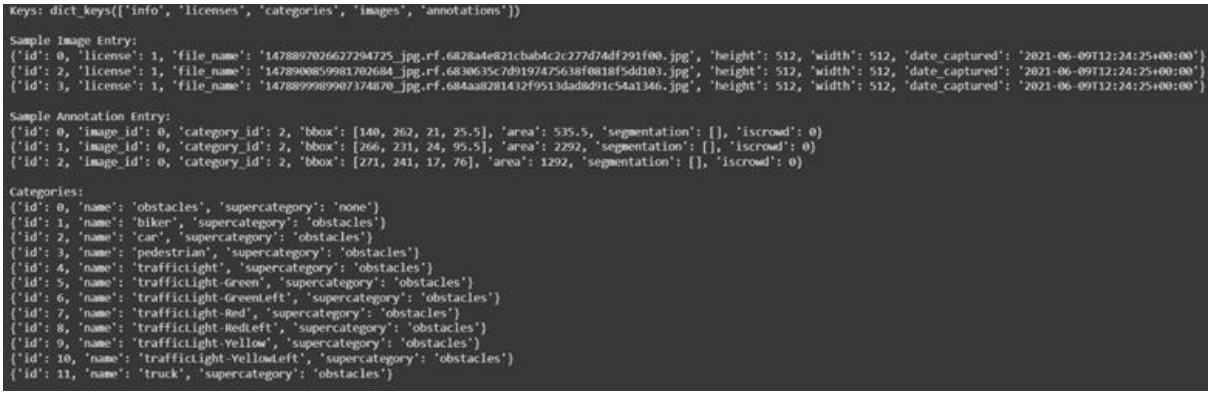
Dataset Exploration Report

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

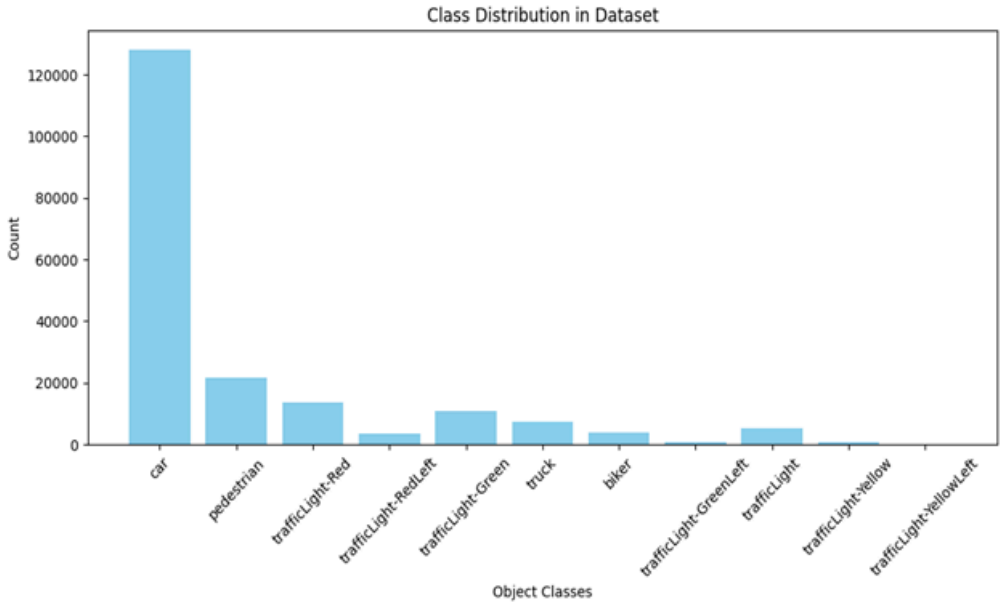
- **Number of Images:** 26300 (deleted unannotated images).
- **Number of Objects (Bounding Boxes):** 11.
- **Classes:**



- **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.

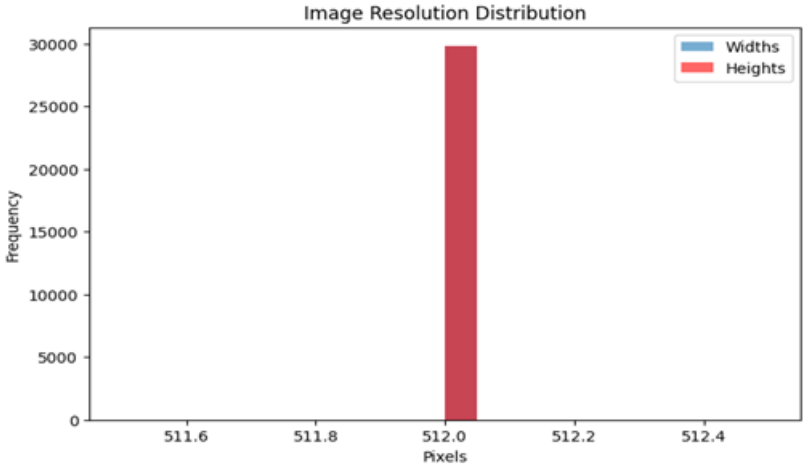


4. Image Quality & Metadata Analysis

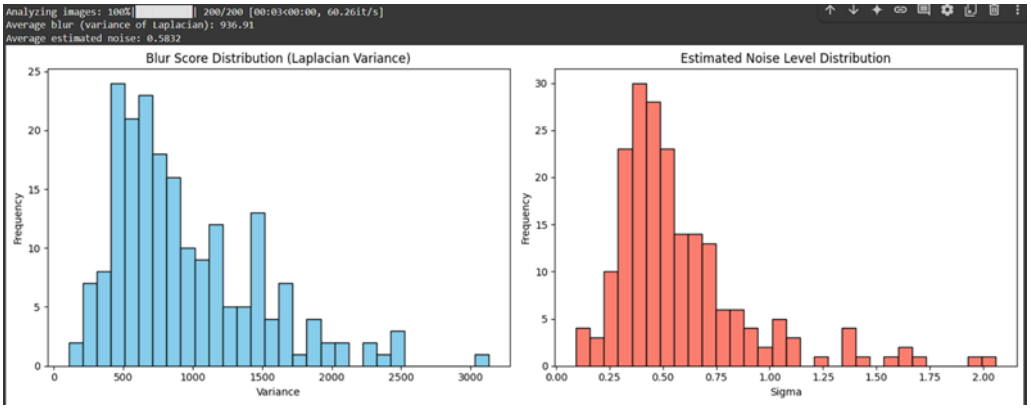
- **Resolution Distribution:**

  Image Resolution Stats:

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

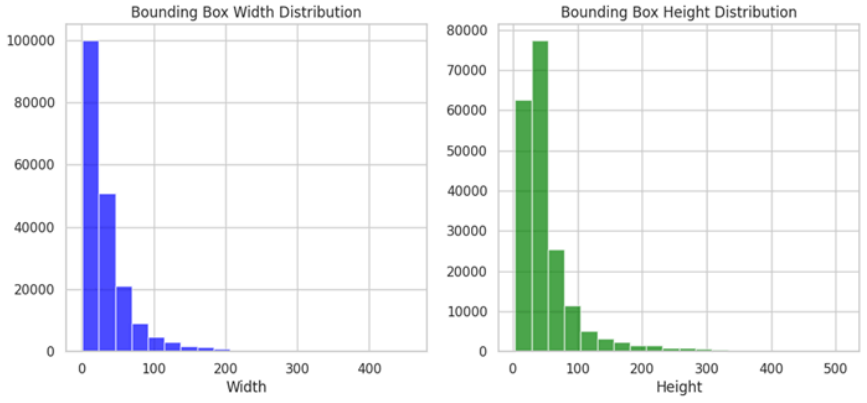


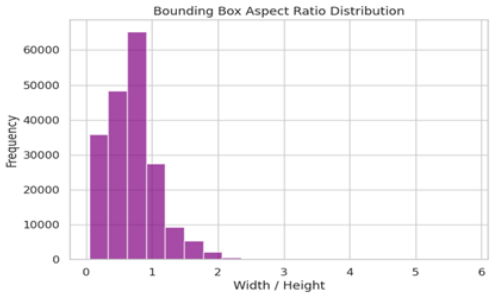
- **Blur & Noise Analysis:** Identification of images with low quality due to motion blur, poor lighting, or noise.



5. Bounding Box Validity Check

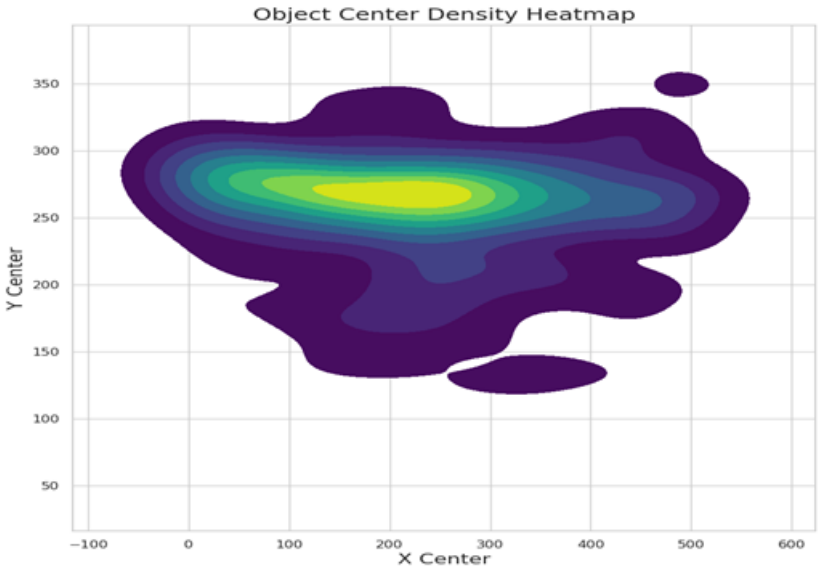
- **Total Bounding Boxes:** 194539
- **Invalid Bounding Boxes:** Zero (number and percentage of bounding boxes exceeding image dimensions or having negative coordinates).
- **Bounding Box Size Distribution:** Statistics on bounding box widths and heights.



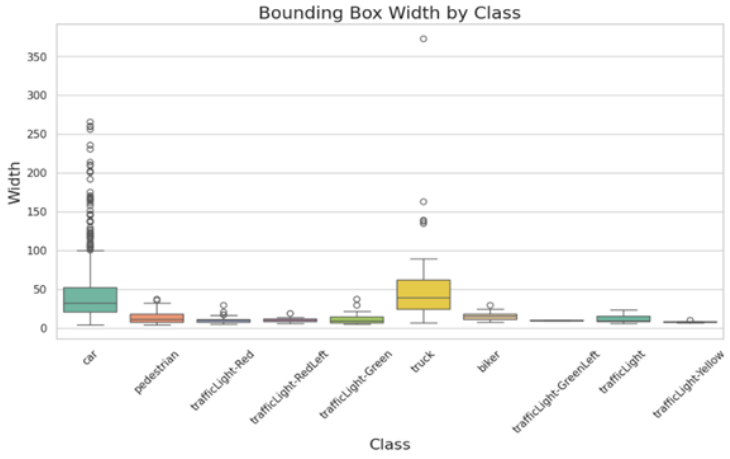


6. Spatial Distribution Analysis

- **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