Dataset for Satellite Object Detection (SOD)

Peng Hu
Advanced Network and Embedded Systems Lab (AEL)
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Canada R3T 5V6
Email: peng.hu@umanitoba.ca

1. Introduction

The satellite object detection (SOD) dataset offers the first comprehensive set of synthetic images of small satellites, aiming to facilitate AI and machine learning (ML) solutions to spacecraft collision avoidance and mitigation for space safety, space sustainability, space situational awareness, etc.

2. Data Generation

To support the SOD task, it is important to have a dedicated dataset in a realistic environment. We first tried the compositing approach to generate the datasets where the Earth's background and satellite images programmed to change the orientations are combined. However, we found that the composite images cannot reflect realistic environments and effects, such as stars, lighting, satellite movements, and other celestial objects, so we decided to build a realistic 3D scenario to generate images with the publicly available satellite models from NASA's Open Data Portal (https://data.nasa.gov). We selected three satellites, i.e., Aqua, Cloudsat, and AcrimSat, to be used in the 3D scenario as their sizes are representative of the dimensions of LEO satellites with solar array designs. LEO satellites may have a large span of dimensions when considering the OneWeb satellites and Starlink satellites V1 to V2, ranging from 260 kg to 1250 kg.

To consider the background noise and conditions in a realistic scenario, we use the solar system environment and the star map in galactic coordinates from NASA (https://svs.gsfc.nasa.gov/3895). Each image generated has an annotation file containing the bounding boxes of three satellite objects. We installed a camera on a satellite, oriented to face forward. The approximate distance between the camera and the satellite is denoted by k. The satellite can appear as a scaled-up version in the camera's field of view, which can be achieved by adjusting camera settings such as the focal length. Based on different values of k, five sets of images are generated to represent the typical proximity scenarios between satellites, where $k = \{8, 14, 20, 26, 33\}$ km. Each set contains 50 images, with a total of 250 images. Fig. 1 shows some sample images when $k = \{8, 20, 33\}$ km in different illumination and background conditions. The field of view (FoV) of the virtual camera is 81.3°. Each image has a matched label file containing the coordinates of up to three satellite objects in view. For example, the label file for the image `450.jpg` is `450.txt`.

The sample images from the dataset we generated can be seen in Fig. 1, and Fig. 2.

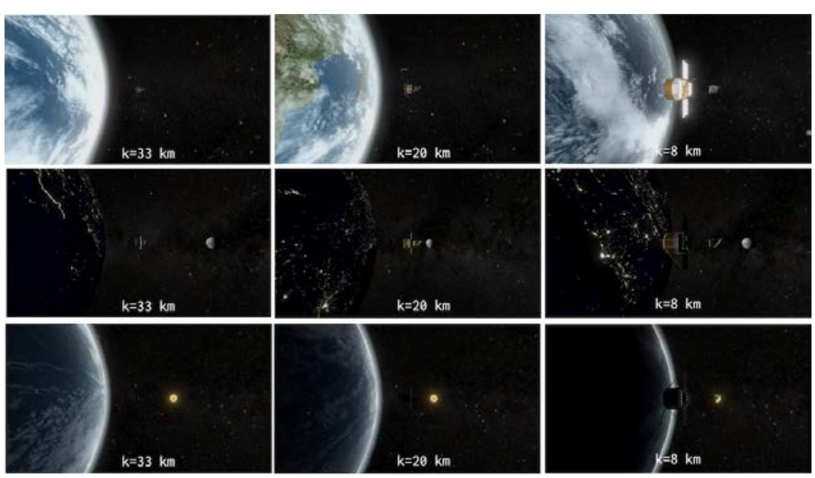

Figure 1. Sample camera views of LEO satellites in different illumination and background conditions with different values of k.

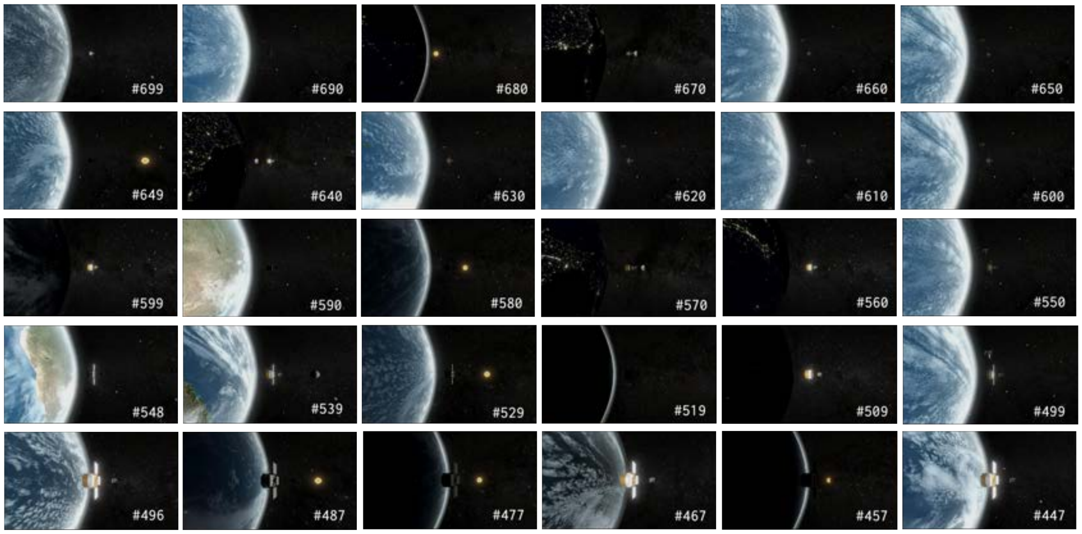

Figure 2. Additional sample images in different illumination and background conditions labeled with the image index

3. Dataset Structure

Each image in the dataset is in the JPEG format and has a size of 1920 $\times$ 1080 pixels. The filenames are based on the image indices. The relationship between image file indices and the value of k is shown in Table 1.

TABLE 1. IMAGE INDICES AND VALUE OF k

Image File Index	k (km)
447–496	8
499–548	14
550–599	20
600–649	26
650–699	33

Figure 3. Dataset directory

As shown in Fig. 3, the `images` folder has two subfolders `train` and `val`, which are the default training and validation sets. The training set has 200 images, while the validation set has 50 images. The labels of these images are in the `labels` folder.

In Fig. 3, the `train.txt` and `val.txt` also show the indices of the default training and validation images.

4. Data Repository

The data repository can be found at https://github.com/AEL-Lab/satellite-object-detection-dataset.git, published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

5. Future Work

This dataset can support research in SOD for space sustainability, space safety, and space situational awareness; however, we recognize that there is still room for improvement. In the future, we plan to enhance the dataset in various aspects, including illumination conditions such as sun glare and camera settings related to the distance k.

Acknowledgments

Zelin Han provided important contribution to the creation of the dataset with software programs that incorporates satellite and solar system models.

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2022-03364]. We would like to thank the support from the National Research Council Canada's Digital Technologies Research Center.