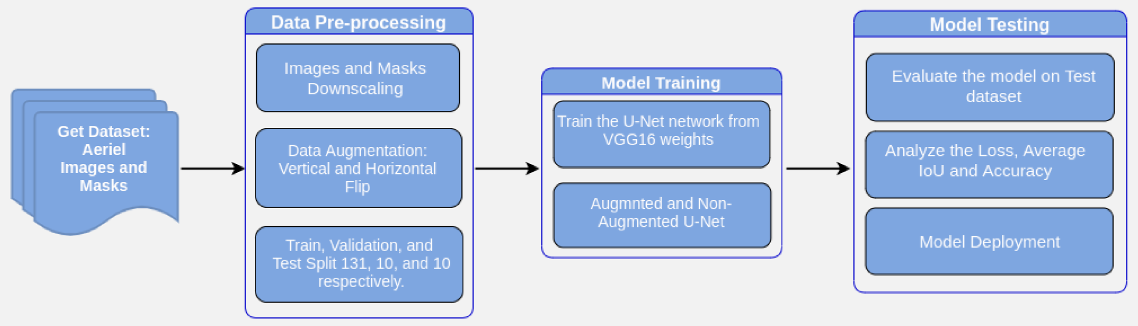

Figure: Research Analytic Pipeline

Evaluation Metrics

- Monitor the train and validation loss. Also, compute the average IoU for test ground truth and predicted masks.

Hyper-Parameters

- Optimizer: Adam
- Loss function: binary cross entropy
- Batch size: 6
- Number of epochs: 120
- Early stoppers: 30
- Learning rate: 0.001
- Dropout rate: 0.5

Network Architecture

- Contracting path to capture image context.
- Symmetric expansion path to which up-sample the images for precision localization.

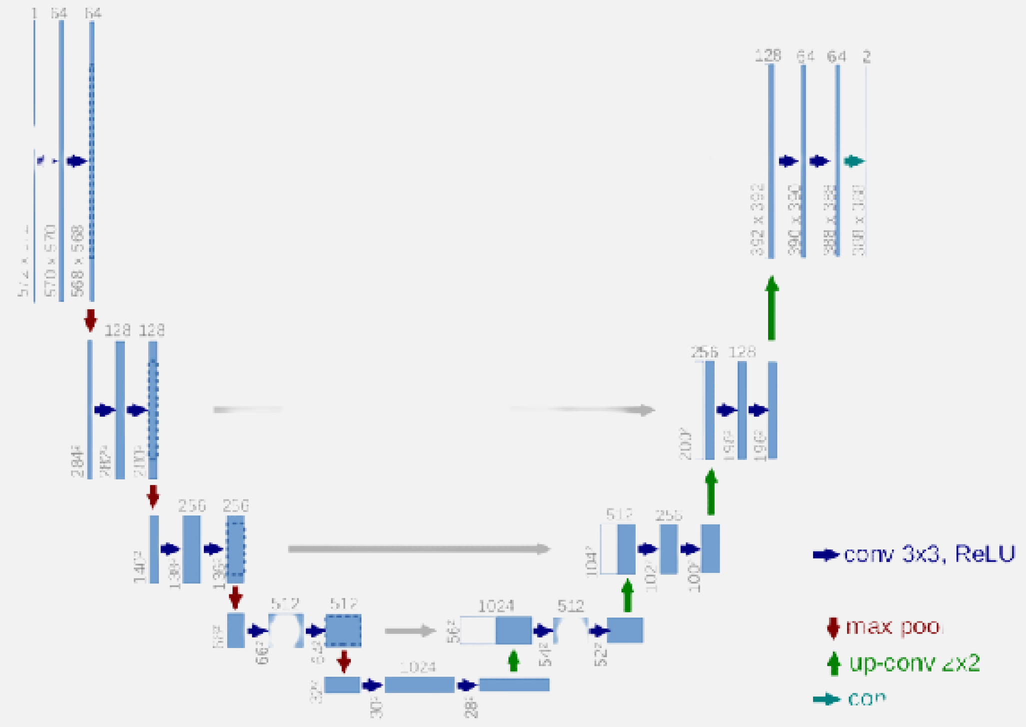

Figure: U-Net Network Architecture

Network Modification

- Removed the 1024 convolutional layer since the dimensions of our images and masks are 512×512 and 256×256.
- Loaded VGG16 pre-trained weights on the encoding side.
- Added dropout between convolutional steps to prevent overfitting.

Results

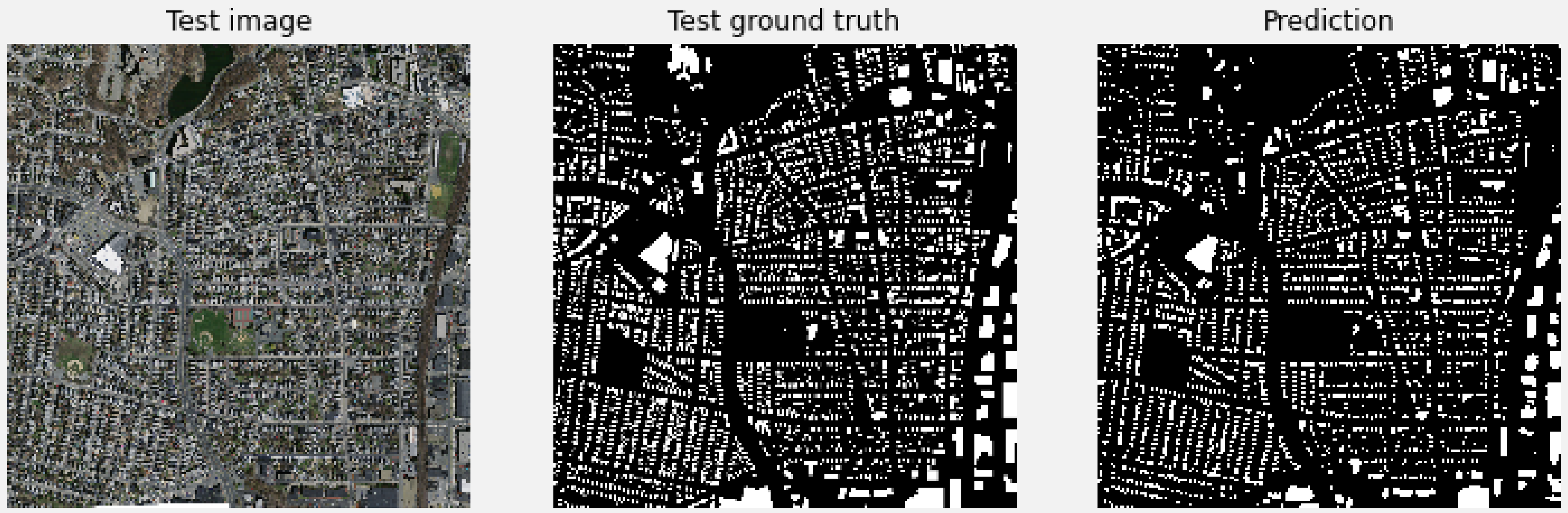

Figure: Original Image, Ground Truth and Predicted Output of U-Net Traine dwith Augmenttaion: IoU = 0.51

Results Comparison

Model	Image Size	average IoU	Val Loss
U-Net	256×256	0.33	0.19
	512×512	0.48	0.15
U-Net(Augment)	256×256	0.38	0.17
	512×512	0.47	0.16

Table: U-Net compared to augmented U-Net results.

Conclusion and Future Work

- Augmented U-Net outperforms the U-Net on some individual images.
- U-Net gives better average IoU compared to augmented U-Net.
- We achieved an average IoU of 0.48 on U-Net and 0.47 on augmented U-Net trained with 512×512 images.
- We also deployed the deep learning segmentation model to enable others to use them.
- In the next step, we will explore other data augmentation techniques such as increasing/decreasing the image contrast, zooming in or zooming out, e.t.c
- Explore other segmentation models such as DeepLab, Segnets, Mask RCNN, Fully Convolutional Neural Network, e.t.c.
- Train the U-Net model from scratch and compare it to our results.

References

[1] Volodymyr Mnih.
Machine Learning for Aerial Image Labeling.
PhD thesis, University of Toronto, 2013.