

# 1 Statistics

The annotation table obtained from the image labelling is in the form of rectangle location and dimensions of the bounding box dimensions. The datasets are divided into three sets

1. *Training Dataset*: annotation data obtained using the consumer drone and used for training the model.
2. *Testing Dataset*: annotation data obtained using the consumer drone and used for testing the model.
3. *GeoScience Dataset*: annotation data obtained using the data obtained from the GeoScience Department using a commercial drone.

Bounding box heights, widths and aspect ratios can be obtained from the annotation data. The testing data and GeoScience data will be tested against the training data to see whether they are from the same population. If any populations are the same, the analysis done on one population should have the same evaluation results. If the populations are different, the analysis on both will be different, and the algorithms would need to be adapted to cater for this difference.

## 1.1 Training Dataset

Training bound boxes have a similar average width and average height pixel sizes. The average size is around 76 pixels, with minimum pixel size of around 28 pixels. The smallest bounding box matches the  $32 \times 32$  VGG16 pixel field of view within the *conv5* output feature map. The highest number of bounding boxes fall under a  $2 \times 2$  VGG16 feature map.

Table 1: Consumer Drone Bounding Box Statistics - Train data

	Mean	Std. Dev.	Min	Max	Population
Width	73.848	28.359	25	206	524
Height	79.293	27.197	31	207	524

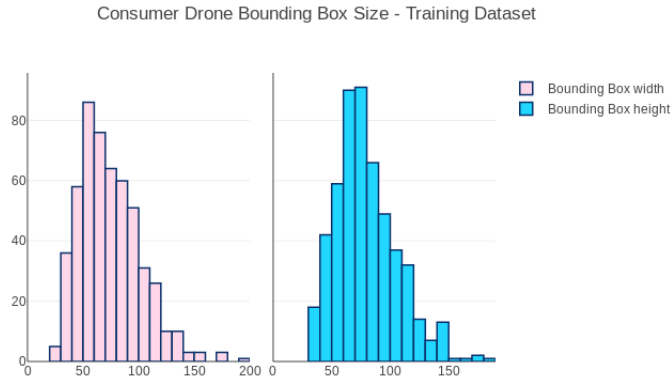


Figure 1: Consumer Drone Bounding Box Height and Width Histograms- Train data

As per qqplot show in figure 1.1 the data does not follow a normal distribution, so non-parametric tests will be performed to check that samples are obtained from the same population.

The aspect ratio of the training set as shown in the figure 1.1 shows that there is no preference between  $\log(-0.5)$  to  $\log(0.5)$ . This shows that there is no orientation preference of the bounding boxes in the training data set. As can be expected from aerial imagery data.

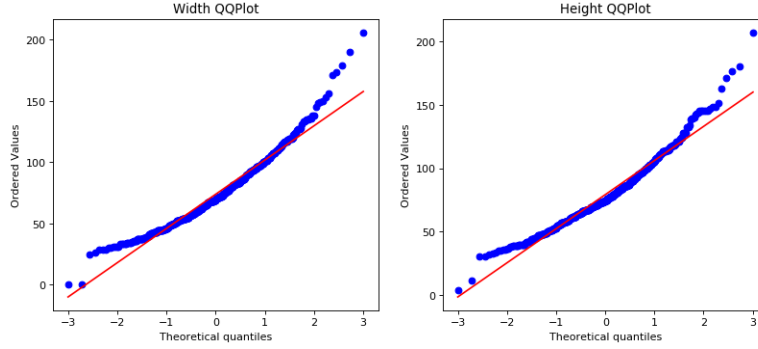


Figure 2: Consumer Drone Bounding Box Height and Width QQplot - Train data

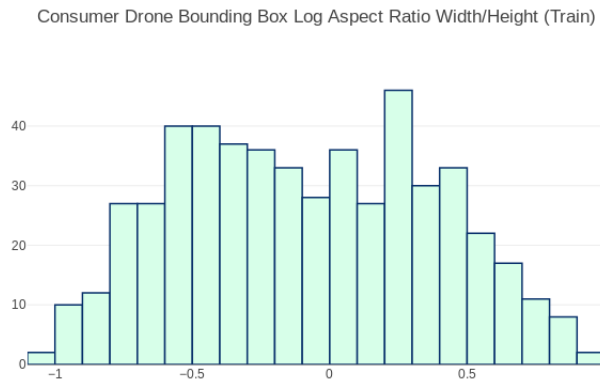


Figure 3: Consumer Drone Bounding Box Log Aspect Ratio Histogram - Train data

## 2 Testing Dataset

Testing bound boxes have a minor tendency to be larger in height than in width. The average size is also around 76 pixels, with minimum pixel size of around 29 pixels. Again this is within the field of view of the VGG16 pixel single feature vector of the *conv5* output feature map.

Table 2: Consumer Drone Bounding Box Statistics - Test data

	Mean	Std. Dev.	Min	Max	Population
Width	71.654	29.435	26	199	246
Height	81.232	34.203	30	272	246

As per qqplot show in figure 1.1 the data does not follow a normal distribution. This confirms the non-parametric tests to be performed on the population sampling.

The aspect ratio of the training set as shown in the figure 1.1 shows that there is no preference between  $\log(-0.5)$  to  $\log(0.5)$ . This shows that there is no orientation preference of the bounding boxes in the training data set. The smaller values at the 0 for both the training set and the test set indicate that the objects bounding boxes are rarely a box shape (i.e. equal width and height) which indicates that the objects have an elongated shape. This is to be expected as the images labelled are of beverage bottles and containers, which are predominantly manufactured to be stacked side-by-side, and ergonomically designed to be handled by a single hand.

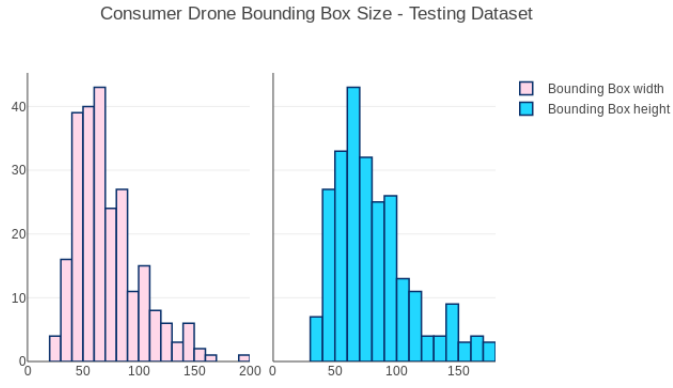


Figure 4: Consumer Drone Bounding Box Height and Width Histograms- Test data

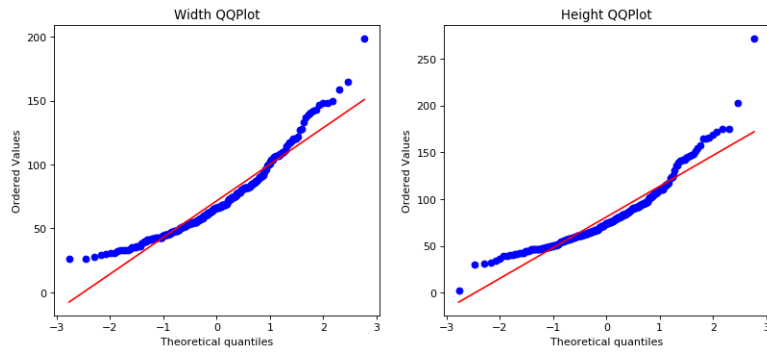


Figure 5: Consumer Drone Bounding Box Height and Width QQplot - Test data

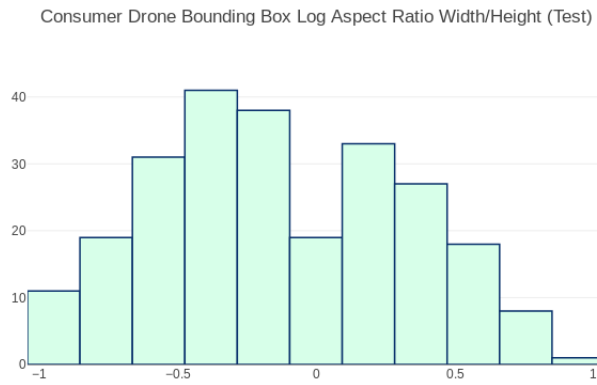


Figure 6: Consumer Drone Bounding Box Log Aspect Ratio Histogram - Test data

### 3 GeoScience Dataset

The Geoscience annotation also show no orientation preference with equal average width and height. The average size is also around 37 pixels, with minimum pixel size of around 11 pixels. This is a marked distinction from the previous datasets. The smaller bounding box sizes within this dataset might cause difficulties in the VGG16 field of view.

Table 3: GeoScience Drone Bounding Box Statistics - Data

	Mean	Std. Dev.	Min	Max	Population
Width	36.038	11.921	11	120	793
Height	37.318	10.420	10	82	793

The average size of the bounding boxes of the GeoSciences is half that of the two other datasets.

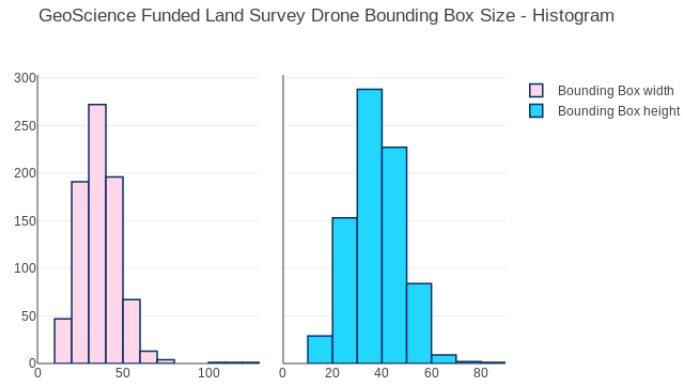


Figure 7: GeoScience Bounding Box Height and Width Histograms

As per qqplot show in figure 3 the data does not follow a normal distribution. This confirms the non-parametric tests to be performed on the population sampling.

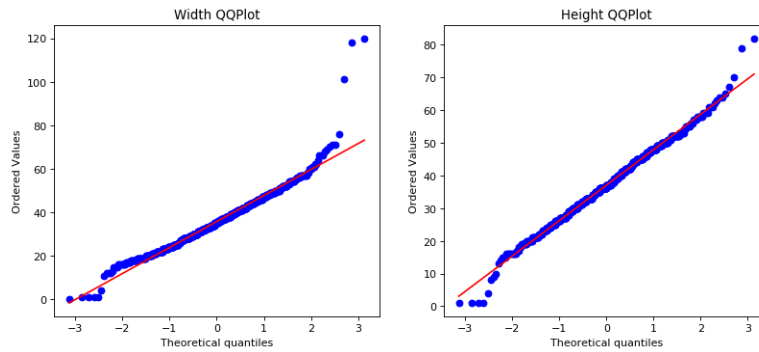


Figure 8: GeoScience Bounding Box Height and Width QQplot

The aspect ratio figure 3 has a similar shape as the train and data set, confirming the nature of the images being labelled.

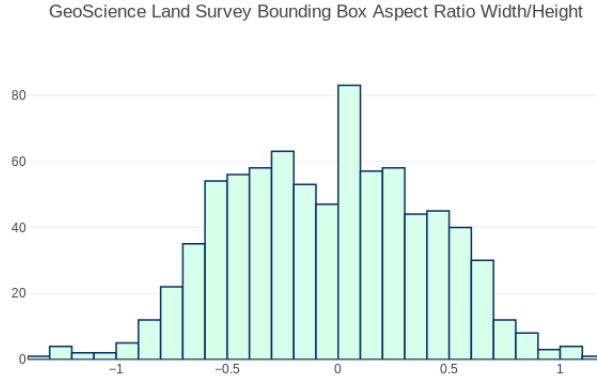


Figure 9: GeoScience Bounding Box Log Aspect Ratio Histogram

### 3.1 Population Testing

The train, test and Geoscience population will be tested for independance. Since the populations are not normally distributed, a non-parametric test will be used. A two-tail Mann Whitney test will be performed. The whole data in the datasets will be used. The scale will be in pixel size. A confidence interval  $\alpha = 0.05$  will be used.

The null hypothesis will be formulated :

**Null Hypothesis 1**  $H_0$ : *the distribution of the pixel sizes of the two datasets are equal*

**Hypothesis 1**  $H_A$ : *the distribution of the pixel sizes is not equal*

The two comparisons will be made between train vs test and train vs geoscience dataset.

#### 3.1.1 Train vs Testing Population

Testing the two datasets using Mann Whitney results with  $U = 69299.00$  and a p-value of 0.092 (for width) and  $U = 66246.50$  and a p-value of 0.533 (for height) . Therefore we fail to reject the Null Hypothesis  $H_0$ . The train and testing data set come from the same population. This is to be expected as the data comes from the same sensor and using the same land-survey techniques.

#### 3.1.2 Train vs GeoScience Population

Testing the two datasets using Mann Whitney results with  $U = 383029.00$  and a p-value of 0.0 (for width) and  $U = 400391.50$  and a p-value of 0.0 (for height) . Therefore we reject the Null Hypothesis  $H_0$ . The train and geoscience dataset come from the different populations.

## 4 Statistical Results

The statistical tests show that the geoscience and the consumer drone data have different image sizes. This shows that the surveying techniques and parameters are different between geoscience data and consumer drone data. Comparing the geoscience images to that of the consumer drone, it can be deduced that the altitude of the surveys are different. This can also be the result of the drone operator knowledge of higher camera resolution.

An algorithm which is dependant on the size of the picture for successful recognition, will have difficulty correctly infering the smaller sized dataset. The ability of the human labeller to correctly label the objects,

indicates that the shape information is present. This has to be exploited by the CNN. The algorithm must be scale invariant or specifically tuned to the zoom levels of the geoscience dataset.