Semantic Segmentation on Aerial Drone Images using U-Net

Star Chen, Muyang Xu, Chengyu Zhang

April 2021

1 Introduction

Aerial Drone, or unmanned aerial vehicle (UAV), which came in as crucial military assets for scouting, has recently gained an increased popularity in civil use. One key feature of an aerial drone is its ability to produce high quality nadir view images and videos through the camera attached to the drone. To further simplify the drone operation and improve customers experience, we deem that further automation is needed. That is because that object detection in aerial drone images is vital for both drone piloting (e.g., identifying and avoiding obstacles during flight) and image capturing (e.g., identifying key objects and locking camera views during flight).

Correspondingly, we intend to implement the image semantic segmentation, where given images, the algorithm is able to identify different objects in those images and apply colored masks to objects. In this way, the previously proposed improvements in drone automation could be achieved. While image semantic segmentation is relatively ripe, most are implemented for daily scenes shot in usual view directions and few are on the top-down birds eye view images. Therefore, in this project, image semantic segmentation, specifically U-Net, would be applied to the aerial drone images to get an ideal object detection in the nadir view images.

2 Objectives

This project will use the “Semantic Drone Dataset” found on the Institute of Computer Graphics and Vision for the dataset. The dataset consists of 400 images of different houses from nadir (bird’s eye) view acquired at an altitude of 5 to 30 meters above the ground[3]. It is created initially to understand urban scenes for increasing the safety of autonomous drone flight and landing procedures[3]. Since our project intends to do object detection for drone images in civil use, mainly in neighborhoods, this dataset is ideal for achieving such a task. The dataset is detailed into three parts[3]:Original full-color images of high resolution ($6000*4000$ pixels), RGB masks for each of the images, and a CSV file containing the labels' information each of the RGB colors represents.

In terms of pre-processing stage, a mask of an image must contain the class's code for each object, and each pixel in the image should be represented with a class number in the mask. So, our data pre-processing would mainly involve the transformation of the RGB masks into gray-scale masks. Since the 400 image dataset is not divided into training and testing sets, we manually split the dataset with the initial ratio being $0.8 : 0.1 : 0.1$ ($train : val : test$), subject to be tuned. While we are not sure if the 400 images would suffice to generate an ideal result, we may well apply k-fold cross-validation.

Our image semantic segmentation approach would mainly follow the U-Net architecture proposed in "U-Net: Convolutional Networks for Biomedical Image Segmentation".[1] The reason to choose U-Net is based on its claimed high performance upon limited datasets. Also, as discussed before, the aerial drone images are shot from a bird's eye view, different from what's usually seen. While U-Net is originally applied on the biomedical image shot from the microscope, we could draw an analogy. Both image sets are a top-down view of the objects, which might make U-Net more useful in the aerial image segmentation task.

The evaluation, as discussed, would be based on the previously split test set. Strategies are mentioned above. Moreover, we would like to generate our testing data with other high-resolution drone footage found on the Internet or shot by ourselves. The tool to manually apply an RGB mask would be the Semantic Segmentation Editor on GitHub. [2]Through this way, we hope that we could test the generality of our approach and get ourselves the experience and knowledge of generating our own dataset.

Through this project, we hope to: generate an ideal result (high accuracy and robustness) towards the semantic segmentation in aerial drone images, have an understanding of the concepts and application of image semantic segmentation, as well as the structure of a semantic segmentation algorithm, the experience of generating dataset and lastly, an end-to-end machine learning project experience.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation* (2015), arXiv:1505.04597 [cs.CV].

[2] Evaluation Data Generator: Semantic Segmentation Editor https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor.

[3] Semantic Drone Dataset https://www.tugraz.at/index.php?id=22387 *Alternative Dataset Resource: Aerial Semantic Segmentation Drone Dataset* https://www.kaggle.com/bulentsiyah/semantic-drone-dataset.