

Mafat (DDR&D) Challenge

Detection and fine grained classification of objects in aerial imagery

Aviad Moreshet
204311112

Background...

1 The Task

...

2 The Dataset

Most of the dataset does not contain objects at all. Only 51 images are fully tagged. 2161 images are partially tagged, 7121 images were verified to contain no objects at all, and 36 images are untagged.

Regarding labels, the dataset definitely biases towards small vehicles, which appear more than the other classes, combined. Distribution can be seen in Figure 1:

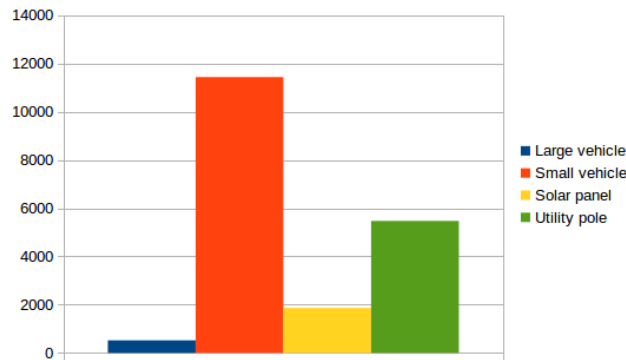


Figure 1: Dataset Coarse Grained Labels Distribution

As mentioned before, large and small vehicles are further tagged with fine-

grained labels, consisted of 3 categories: subclass, feature and color. The large vehicle class has only hundreds of subclass tags, while small vehicle has many thousands, as seen in Figure 2:

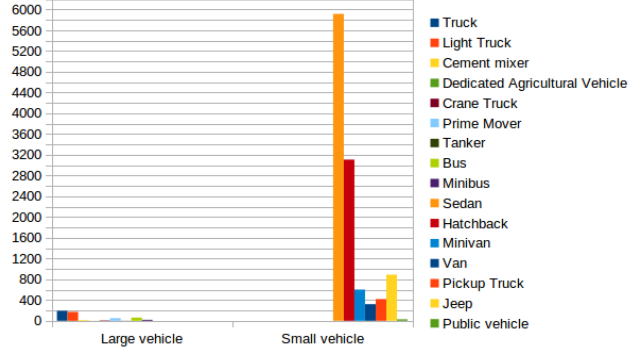


Figure 2: Dataset Subclass Labels Distribution

The same goes with feature tags, as seen in Figure 3:

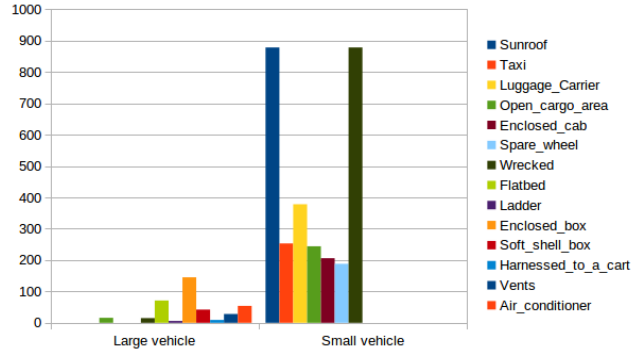


Figure 3: Dataset Feature Labels Distribution

And also with color tags, as seen in Figure 4.

On average, there are 8 objects in an image, while the most crowded image contains 164 objects altogether.

3 Challenges

1. Oriented Bounding Box

The dataset consisted of Oriented Bounding Boxes (Bounding Boxes which are not axis aligned).

All the known formats however supports axis aligned Bounding Boxes only, called Horizontal Bounding Boxes. Therefore, I had to wrap each OBB in HBB.

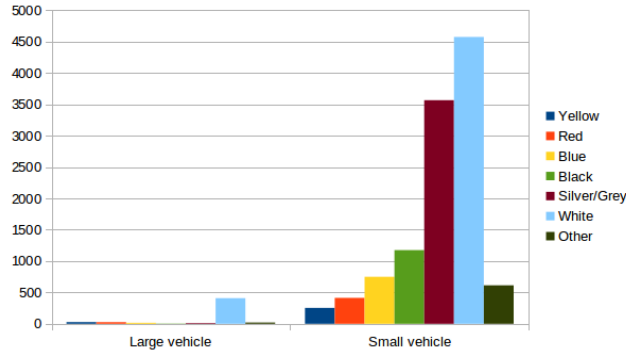


Figure 4: Dataset Color Labels Distribution

2. Classes with a point instead of Bounding Box

The utility pole class objects are marked with a point instead of 4 points Bounding Box. Therefore, for the beginning I've decided to remove all the utility poles objects from the dataset (and also images which contained only utility poles, 533 in total), as they are extremely difficult to detect.

3. Subclasses

I decided to ignore the fine-grained classification for now, since most of the detection networks modules do not support them.

4. Falsely annotated images

Some image annotations contained bounding boxes with coordinates outside of the image. I decided to crop these coordinates to fit inside image.

5. Small objects

The objects appeared in the images are much smaller than standard images networks handle. Therefore, the Region Proposal Network missed all the objects. I had to fine-tune some RPN related hyper-parameters in order to be able to detect the small objects in the dataset.

4 State Of The Art Computer Vision Networks

...

5 Experiments Results

I mostly used default learning-related hyper parameters in the following experiments, and fine-tuned the hyper parameters which would help the network to

better learn the small sized objects of the dataset. There are more than 30 hyper parameters, so I'll mention the default of the most important ones, and of course each experiment will describe the hyper parameters tuned for the experiment. Learning rate was 0.001, optimizer was SGD with 0.9 momentum, weight decay was $5e-4$,