



Technische
Universität
Braunschweig



Object Detection – Exploring DIOR dataset

Group 5

Amulya Harihara Narayana Rao
Vigneshwar Karuppiyah Ramanathan

M.Sc. Computational Sciences in Engineering

Introduction

Problem task

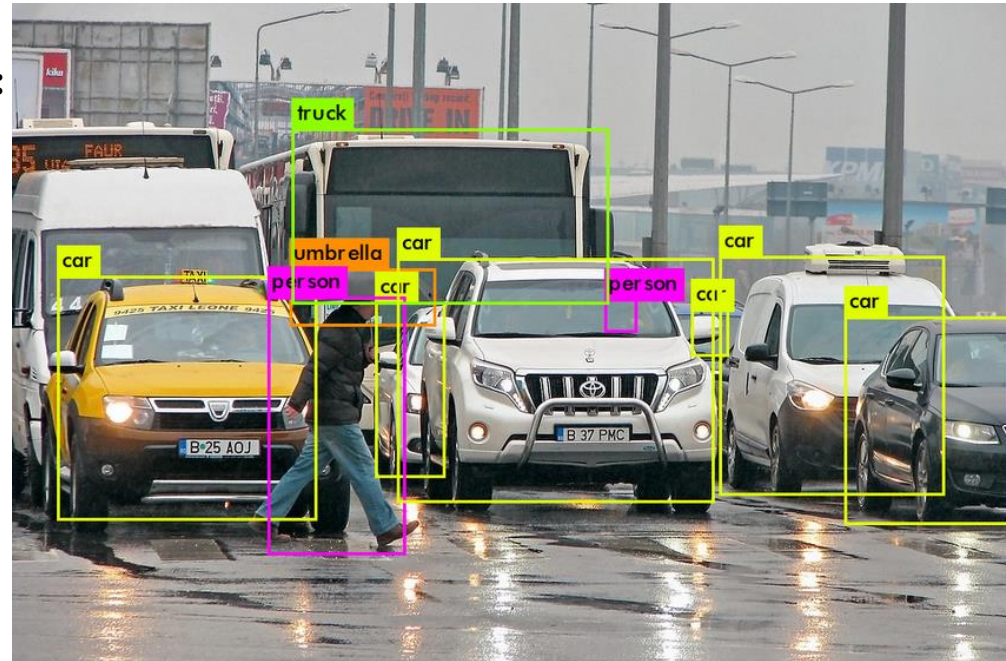
Object detection in computer vision is the process of identifying and locating objects within digital images or videos.

Standard steps involved in object detection:

1. Input data
2. Feature extraction
3. Classification
4. Localization
5. Post processing
6. Output

This task can be implemented using,

- Deep learning approaches
- Traditional computer vision methods`



Source : Research Gate

DIOR Dataset

- Large-scale, publicly available benchmark for Object Detection in Optical Remote sensing images.

Comparison between the proposed DIOR dataset and nine publicly available object detection datasets in earth observation community.

Datasets	# Categories	# Images	# Instances	Image width	Annotation way	Year
TAS	1	30	1319	792	horizontal bounding box	2008
SZTAKE-INRIA	1	9	665	~800	oriented bounding box	2012
NWPU VHR-10	10	800	3775	~1000	horizontal bounding box	2014
VEDAI	9	1210	3640	1024	oriented bounding box	2015
UCAS-AOD	2	910	6029	1280	horizontal bounding box	2015
DLR 3K Vehicle	2	20	14235	5616	oriented bounding box	2015
HRSC2016	1	1070	2976	~1000	oriented bounding box	2016
RSOD	4	976	6950	~1000	horizontal bounding box	2017
DOTA	15	2806	188282	800-4000	oriented bounding box	2017
DIOR (ours)	20	23463	192472	800	horizontal bounding box	2018

Source : [\(Paper\)](#)

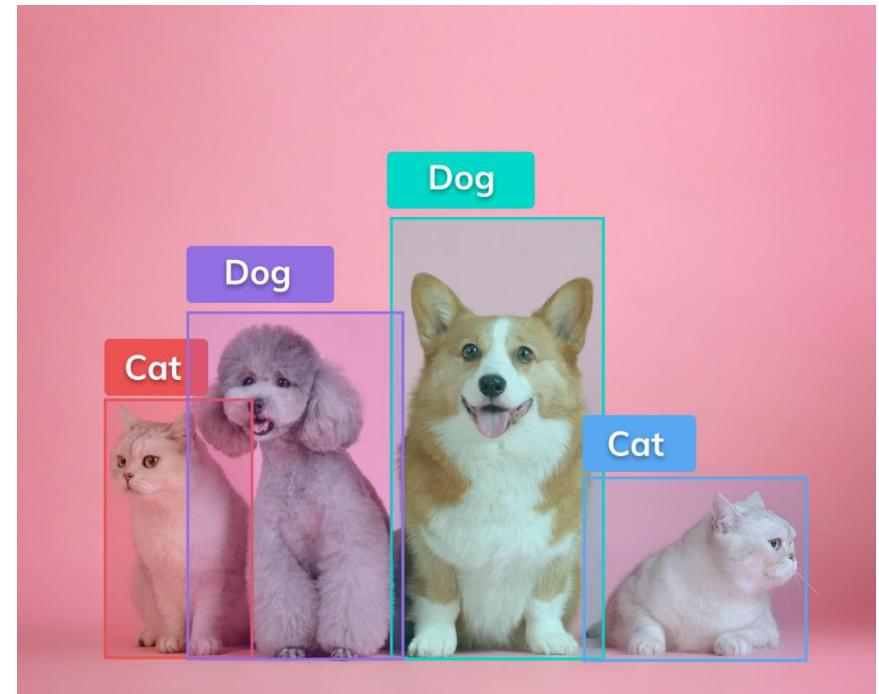
Disadvantages of DIOR's predecessors and other publicly available dataset:

- Low numbers of images
- Small scale object categories
- Image diversity & variations are insufficient
- This blocks the development of deep learning based methods

- contains 23463 images and 192472 instances, covering 20 object classes
- each category contains about 1200 images.

Inputs and outputs of a object detection model

- **Inputs** are images or videos
- **Output** is bounding boxes with class labels
- **Regression problem** because the class probabilities as the output space is continuous in nature
- **Output representation** is done in following ways:
 1. Top-left, Bottom-right coordinates
 2. Top-left coordinate, width, height
 3. Top-left coordinate, width, height and Theta
 4. 'n' number of (x , y) coordinates

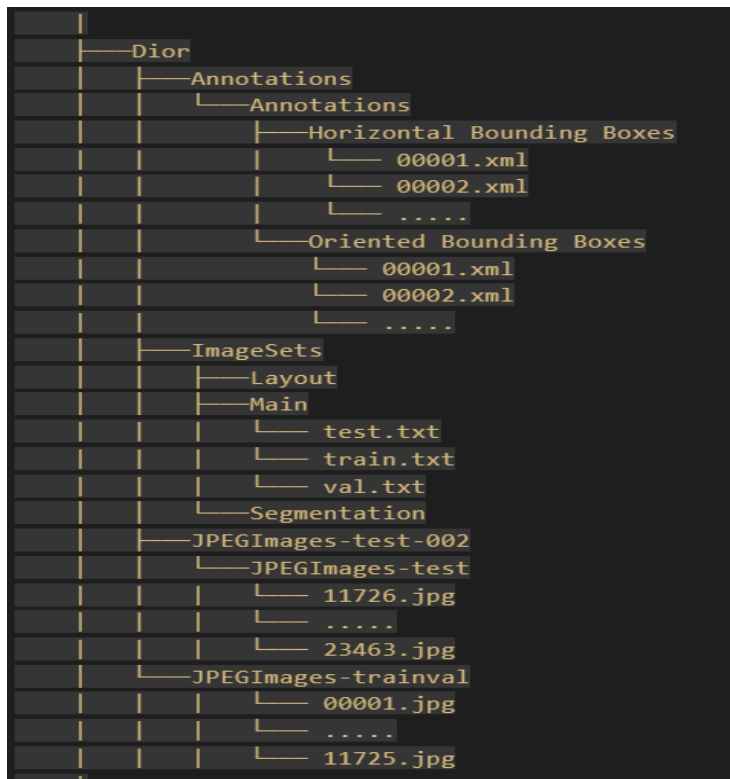


Source : V7 Labs

General structure of DIOR dataset

DATASET

- Test dataset contains 11738
- Train Val dataset contains 11725



ANNOTATION FORMAT

- Pascal VOC (XML File)
- *from xml.etree import ElementTree*

```
1  <annotation>
2      <filename>00001.jpg</filename>
3      <source>
4          <database>DIOR</database>
5      </source>
6      <size>
7          <width>800</width>
8          <height>800</height>
9          <depth>3</depth>
10     </size>
11     <segmented>0</segmented>
12     <object>
13         <name>golffield</name>
14         <pose>Unspecified</pose>
15         <bndbox>
16             <xmin>133</xmin>
17             <ymin>237</ymin>
18             <xmax>684</xmax>
19             <ymax>672</ymax>
20         </bndbox>
21     </object>
22 </annotation>
23
```

Types of annotation output

The final object detection output can be visualized in two ways, namely:

Horizontal Bounding Box



4 Parameters

(X_{min} , Y_{min} , X_{max} , Y_{max})

Oriented Bounding Box



4 Parameters + Angle parameter

(X_{left_top} , Y_{left_top} , X_{right_top} ,
 Y_{right_top} , X_{left_bottom} , Y_{left_bottom} ,
 X_{right_bottom} , Y_{right_bottom})

Source : Research Gate

Deep learning, Object detection (DIOR Dataset) | Group 5 | page 6

Classes available in DIOR Dataset (20 Classes)

- Airplane
- Airport
- baseball field
- basketball court
- Bridge
- Chimney
- Dam
- expressway service area
- expressway toll station
- Harbor
- golf course
- ground track field
- Overpass
- Ship
- Stadium
- storage tank
- tennis court
- train station
- Vehicle
- Wind mill

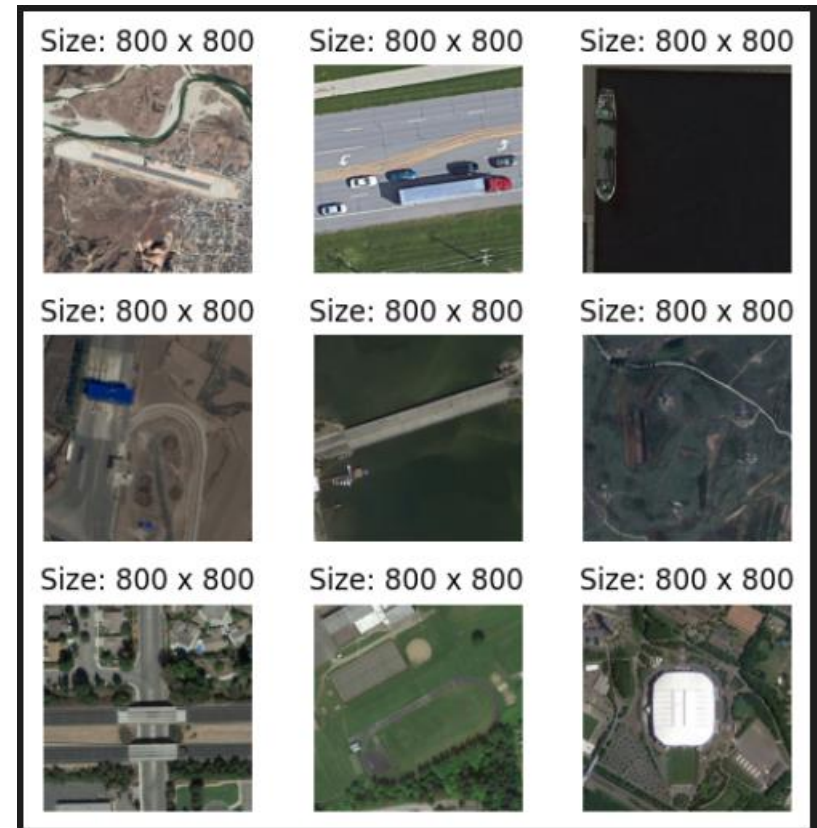


Source : [\(Paper\)](#)

Sample Images

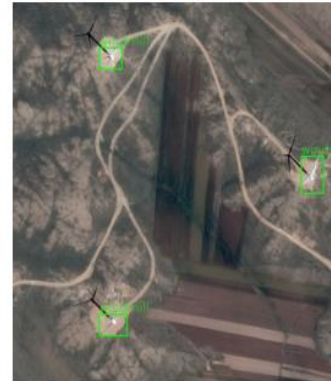
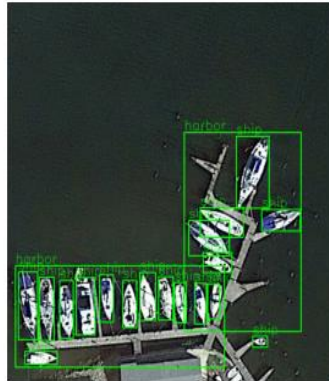
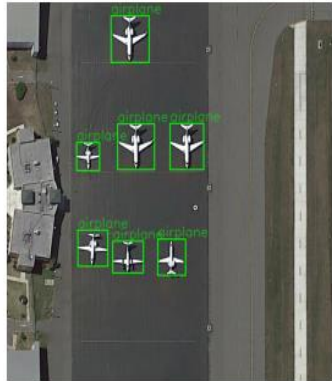
- 800 * 800 pixel images
- 3 Channels
- Spatial resolution 0.5 m to 30 m
- Large range of size variations of object instances
- Richer variations in viewpoint, translation, illumination, background, object pose and appearance, occlusion, etc.

Total number of possible combination of classes is 526



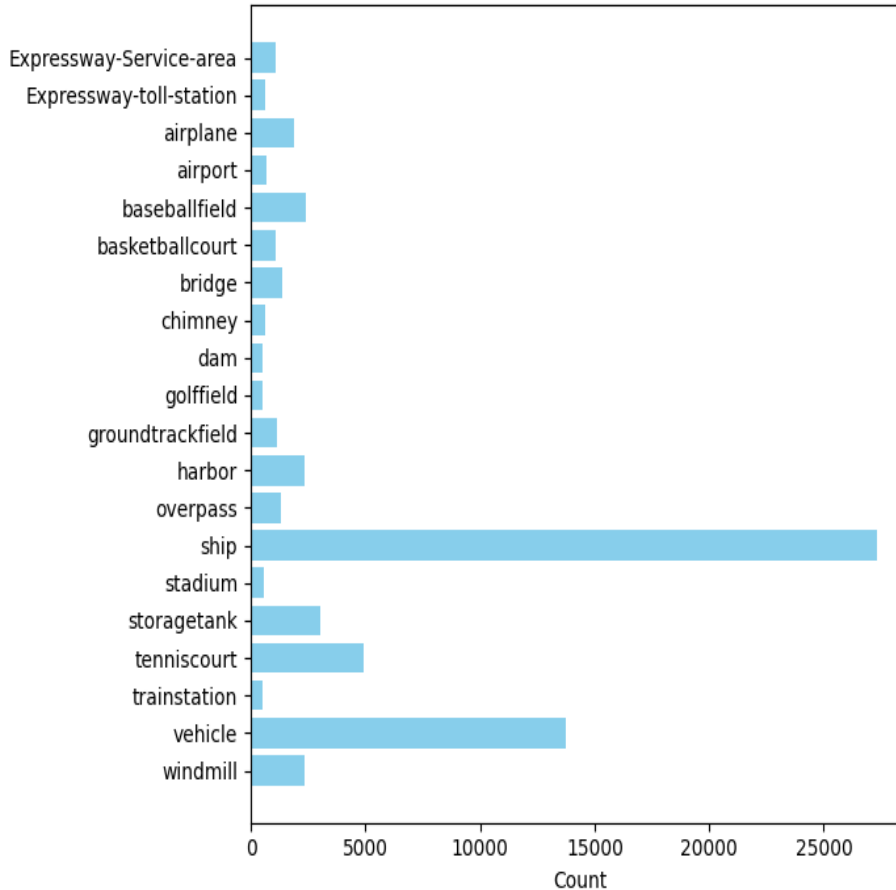
Source : [\(Paper\)](#)

Sample Images with bounding boxes and labels

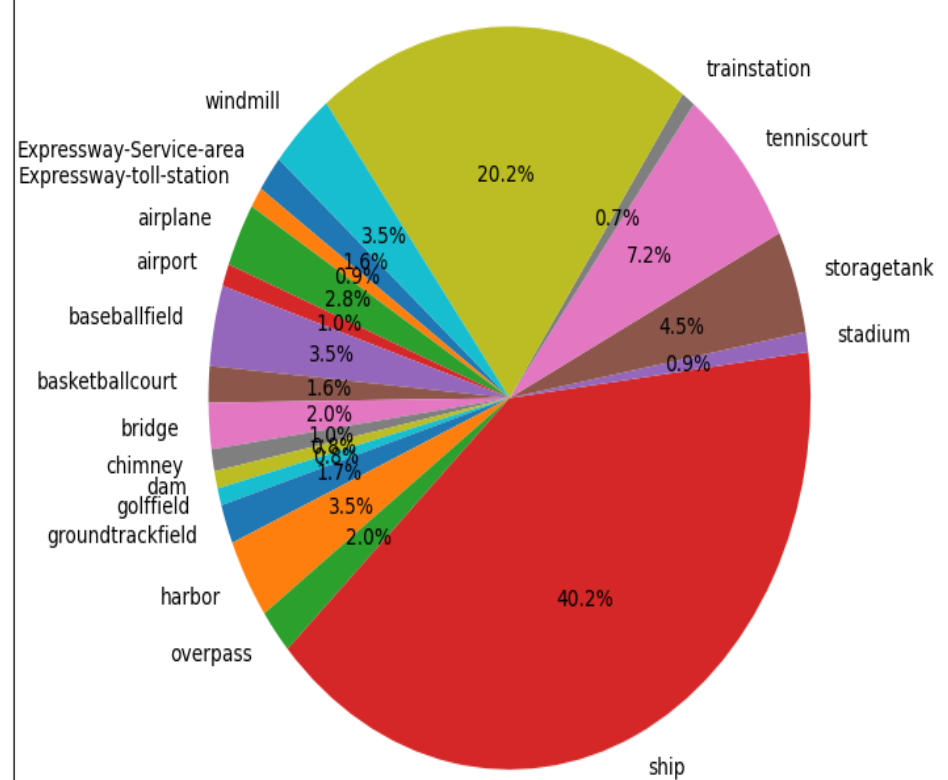


Object Instance counts in Train Val dataset

Object Instance Counts in trainval dataset

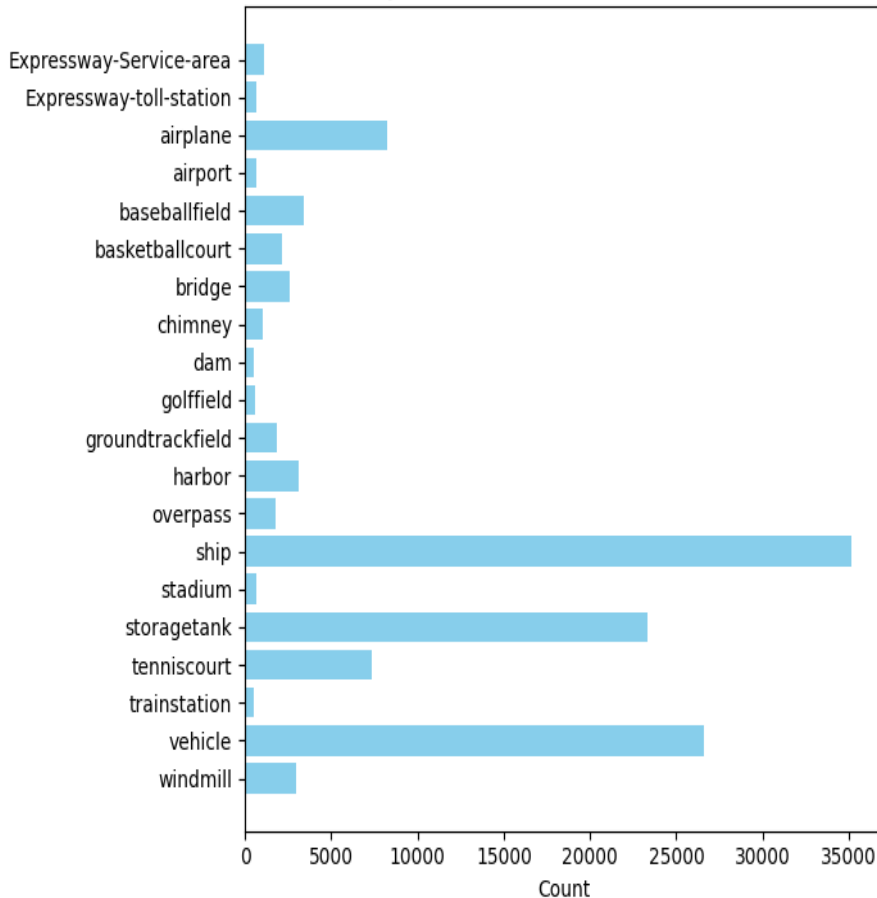


Object Instance Distribution in trainval dataset

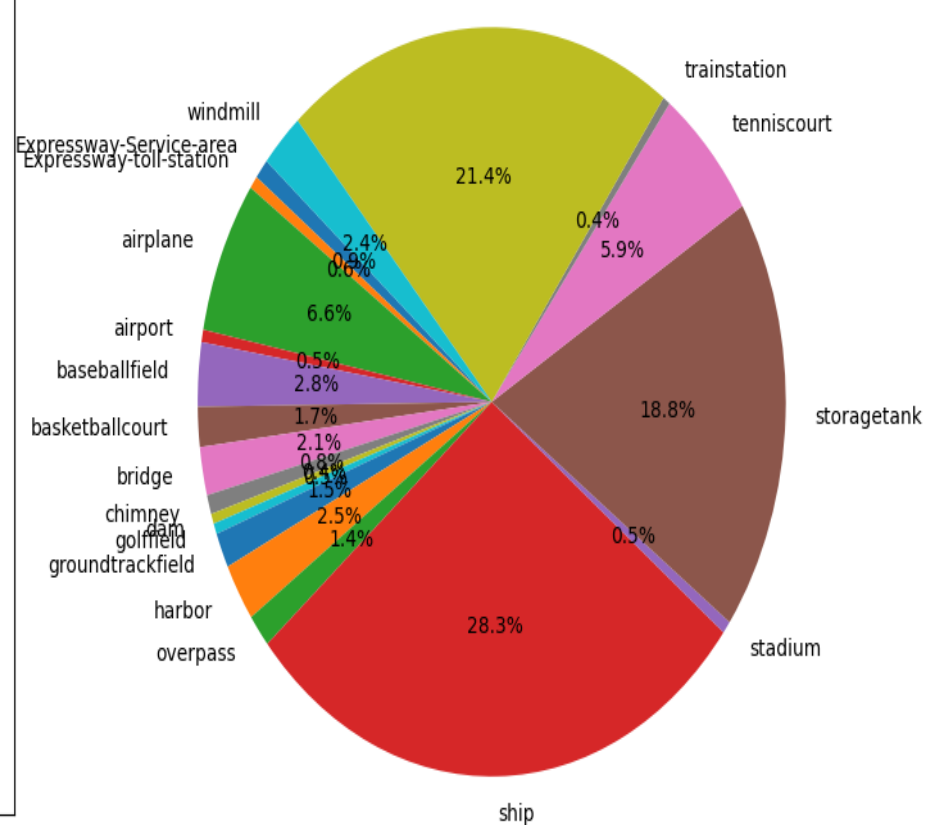


Object Instance counts in Test dataset

Object Instance Counts in test dataset



Object Instance Distribution in test dataset



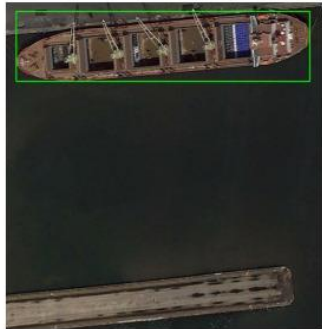
Aspect ratios

Image Aspect Ratio

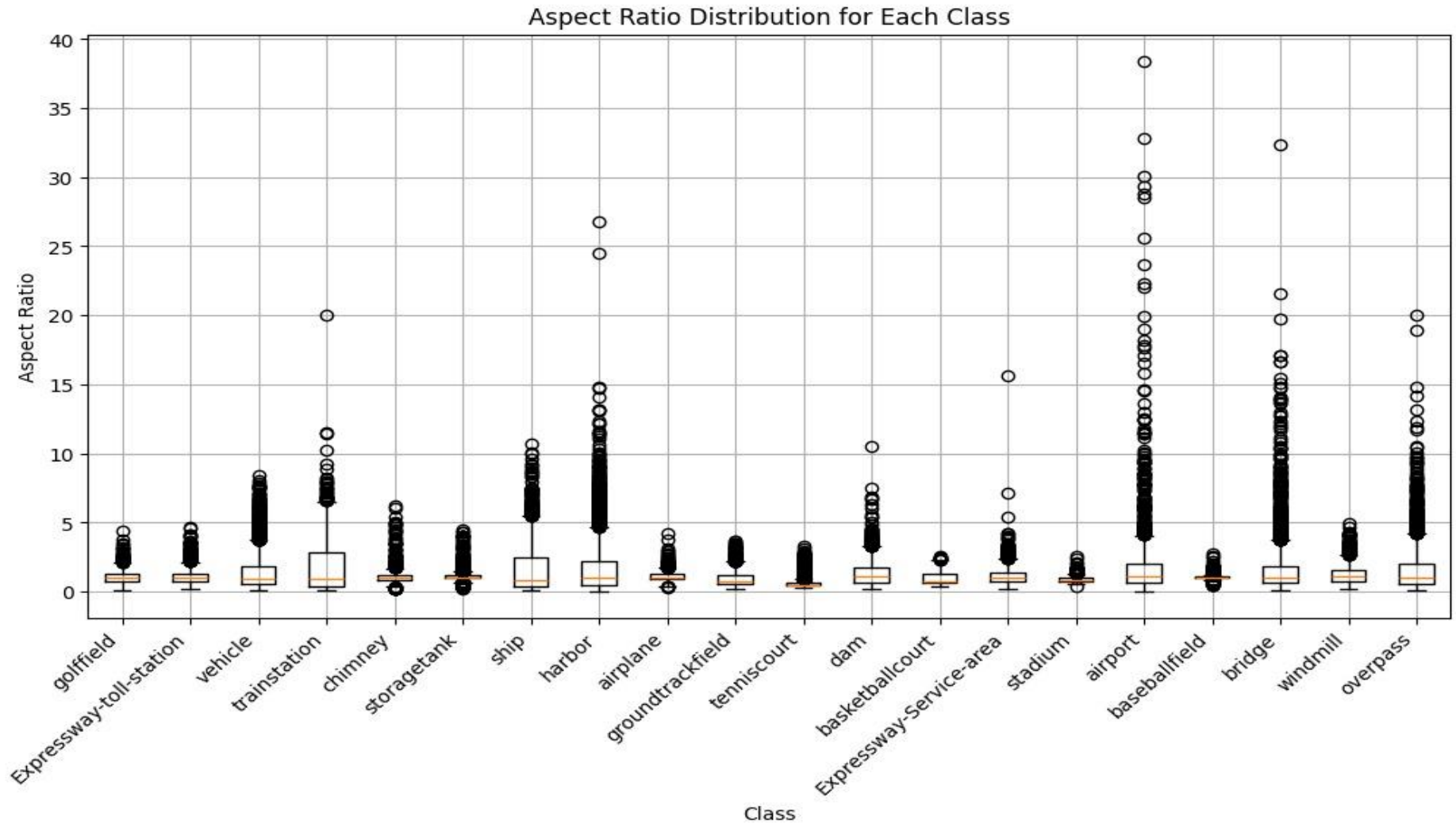
- Every image is of shape 800*800 pixels
- Therefore the Aspect ratio is exactly one
- It's a square

Object Aspect Ratio

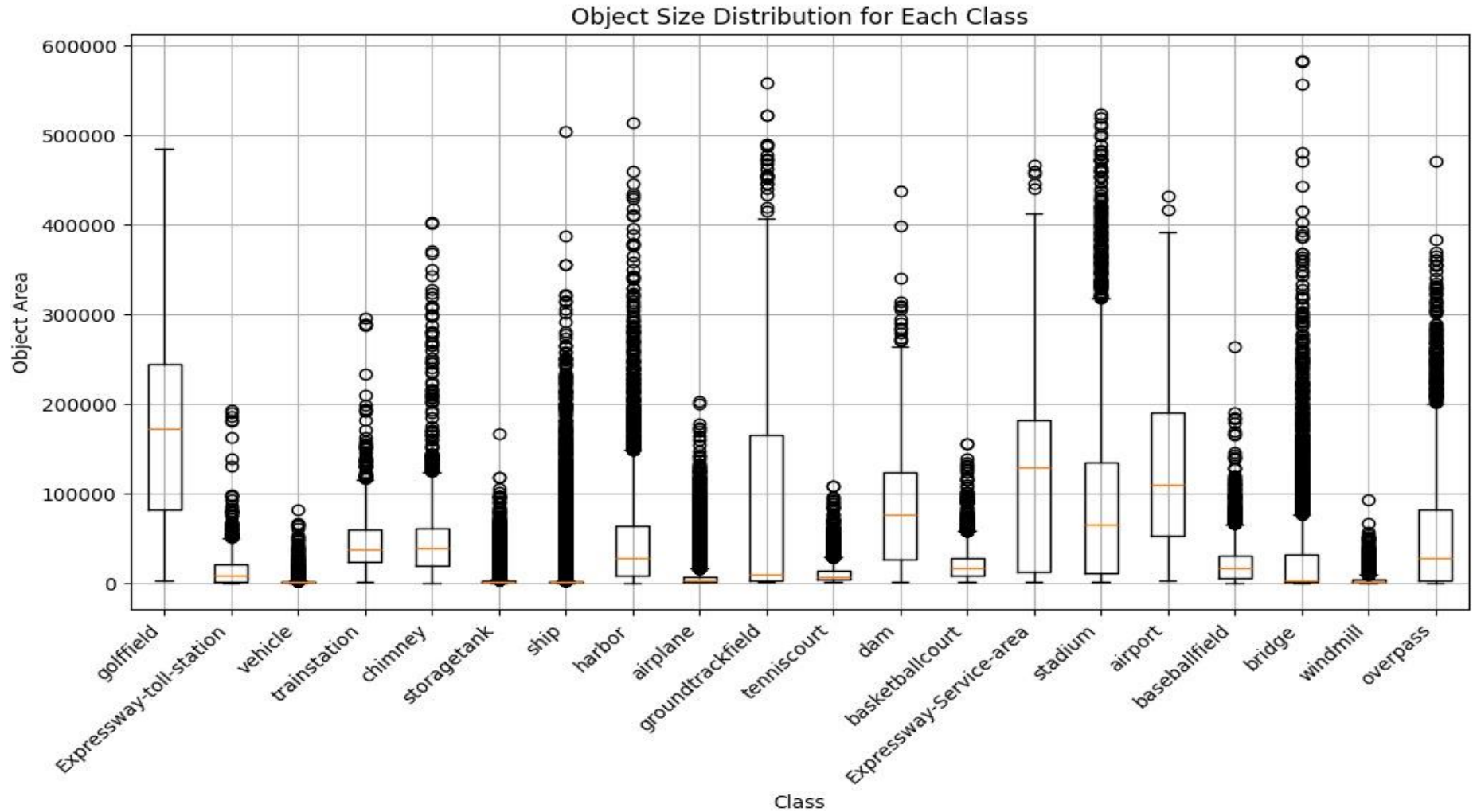
- Ratio of the width to the height of a specific object within an image



Object aspect ratio distribution



Object size distribution for each class



Potential problems and things to look out for

- Imbalance in class distribution may make the model biased to one class than the others.
- Annotation format of each image might have to be changed based on the model we are using.
- Look for preprocessing techniques.

Reference

- Object Detection in Optical Remote Sensing Images: A Survey and A New Benchmark: Ke Li, Gang Wan, Gong Cheng*, Liqiu Meng, Junwei Han* [\(Paper\)](#)
- [Dataset](#)

Thank you