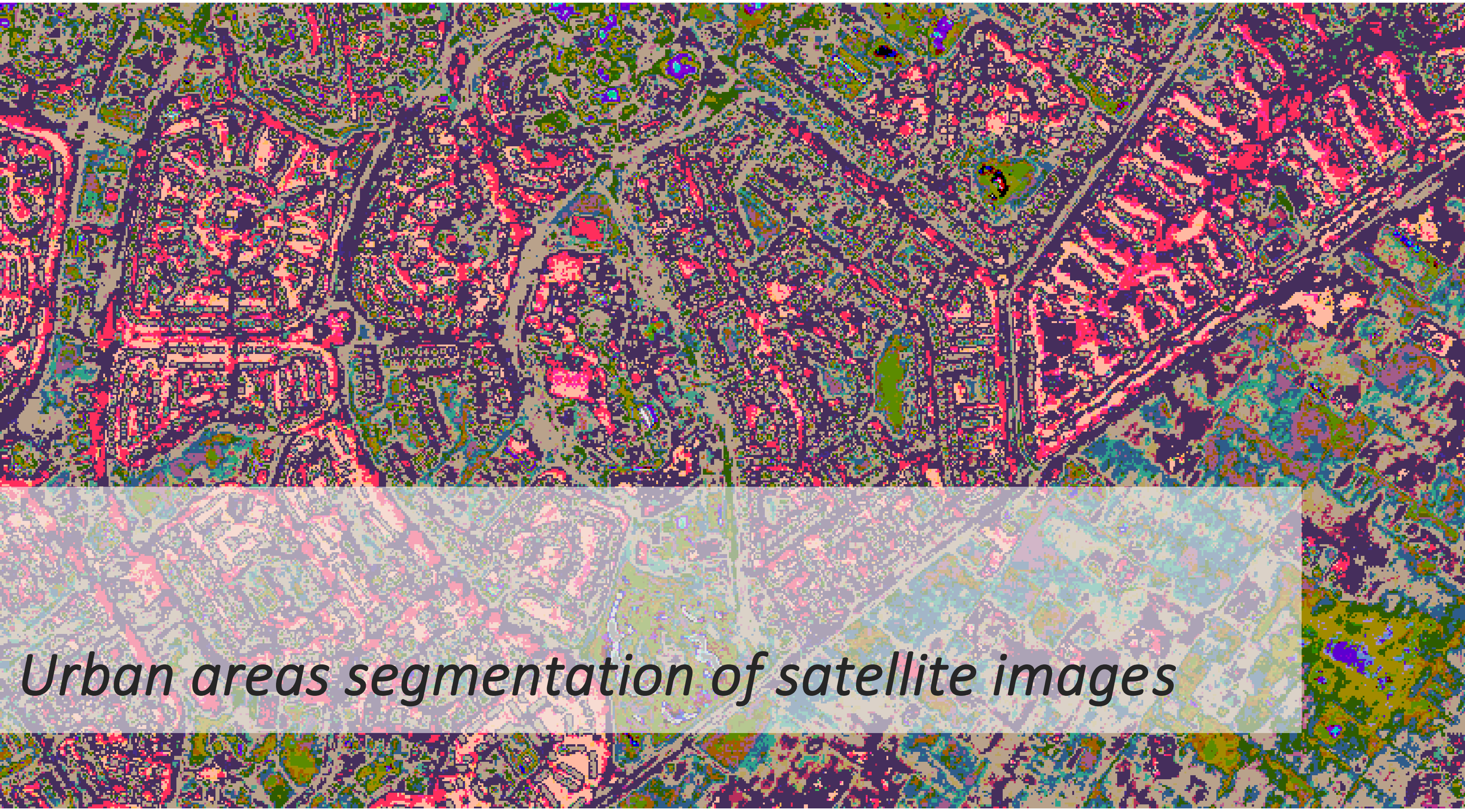

Urban areas segmentation of satellite images

This report has been prepared by Jorge Martinez Navarrete, Information Technology Officer at UN office of Information and Communications Technology, and Gad Mohamed Gad, Technology Trainee at the ESCWA Technology Center.

15/1/2021

Table of Contents

Abstract

Automatic detection of urban areas footprints from high-resolution satellite imagery is an important topic that is receiving greater attention in applications like urban development planning. In this paper, I will present an end-to-end method of urban-area segmentation starting from acquiring and preprocessing satellite images, combining unsupervised learning (clustering) with manual labeling to generate a labeled dataset of urban area maps of the acquired images, finally, building and training a binary segmentation model to predict urban area maps of new satellite images.

Introduction

Satellite imagery analysis has gained a lot of interest in the last decade due to the availability of near-real-time high-resolution satellite imagery sources and the rapid development of image processing techniques. Most approaches addressing this problem use deep learning-based algorithms which has obtained state-of-art results compared to traditional image processing.

The problem of extracting land use and land cover information (e.g., buildings, roads, farms, forests, etc.) is often modeled by machine learning as a single or multi-class segmentation problem.

The goal of this paper is to explain the methodology of building a machine learning algorithm that segments satellite images into predefined classes of land use. What characterizes this work are the following:

This work only uses raw satellite data and assumes no labeled dataset for the task at hand is available. This is useful for rare or custom scenarios where no labeled dataset is available. A method for effective semi-auto labeling of raw satellite images is presented as part of the analysis.

The analysis performed in this work is limited only to public data and tools. Thus, the whole pipeline can be replicated exactly as described.

The code of this analysis is open-sourced [1].

Methodology

Data gathering

The satellite image used in this analysis is a 4-band multispectral orthorectified image of Sheikh Zayed city, Giza, Egypt acquired by PlanetScope mission [2] and downloaded from Planet [3]. The four bands of the images are:

1- Band 1 = Blue
2- Band 2 = Green
3- Band 3 = Red
4- Band 4 = Near-infrared

Compositing the color bands (first three bands) in an RGB image yields a natural view, as shown in Fig. 1.

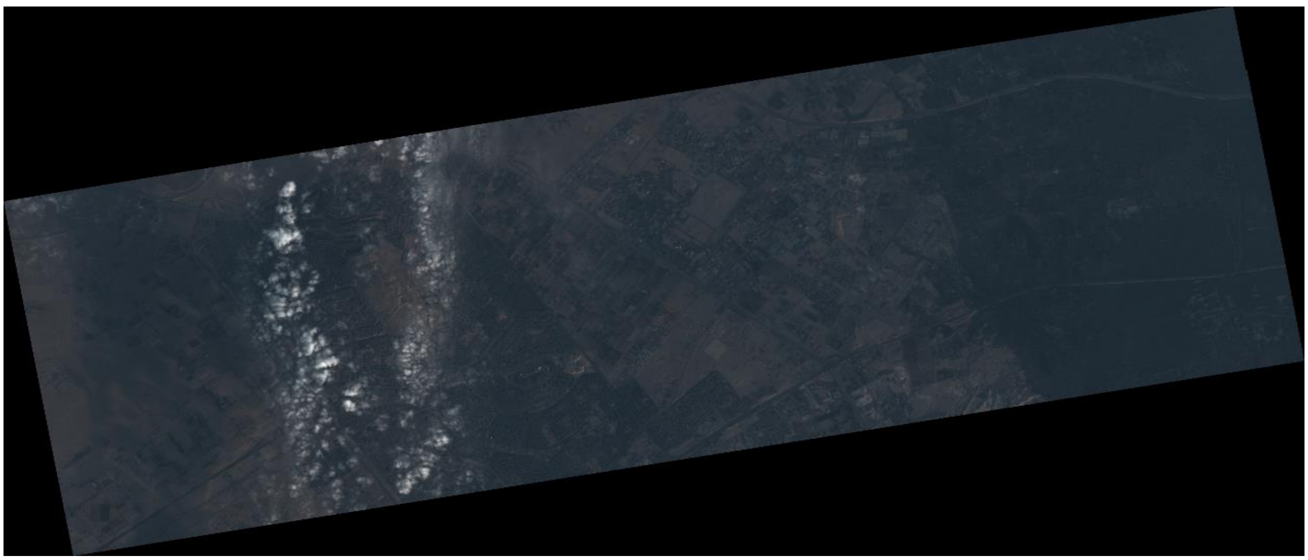

Figure 1image of Sheikh Zayed city, Giza, Egypt acquired by PlanetScope and downloaded from Planet.com

Alongside with the image, there are usable data masks (UDM), PlanetScope has two versions of it:

- UDM a simple 1-band usable data mask
- UDM2 a detailed 8-band usable data mask.

The 8 bands of UDM2 are:

- Band 1 = clear map mask
- Band 2 = snow map mask
- Band 3 = shadow map mask
- Band 4 = light haze map mask

- Band 5 = heavy haze map mask
- Band 6 = cloud map mask
- Band 7 = confidence map mask
- Band 8 = unusable map mask (equivalent to UDM)

In this analysis, only the 1^{st} UDM2 band will be considered since it contains clear vs unclear pixels. The 1^{st} UDM2 band is shown in Fig. 2.

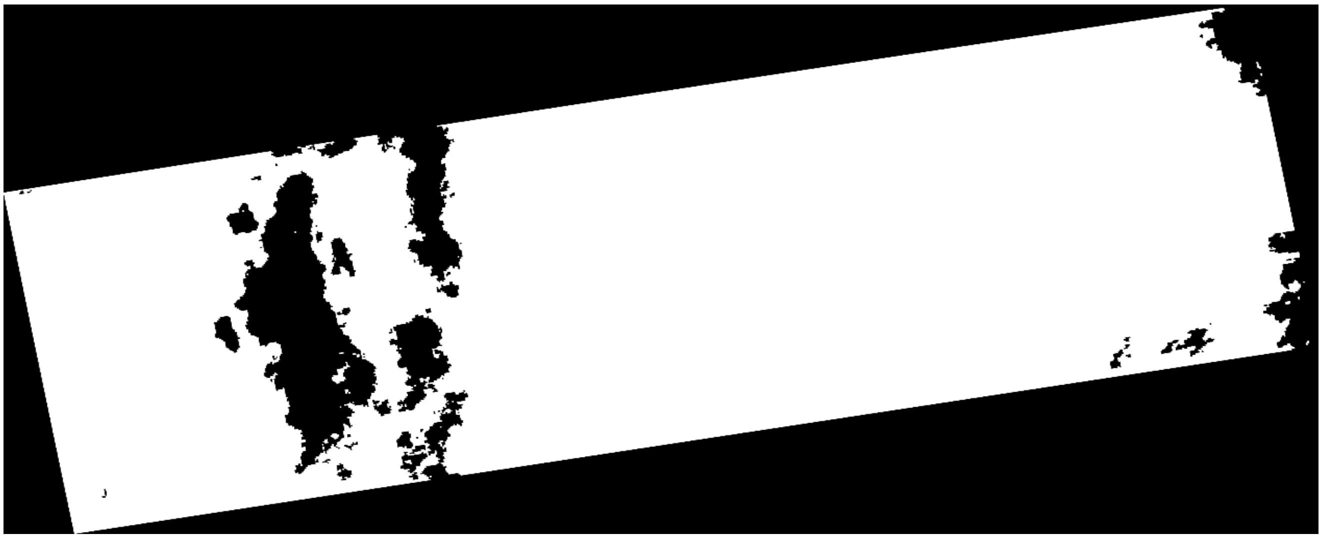

Figure 2 The 1st UDM2 band of the downloaded satellite image

Data preprocessing

Applying the 1^{st} UDM2 band mask on the image to eliminate unclear pixels so that our analysis is performed on clear, meaningful pixels, as in Fig. 3.

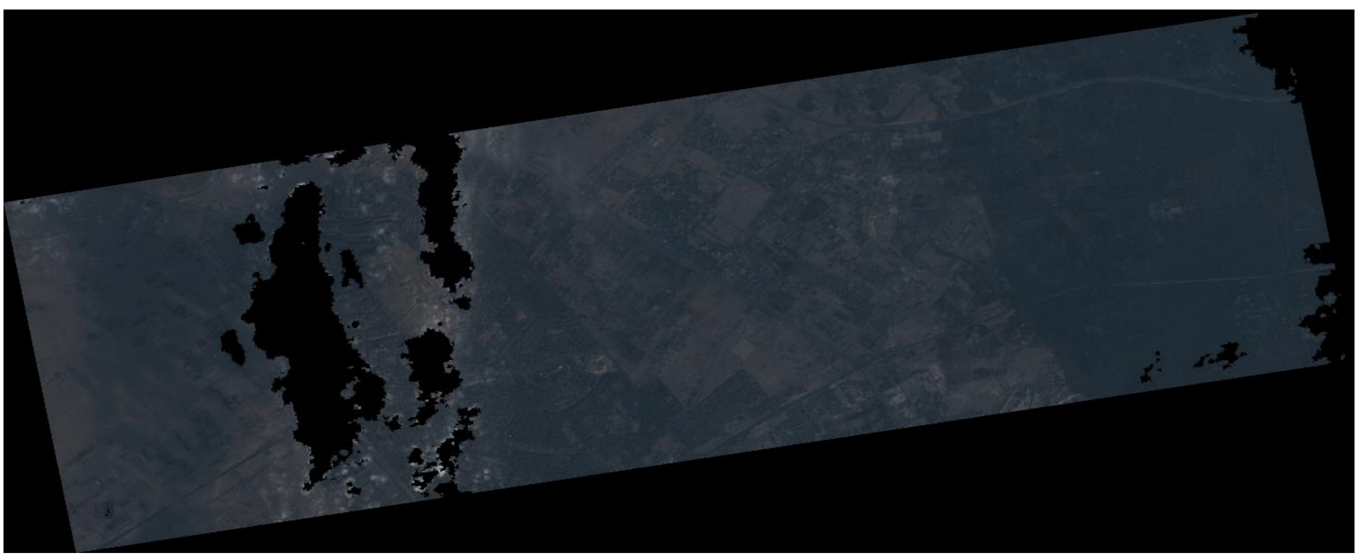

Figure 3 The resulting image from masking out the unclear pixels from the satellite image

Annotating data

The approach used for annotating data is a combination of unsupervised learning and manual labeling

i. Unsupervised learning

Unsupervised learning refers to the set of algorithms that can use the properties of the data to perform operations like clustering, without guidance. On popular unsupervised algorithm is the K-mean clustering algorithm which produce takes a set of features (e.g., image) and several clusters K, and returns the given image clustered into K clusters. In this step of the analysis, K-mean is applied on the 3-band composed satellite image and K is chosen to be a large number (as large as the computational power can handle since K-mean is quite computationally expensive algorithm). Figure 5 shows the output of applying K-mean with K set to 12.

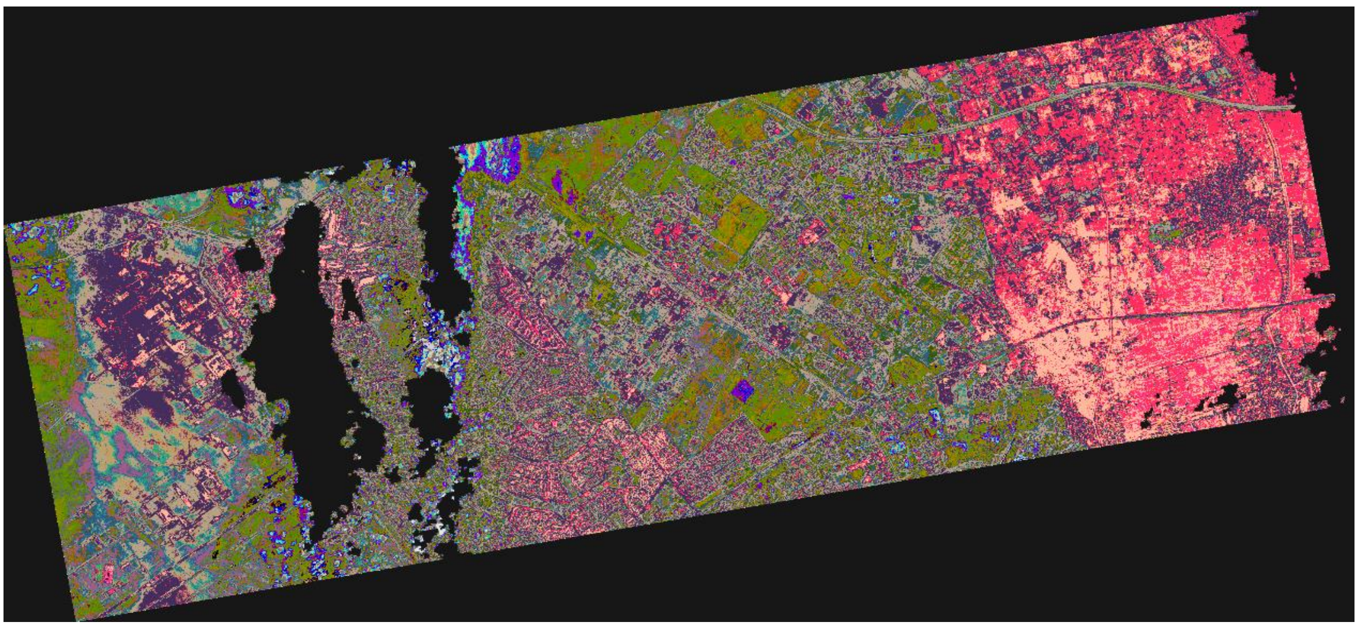

Figure 4 The output from applying K-mean clustering to the satellite image with k set to 12

ii. Manual labeling

Figure 6 shows the satellite image after applying histogram equalization (for visual clearness) and red dots drawn manually on urban areas. by masking the areas in the K-mean map where the red dots and extracting the most frequent pixel value assigned by the K-mean algorithm to those areas, we can know the color (RGB) that K-mean assigns to urban areas, if we masked that color in the whole K-mean map, we will get an urban binary mask for the whole image, as in Fig. 7.

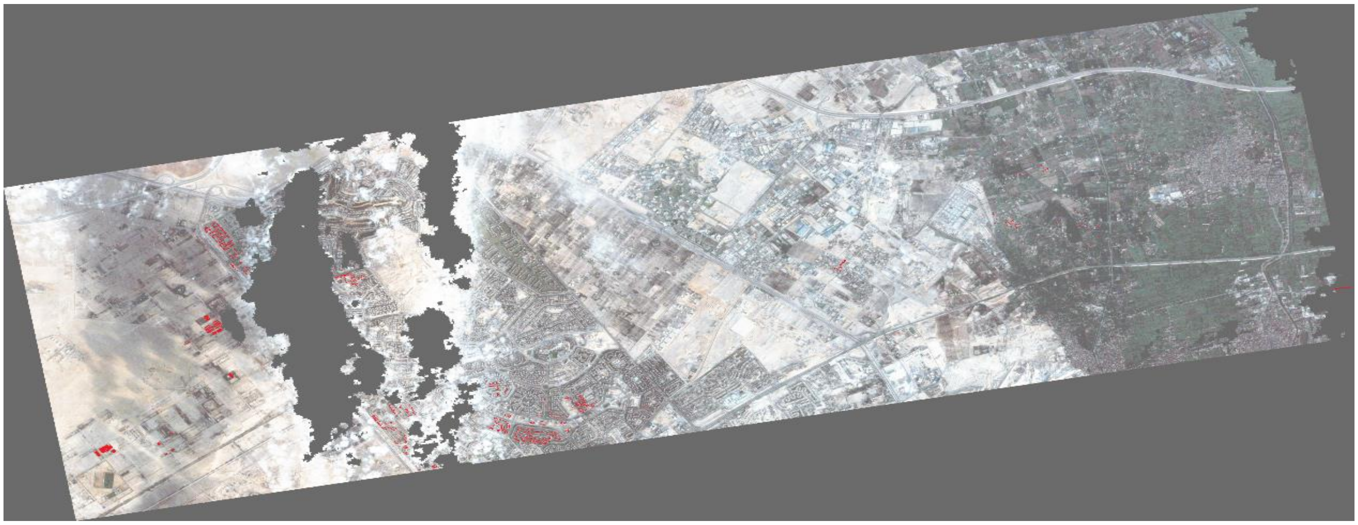

Figure 5 Manually annotated image (partially) that will be combined with the K-mean map to generate the urban area map

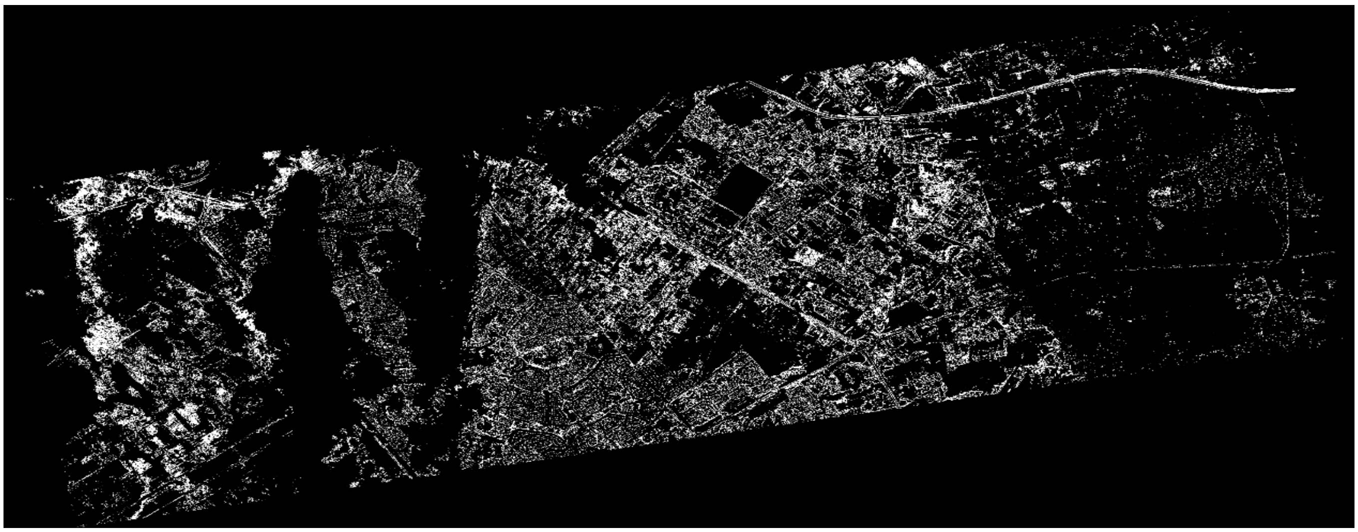

Figure 6 The generated urban area mask

Supervised learning

Supervised learning refers to the set of algorithms that use examples of inputs and their corresponding output to learn how to predict a similar output for a similar input. In the "Annotation data" section, unsupervised learning combined with manual labeling were used to generate an urban area binary segmentation map. In this section, an encoder-decoder architecture called SegNet will be trained on the Satellite image in Fig. 1 (as input) and the urban area binary segmentation map in Fig. 6 (as output).

i. Splitting data

Regardless of the size of the dataset available, it should be split between training and validation data where the former is used to train the model and the latter is used to evaluate the model. Both the training and validation datasets should be representative of the data. In our case, all our data is only a single image, which contains part of Sheikh Zayed city (which has a high living standard), an industrial area (which contains unique shaped warehouses), and a high-density residential area. To make sure all these areas are represented in both training and validation sets, the large image is firstly split into 10 horizontal equal-sized strips, then the strips are shuffled, and each of them is assigned to one of the two sets depending on the splitting ratio. For example, in this analysis an 0.2 splitting ratio is used, so 2 strips are denoted as the validation set and 8 strips as the training set. Finally, validation set strips are concatenated and training set strips are concatenated to have a single image for each set.

ii. Augmenting data

Augmentation refers to the set of transformation operations applied on the dataset (training & validation) to increase its size and improve the model's generalization by preventing it from overfitting (remembering but not learning). Augmentation includes many operations, but not all of them can be applied to every application. Fortunately, for satellite images a lot of augmentation can be performed. We applied operations like rotation, width & height shifting, shearing, zooming, and horizontal & vertical flipping.

iii. Building and training the model

SegNet [4] model is set to take input size of (96, 96, 3) and an output size of (96, 96, #classes) where #classes in our case is 1 (because we detect one class only) as shown in Fig. 7.

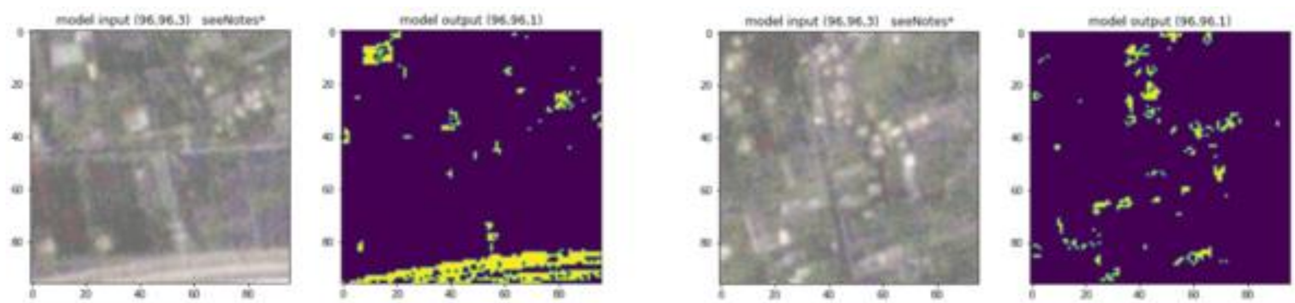

Figure 7 two samples of inputs images to SegNet and their corresponding labels

References

1- Gad, G., 2021. *Gadm21/Satellite-Imagery-Analysis*. [online] GitHub. Available at: <https://github.com/gadm21/satellite-imagery-analysis> [Accessed 15 January 2021].

2- Earth.esa.int. 2021. *Planetscope - Earth Online*. [online] Available at: <https://earth.esa.int/eogateway/missions/planetscope> [Accessed 15 January 2021].

3- Planet. 2021. *Daily Satellite Imagery And Insights*. [online] Available at: <https://www.planet.com/> [Accessed 15 January 2021].

4- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481-2495.