Buildings Segmentation using Semantic Segmentation

Sibusiso Mgidi
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Student no: 1141573

Thabo Rachidi
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Student no: 1632496

***Abstract*—**

***Index Terms*—semantic segmentation,**

1. Introduction

2. Background and Related Work

Semantic segmentation is a process of matching each pixel in an image to a class label. It is used in cases where a thorough understanding of images is needed such as diagnosing medical conditions, navigation for self-driving cars, photo and video editing and navigation for robots.

To know how deep learning is used to tackle semantic segmentation we need to understand that its not just an isolated field of machine learning but a step in a larger process of object detection. It helps with the progression from a normal image to detecting just not what an object is but how many objects there are in the image. The natural progression is image classification, object localisation, semantic segmentation and instance segmentation.

Early work of aerial image labeling focused on ad-hoc and knowledge-based approaches [1] but this paper will focus on approaches that have used machine learning. Machine learning has led to more recent progression when interpreting aerial images and in general computer vision problems such as labelling in images using semantic segmentation.

3. Research Methodology

3.1. Research design

3.1.1. Source Dataset.

We used the Massachusetts buildings dataset from Kaggle[1]. The dataset consists of 151 aerial images of the Boston area. The images cover an area of 2.25 square kilometres, so the entire dataset covers roughly 340 square kilometres. The images are of high quality with dimensions of 1500 x 1500 pixels. The data is already split into a training set of 137 images, a testing set of 10 images and a validation set of 4 images. Each image has a corresponding mask that are used as targets. These target maps were obtained by rasterising building footprints obtained from the OpenStreetMap project.The dataset covers urban and suburban areas with buildings of all sizes.

1. https://www.kaggle.com/balraj98/massachusetts-buildings-dataset

3.1.2. Data Pre-processing.

To reduce computation time and memory usage we down scaled the 1500 x 1500 pixel images to 1024x1024 and later on to 512 x 512 pixels.

3.2. Dataset split and data augmentation

We reduced the training set to 131 images and transferred them to the validation set. This gave us a total of 10 images for the validation set. This was done to help reduce the training time and the amount of memory needed to store the images.
To help increasing the training data we performed data augmentation of the images by flipping the images horizontal and vertical, and changing the contrast of the images.

3.3. Training the Model

Due to insufficient memory on our local machines we used Google Colaboratory which gives us 16Gb of memory to train our model.
We used the UNet architecture that modifies and extends a Fully Convolutional Network(FCN). It is a U-shaped architecture that uses a encoder-decoder schem. It consists of three sections a downsampling, bottleneck and upsampling section.

epochs, batch sizes.

4. Results and Discussion

4.1. Evaluation Metrics

Precision.

Confusion Matrix.

Accuracy.

4.2. Experimental Results

4.3. Discussion

5. Conclusions and Future Work

References

[1] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, University of Toronto, 2013.