DATA SCIENCE & AI LAB (BSCSS3001)

MILESTONE - 2: Dataset Preparation

GROUP NO. 2

PRASHASTI SARRAF **(21f1001153)**
TANUJA NAIR **(21f1000660)**
BALASURYA K **(22f3002744)**
KARAN PATIL **(22f2001061)**
JIVRAJ SINGH SHEKHAWAT **(22f3002542)**

IITM BS Degree Program
Indian Institute of Technology,
Madras, Chennai,
Tamil Nadu, India, 600036

Vision Assist: Real-Time Navigation Support for the Visually Impaired

Dataset Preparation

1. Abstract

This milestone focuses on preparing the dataset required for training and fine-tuning a lightweight object-detection model (YOLOv8) intended to assist visually impaired individuals by recognizing street-level obstacles, pedestrians, and vehicles in real time.

We used the MS COCO 2017 train split as the base dataset because of its rich, diverse set of urban scenes and standardized annotations. From this, a 500-image curated subset was extracted for rapid experimentation. Each image contains detailed bounding-box annotations in COCO JSON format corresponding to multiple object classes such as person, car, truck, traffic light, and stop sign.

This curated dataset will later be extended with custom video-frame images collected from real-world navigation scenarios, annotated using ReDeTR for object boundaries and categories. Together, these two sources will form a robust, multi-context dataset for fine-tuning YOLOv8 to achieve reliable performance on wearable or mobile hardware.

2.Dataset Overview

Aspect	Description
Dataset Name	COCO 2017 (Subset + Custom Frames planned)
Source	Official MS COCO 2017 dataset (http://images.cocodataset.org/annotations/annotations_trainval2017.zip)

Subset Used	1000 images from train2017
Annotation Format	COCO JSON (instances_coco_sample.json)
Image Resolution	640 × 480 (standardized)
Annotation Fields	image_id, category_id, bbox, area, iscrowd
Total Annotations	2,315 bounding boxes
Object Classes	5 (main): person, car, truck, traffic light, stop sign
Split Ratio	Train 80%, Test 10%, Validation 10%
Storage Location	Drive Link for Dataset Samples and Results and VisionAssist-Dataset (50k+ samples)
License	COCO Creative Commons Attribution 4.0 License
Intended Use	Visual perception module training for assistive navigation

3.Data Preparation Process

Dataset Acquisition:

Downloaded COCO 2017 annotations (annotations_trainval2017.zip) and accessed the train2017 image set via Hugging Face Datasets (detection-datasets/coco).

Sampling and Curation:

A random sample of 1000 images was extracted to reduce processing load. These images span a diverse set of real-world scenes including pedestrians, roads, crosswalks, and vehicles.

Annotation Extraction:

For each image, all associated object annotations were extracted from the COCO JSON file and re-mapped to a compact subset JSON:
/content/drive/MyDrive/M2-Dataset/instances_coco_sample.json
Corresponding images were stored under:
/content/drive/MyDrive/M2-Dataset/images/coco_sample/

Label Formatting for YOLOv8:

Bounding boxes were converted into YOLOv8-compatible TXT files using the format:
<class_id> <x_center> <y_center> <width> <height>
with coordinates normalized to [0, 1].

Data Integrity Check:

Verified that each image had a corresponding annotation file, ensuring no missing entries or empty boxes.

4.Sample Annotation Structure


```
{
  "images": [
    {
      "license": 4,
      "file_name": "000000370701.jpg",
      "coco_url": "http://images.cocodataset.org/train2017/000000370701.jpg",
      "height": 640,
      "width": 640,
      "date_captured": "2013-11-17 08:00:35",
      "flickr_url": "http://farm6.staticflickr.com/5451/9490830088_7279abceef_z.jpg",
      "id": 370701
    },
    {
      "license": 3,
```

```
    "file_name": "000000413736.jpg",
    "coco_url": "http://images.cocodataset.org/train2017/000000413736.jpg",
    "height": 640,
    "width": 480,
    "date_captured": "2013-11-20 23:20:47",
    "flickr_url": "http://farm1.staticflickr.com/23/89721544_0c3845a923_z.jpg",
    "id": 413736
  },..],
"annotations": [
  {
    "segmentation": [
      [...
        139.01,
        631.51
      ]
    ],
    "area": 12328.5508,
    "iscrowd": 0,
    "image_id": 215867,
    "bbox": [
      126.62,
      438.82,
      295.92,
      194.07
    ],
    "category_id": 2,
    "id": 126302
  },...,
],
"categories": [
  {
    "supercategory": "person",
    "id": 1,
    "name": "person"
  },
  {
    "supercategory": "vehicle",
    "id": 2,
```

```
    "name": "bicycle"
  },
  {
    "supercategory": "vehicle",
    "id": 3,
    "name": "car"
  },

}
```

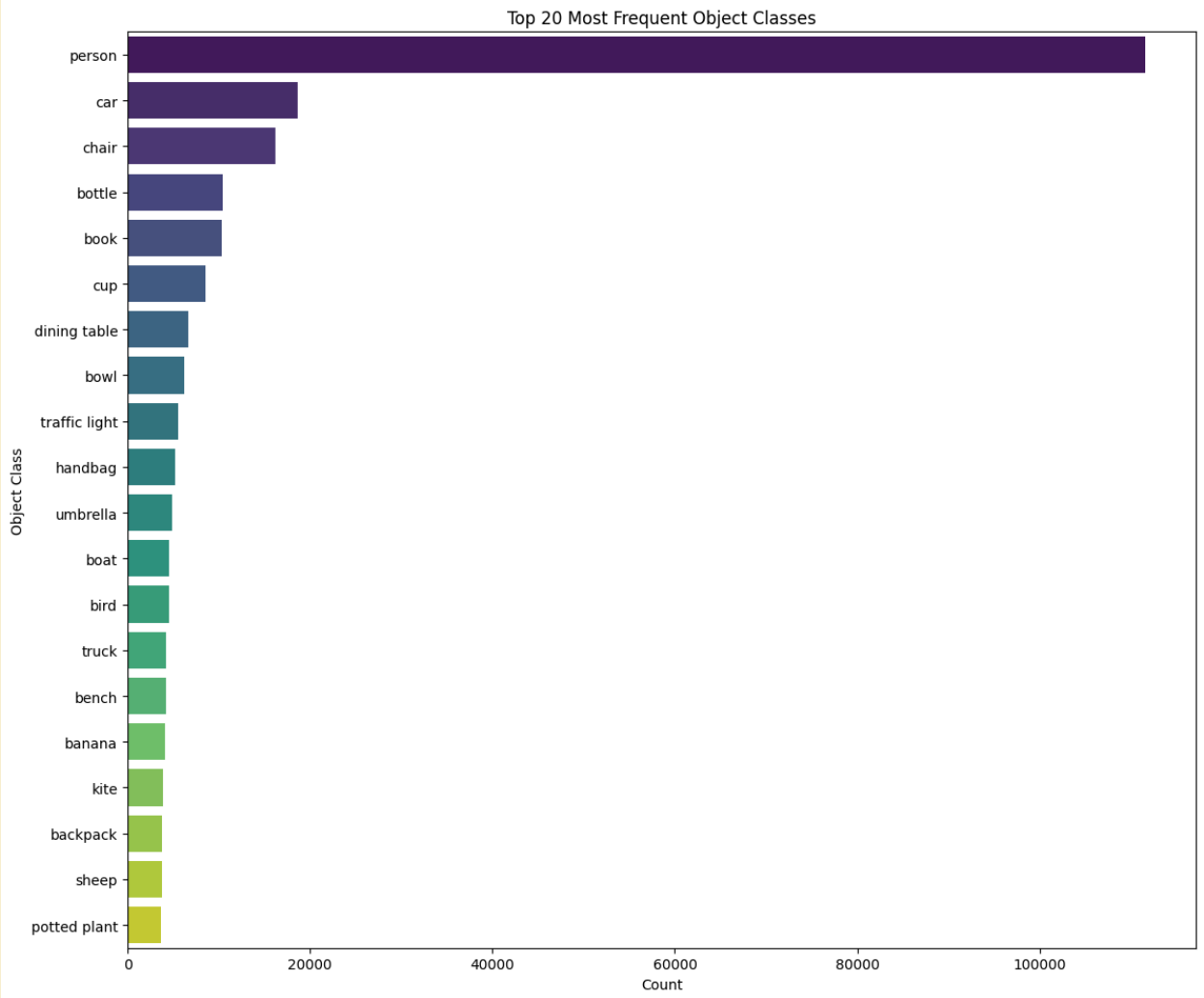

data.yaml configuration:


```
names:
- person
- bicycle
- car
- motorcycle
- airplane
- bus
- train
- truck
- boat
- traffic light
- fire hydrant
- stop sign
- parking meter
- bench
- bird
.
.
.
.

nc: 80
test: path/to/images/test
train: path/to/images/train
val: path/to/images/val
```

5.EDA Results

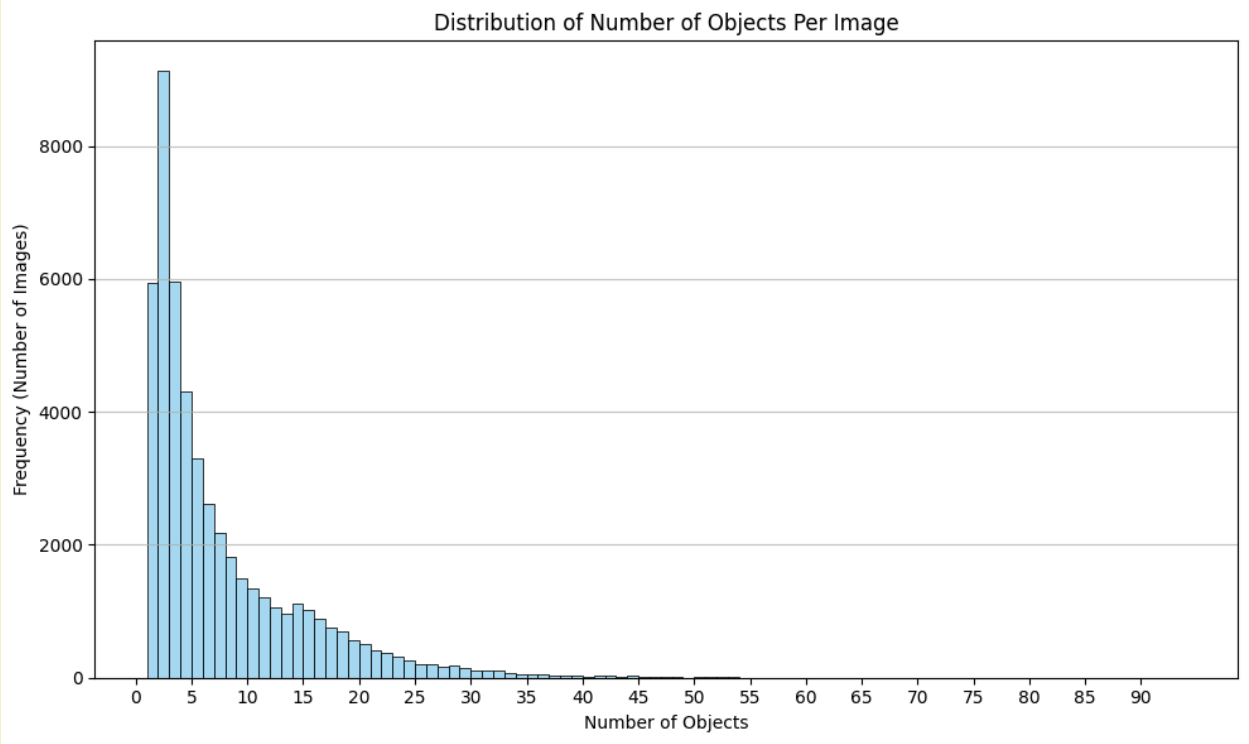
Class Distribution



Insight: Significant Class Imbalance

The COCO dataset exhibits a strong class imbalance, with the person class being overwhelmingly dominant. This means the model will be very good at detecting people, but potentially less reliable for rarer, yet critical, navigation objects like 'traffic light' or 'bench'. To address this, we are supplementing the training data with frames from YouTube POV walkthrough videos. This custom dataset will naturally include more examples of vital navigation objects like benches, stairs, and signs from the first-person perspective, helping to create a more balanced and effective model for our specific use case.

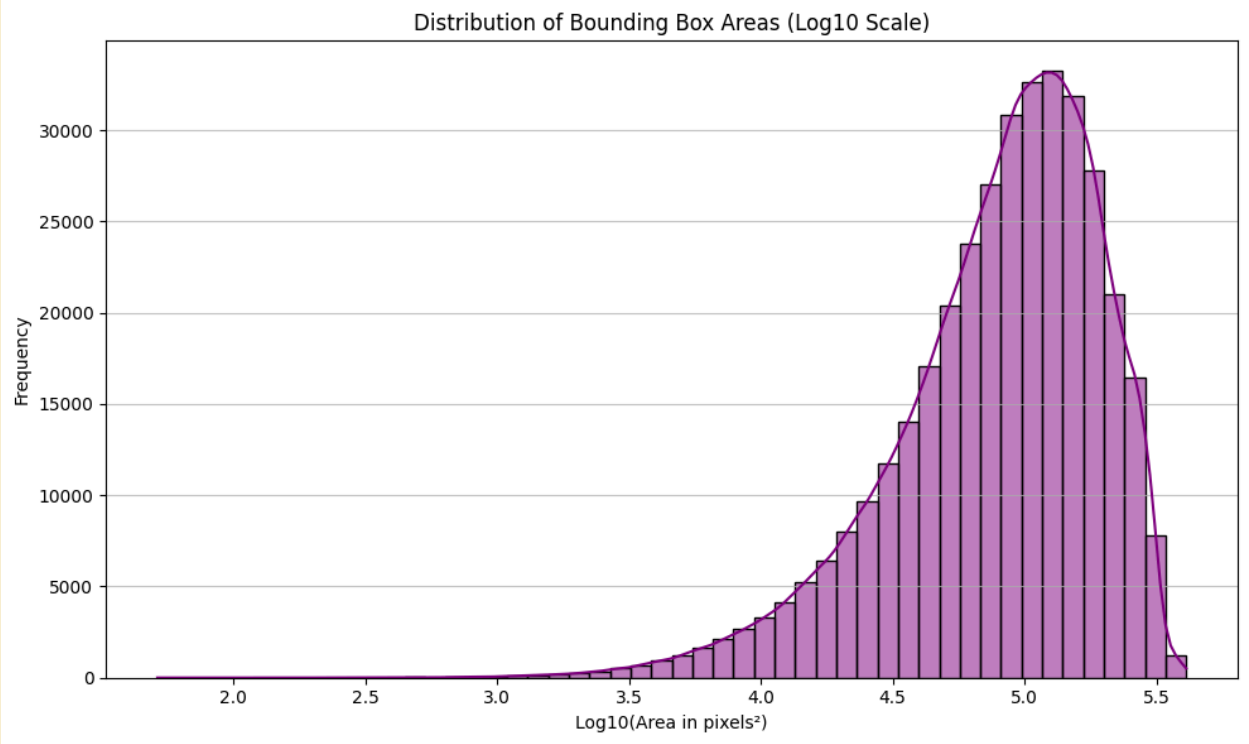
Num. Objects Per Image



Insight: Varied Scene Complexity

The dataset contains a healthy mix of simple and complex scenes. While most images have fewer than 10 objects, the long tail of the distribution shows a significant number of crowded images with up to 90 objects. This is beneficial as it will train the model to perform reliably in both sparse environments (like an empty corridor) and busy ones (like a crowded street), reflecting real-world navigational challenges.

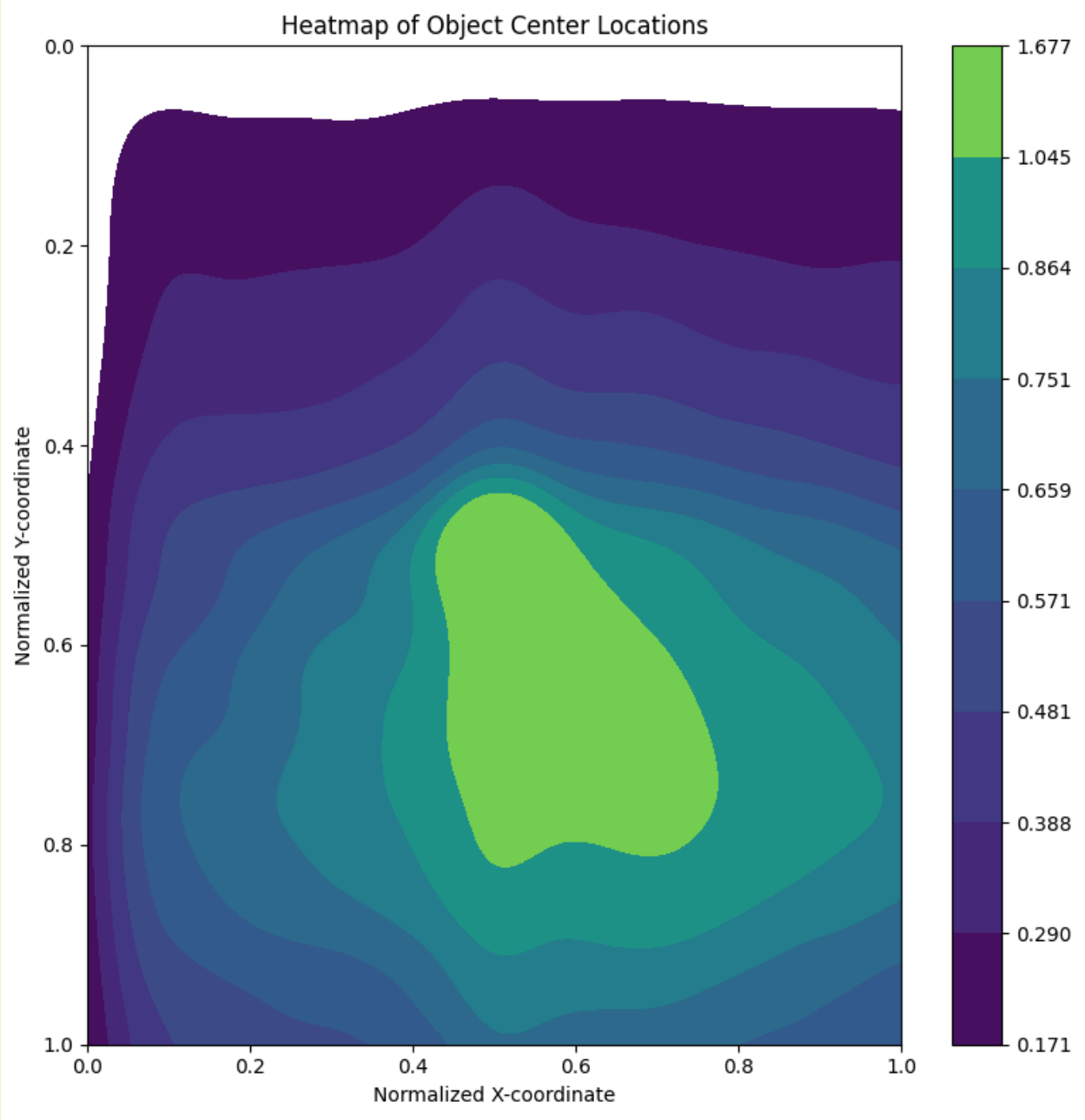
BBOX Areas



Insight: Good Distribution of Object Sizes

The bounding box areas, when viewed on a log scale, show a balanced, bell-shaped distribution. This indicates the model will be trained on a good variety of small, medium, and large objects. For a navigation aid, detecting small objects is crucial as they often represent distant hazards. While the distribution is good, we must carefully evaluate the model's performance on smaller-scale objects to ensure it can provide timely warnings to the user.

Object Locations Heatmap

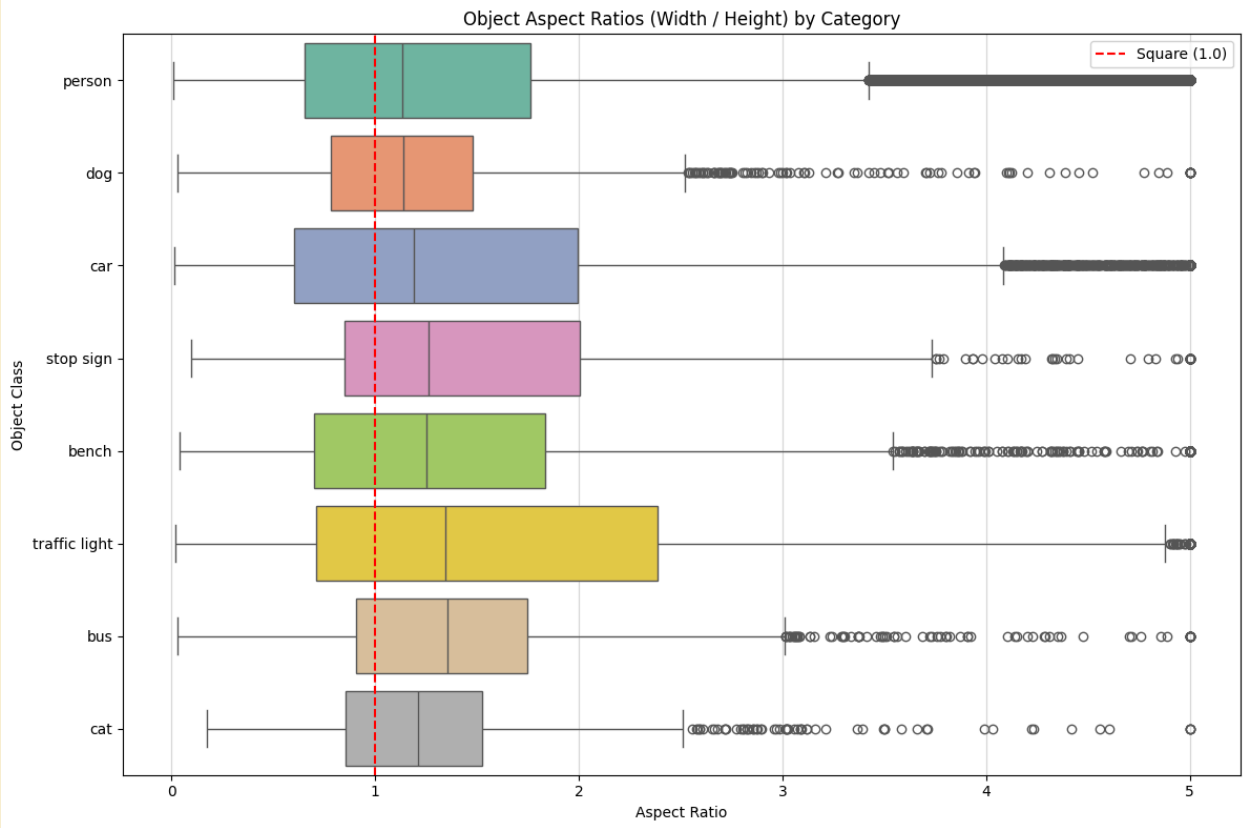


Insight: Central and Lower Bias in Object Locations

The heatmap reveals a significant data bias: objects in the COCO dataset are heavily concentrated in the center and lower half of the image frame. This means the model will be well-trained to detect objects directly ahead but may be unreliable at identifying

hazards in the periphery, especially in the upper regions. This is a critical weakness for a navigation aid, as the model could fail to see overhanging obstacles like signs or tree branches. Our custom dataset must therefore include annotated examples of these high-up hazards to ensure user safety.

Object Aspect Ratios



Insight: Distinct Object Shapes by Category

This plot reveals the distinct shapes of different objects by visualizing their aspect ratios. It confirms that objects like *person* and traffic *light* are typically taller than they are wide (aspect ratio < 1), while *car* and *bus* are wider than they are tall (aspect ratio > 1). This is valuable for diagnosing model errors; for example, a failure to detect a person lying down could be explained by the model being primarily trained on standing (tall) examples. This highlights the need to include varied object poses in our custom dataset to build a more robust model.

5.Future Work

- Extend the dataset with >3000 custom frames from real-world navigation videos.
- Annotate using ReDeTR for consistent bounding box quality.
- Merge custom annotations with COCO subset for final training dataset.
- Upload merged dataset to Google Cloud Storage for integration into Vertex AI training pipeline.

6. Conclusion

A clean, well-structured COCO subset has been successfully prepared and documented. This subset is sufficient for initial experimentation and model fine-tuning under resource constraints. The pipeline is fully scalable for future augmentation and custom annotation integration.

Annexure

Github Repository
Drive Data Storage Links(1, 2, 3)