

# Dataset Exploration Report

## 1. Introduction

- **Dataset Source:** COCO (Common Objects in Context)
- **Purpose:** Object detection in autonomous driving scenarios, aiming to identify and classify objects like pedestrians, vehicles, traffic signs, and obstacles in real-time.

## 2. Dataset Composition

- **Number of Images:** 26300 (deleted unannotated images).
- **Number of Objects (Bounding Boxes):** 11.
- **Classes:**

```
Keys: dict_keys(['info', 'licenses', 'categories', 'images', 'annotations'])

Sample Image Entry:
{'id': 0, 'license': 1, 'file_name': '1478897026627294725_jpg.rf.6828a4e821cbab4c2c277d74df291f00.jpg', 'height': 512, 'width': 512, 'date_captured': '2021-06-09T12:24:25+00:00'}
{'id': 2, 'license': 1, 'file_name': '1478900859981702684_jpg.rf.6830635c7d9197475638f0818f5dd103.jpg', 'height': 512, 'width': 512, 'date_captured': '2021-06-09T12:24:25+00:00'}
{'id': 3, 'license': 1, 'file_name': '1478899989907374870_jpg.rf.684aa8281432f9513dad8d91c54a1346.jpg', 'height': 512, 'width': 512, 'date_captured': '2021-06-09T12:24:25+00:00'}

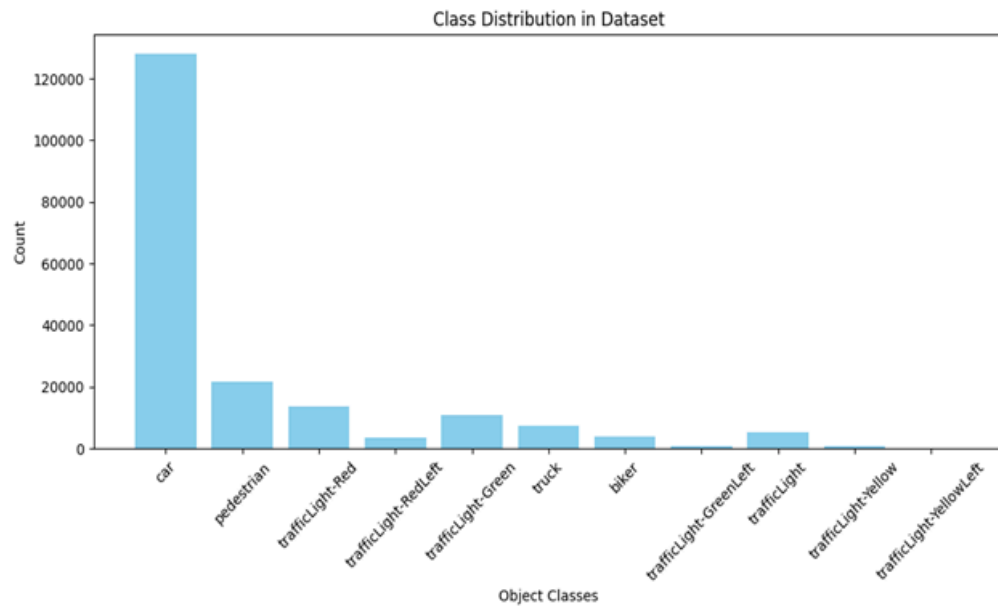
Sample Annotation Entry:
{'id': 0, 'image_id': 0, 'category_id': 2, 'bbox': [140, 262, 21, 25.5], 'area': 535.5, 'segmentation': [], 'iscrowd': 0}
{'id': 1, 'image_id': 0, 'category_id': 2, 'bbox': [266, 231, 24, 95.5], 'area': 2292, 'segmentation': [], 'iscrowd': 0}
{'id': 2, 'image_id': 0, 'category_id': 2, 'bbox': [271, 241, 17, 76], 'area': 1292, 'segmentation': [], 'iscrowd': 0}

Categories:
{'id': 0, 'name': 'obstacles', 'supercategory': 'none'}
{'id': 1, 'name': 'biker', 'supercategory': 'obstacles'}
{'id': 2, 'name': 'car', 'supercategory': 'obstacles'}
{'id': 3, 'name': 'pedestrian', 'supercategory': 'obstacles'}
{'id': 4, 'name': 'trafficslight', 'supercategory': 'obstacles'}
{'id': 5, 'name': 'trafficslight-Green', 'supercategory': 'obstacles'}
{'id': 6, 'name': 'trafficslight-Greeneft', 'supercategory': 'obstacles'}
{'id': 7, 'name': 'trafficslight-Red', 'supercategory': 'obstacles'}
{'id': 8, 'name': 'trafficslight-Redeft', 'supercategory': 'obstacles'}
{'id': 9, 'name': 'trafficslight-Yellow', 'supercategory': 'obstacles'}
{'id': 10, 'name': 'trafficslight-Yelloweft', 'supercategory': 'obstacles'}
{'id': 11, 'name': 'truck', 'supercategory': 'obstacles'}
```

- **Bounding Box Format:** COCO format [x\_min, y\_min, width, height].

### 3. Class Distribution Analysis (📊)

- **Histogram of Object Categories:** A visualization showing the distribution of object categories to check for class imbalances.

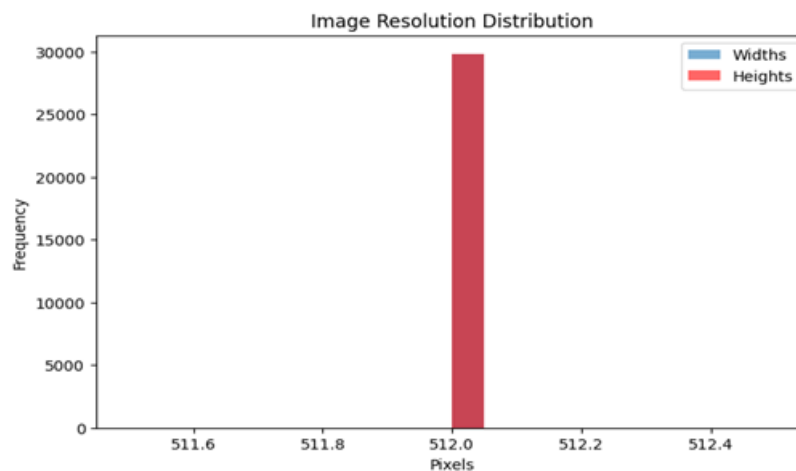


### 4. Image Quality & Metadata Analysis

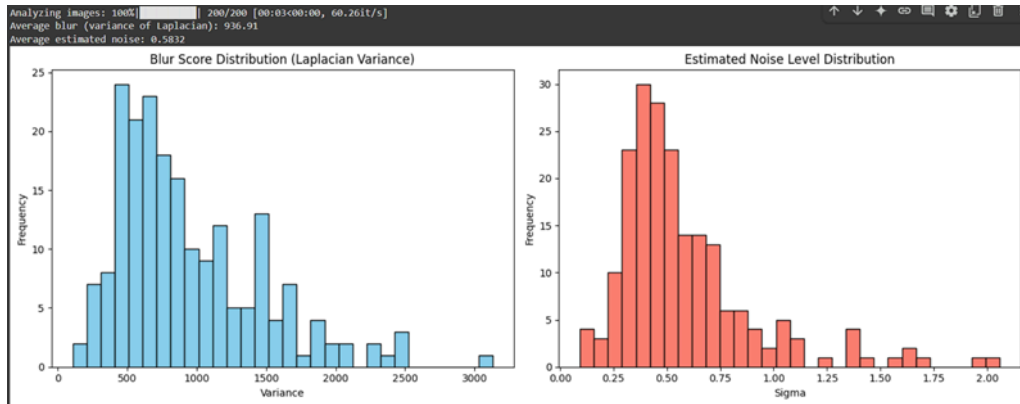
- **Resolution Distribution:**

Image Resolution Stats:

- **Min size: 512x512**
- **Max size: 512x512**
- **Avg size: 512x512**

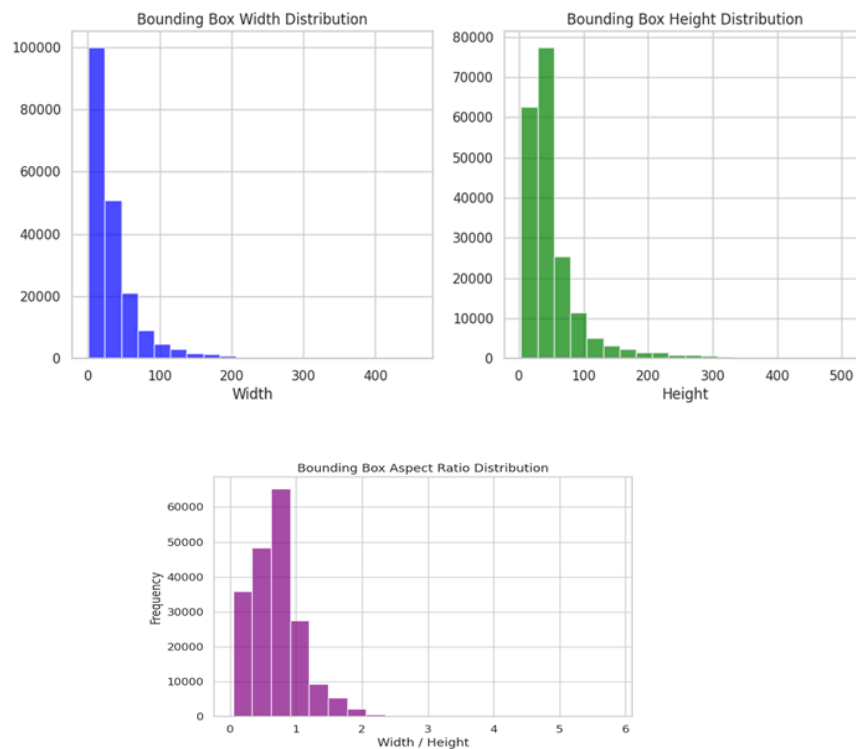


- **Blur & Noise Analysis:** Identification of images with low quality due to motion blur, poor lighting, or noise.



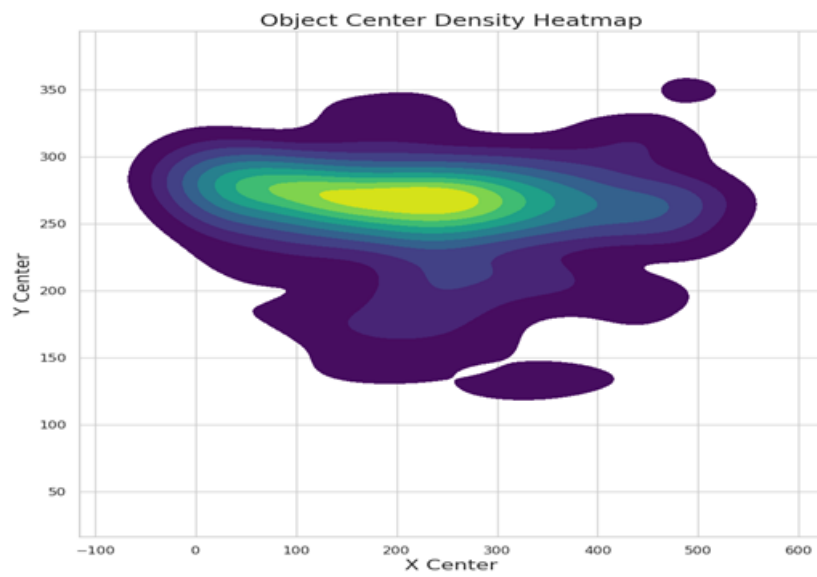
## 5. Bounding Box Validity Check

- **Total Bounding Boxes:** 194539
- **Invalid Bounding Boxes:** Zero (number and percentage of bounding boxes exceeding image dimensions or having negative coordinates).
- **Bounding Box Size Distribution:** Statistics on bounding box widths and heights.

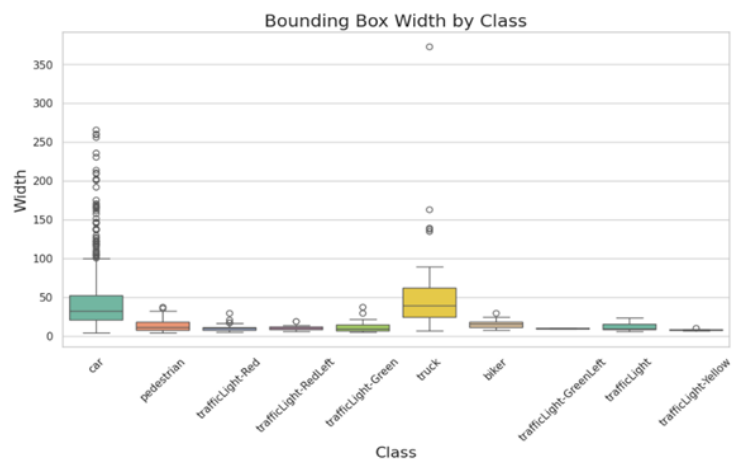


## 6. Spatial Distribution Analysis

## Spatial Distribution Analysis



## Bounding Box Width by Class



## 7. Bias Detection & Visualization:

	class	count	percentage	avg_size
0	car	127873	65.731293	4087.007779
1	pedestrian	21491	11.047142	1426.195942
2	trafficLight-Red	13673	7.028411	396.755713
4	trafficLight-Green	10838	5.571119	308.702008
5	truck	7194	3.697973	8258.766767
8	trafficLight	5101	2.622096	630.617981
6	biker	3704	1.903988	1567.550964
3	trafficLight-RedLeft	3482	1.789872	308.634486
7	trafficLight-GreenLeft	614	0.315618	330.293338
9	trafficLight-Yellow	541	0.278093	347.246753
10	trafficLight-YellowLeft	28	0.014393	311.580579

Normalized Object Positions by Class

