Final Report: YOLOv8n Indoor Object Detection Model for SafeWalk AI

Project Objective

To develop a lightweight and accurate indoor object detection model using YOLOv8n, tailored to the SafeWalk AI mobile application, which assists visually impaired users in navigating complex indoor environments. This model forms a core module in the end-to-end pipeline:

Camera Input → Scene Classification → Object Detection (Indoor/Outdoor) → OCR(optional) → Depth Estimation → Navigation Logic → TTS Output

---

Dataset Strategy & Preparation

1. Primary Datasets Used for expermints:

- **HomeObjects-3K** (Ultralytics):
    - **Status:** *Picked & Merged*
    - **Size**: ~3,000 images
    - **Classes**: 12 common household objects (e.g., chair, bed, TV, fridge, microwave)
    - **Reason for Selection**: Provided high-quality annotations for frequently encountered indoor objects essential for visually impaired navigation.
- **Roboflow Indoor-Spoor:**
    - **Status:** *Picked & Merged*
    - **Size**: ~500+ images
    - **Classes**: Includes common indoor elements with an added **"person"** class
    - **Reason for Selection**: Helped incorporate human presence detection, which is critical for safe pathfinding in indoor environments.
- **Roboflow Washroom-RF1FA:**
    - **Status:** *Picked & Merged*
    - **Size**: ~700+ images
    - **Classes**: Sink, toilet, mirror, towel, hygiene-related fixtures

- **Reason for Selection**: Targeted detection of washroom elements, enhancing the system's awareness of essential sanitary facilities.

- **Indoor Object Detection Dataset (Kaggle):**
  - **Size**: 1,349 images
  - **Classes**: 10 (e.g., door, chair, cabinet)
  - **Status**: *Dropped*
  - **Reason**: While it was a decent dataset, we decided to search for or construct a more tailored dataset that aligns more closely with the specific indoor navigation needs of our project.

- **OpenImages V7 (Filtered for choosen indoor objects):**
  - **Size**: Subset of 5,000 images
  - **Classes**: 28 indoor-related classes
  - **Status**: *Dropped*
  - **Reason**: Although the dataset offers a large class variety and image volume, it requires extensive preprocessing. Additionally, most of the classes are not relevant to our use case, making it difficult to maintain and justify its inclusion.

- **NYU Depth V2:**
  - **Size:**
  - **Classes**: RGB-D dataset (includes depth information alongside RGB images)(more info through link)
  - **Status**: *Dropped*
  - **Reason**: Initially selected for its depth-rich indoor scenes, but dropped due to challenges including:
    - Lack of YOLO-formatted annotations
    - Limited object class diversity aligned with our target indoor classes
    - Dataset more suited for segmentation and depth tasks than object detection
    - Additional effort required to convert depth data and align it with our object detection pipeline

- **Imperial UAV Indoor Dataset:**

- **Size:** Around 13,000 RGB frames
- **Description**: Drone-perspective images capturing indoor obstacles, intended for augmenting ceiling and top-view data
- **Status**: *Dropped*
- **Reason**: The drone's awkward and unusual viewing angles do not correspond to the typical perspective of a visually impaired person navigating indoors, making the data less relevant. Other issues such as inconsistent object visibility and limited practical applicability also contributed to the decision.

- **2. Dataset Merging Process:**
- All picked datasets were provided(downloaded) in YOLOv8 format so no conversion was needed
- Unified into a merged dataset with 19 total indoor classes
- Folder structure:

/merged_dataset_final

├── images/

| ├── train/

| └── val/

└── labels/

├── train/

└── val/

- Mapped class names and removed irrelevant/duplicate labels (e.g., "geyser", "c")
- Created merged_data.yaml with 19 class names

Model Training Pipeline(for the picked model)

Framework: Ultralytics YOLOv8

Model Base: yolov8n.pt (Nano version for speed + mobile optimization)

Notes: *The model was first fine tuned on HomeObjects-3K but then we decided to add more classes since it felt the covered classes are great but not enough even with the new classes*

added there are still a lot of hazards a blind person needs to be notified of this is why those models are always open for more improvements.

Training Command:

```
model.train(
   data="/kaggle/working/merged_dataset_final/merged_data.yaml",
   epochs=100,
   imgsz=640,
   batch=16,
   device=0,
   freeze=10,
   patience=20
)
```

Settings & Features:

- Mixed Precision (AMP)
- Early stopping (patience=20)
- Layer freezing to retain base COCO knowledge
- Fine-tuned on merged dataset for better generalization

Challenges & Solutions

Challenge	Solution
Dataset format mismatches	Custom remapping scripts for unified YOLO format
Missing classes (e.g., elevator)	Marked for future annotation using LabelImg also much more classes need to be addressed like hand railings , stoves and refrigerators in the kitchen
Underfitting with YOLOv8n	Used longer training, data augmentation, and class balancing

Challenge	Solution
FileNotFoundError on Kaggle	Provided full path to YAML file instead of shorthand name
torchsummary crash	Used model.info() instead
Catastrophic forgetting	When we fine tuned the model on homeobjects-3k then finetuned for the second time on the merged datasets from roboflow(without homeobjects) it captured the classes on the merged dataset forgetting the one in homeobject-3k completly

---

Evaluation Metrics

Validation Results (Sample):

metrics = model.val(data="/kaggle/working/merged_dataset_final/merged_data.yaml")

- **mAP@0.5**: 62.25%
- **mAP@0.5:0.95**: 40%
- **Precision**: 75%
- **Recall**: 60%

Other Experiment Results:

Dataset	Classes	Model	mAP@0.5	Notes
Indoor Object Detection (Kaggle)	10	YOLOv8s	~75%	Good initial performance, fast setup
HomeObjects-3K	12	YOLOv8n	~70%	High diversity but class imbalance present
HomeObjects-3K then on the combined Roboflow datasets	19	YOLOv8n	~60%	Did great on the newly introduced classes but was unable to detect the 12 initial classes due to those classes not being present in the 2nd fine tuning phase

				causing catastrophic forgetting .
Merged Final Dataset(HomeObjects+the 2 Roboflow datasets)	19	YOLOv8n	~62.25%	Highest performance across classes finetuned using the merged dataset as a whole on the 1st fine tuning attempt. The chosen one for mobile deployment
Merged Final Dataset(HomeObjects+the 2 Roboflow datasets)	19	YOLOv8s	~65.25	Same as the above a small improvement in mapping precision and recall but a slightly heavier model and a better choice if we can afford the increase in size by 10 to 30 MB.

Inference and Visualizations:

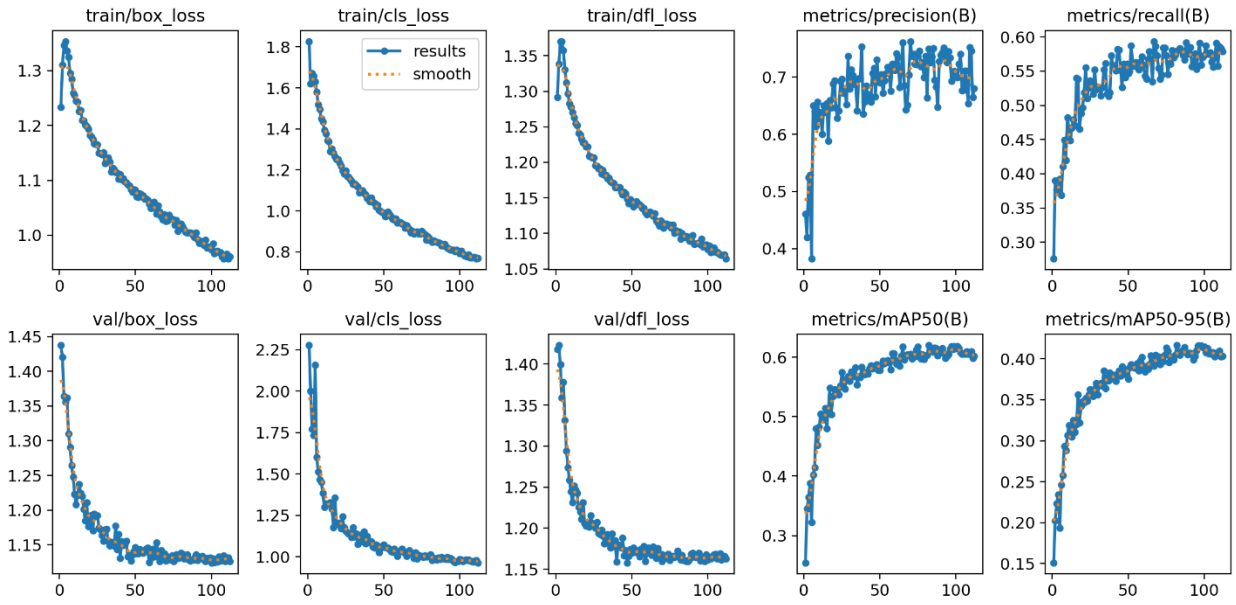
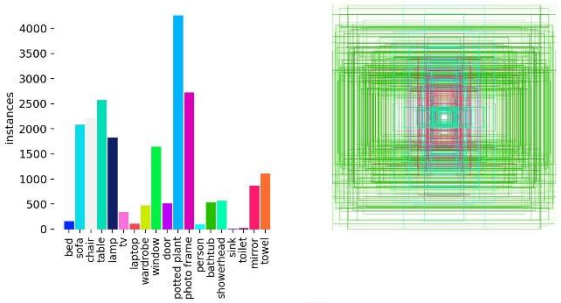
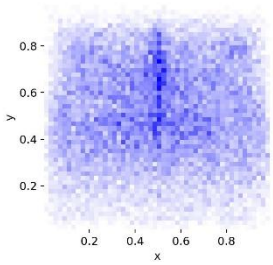
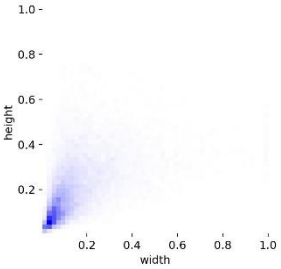
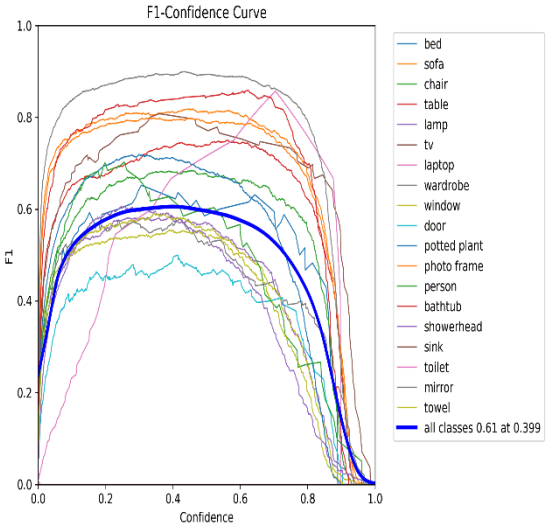


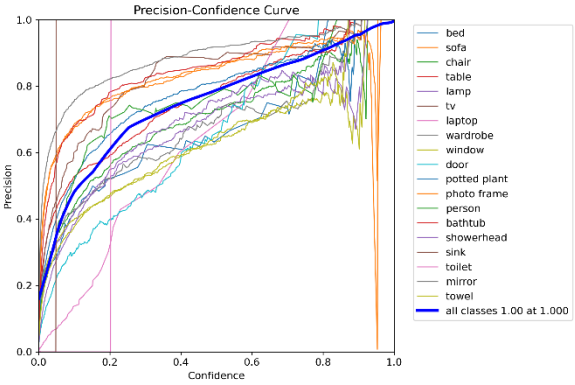
Figure 1-metrics_graphs

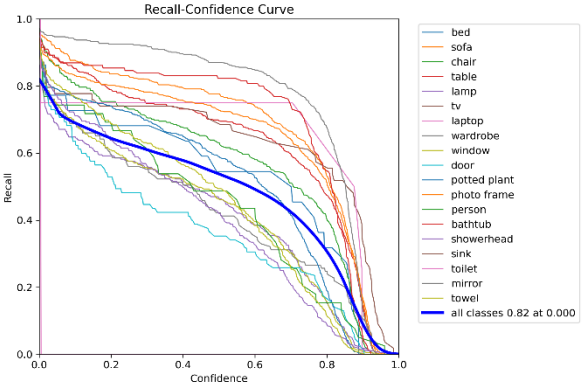

instances
bed
sofa
chair
table
lamp
tv
laptop
wardrobe
window
door
potted plant
photo frame
person
bathtub
showerhead
sink
toilet
mirror
towel

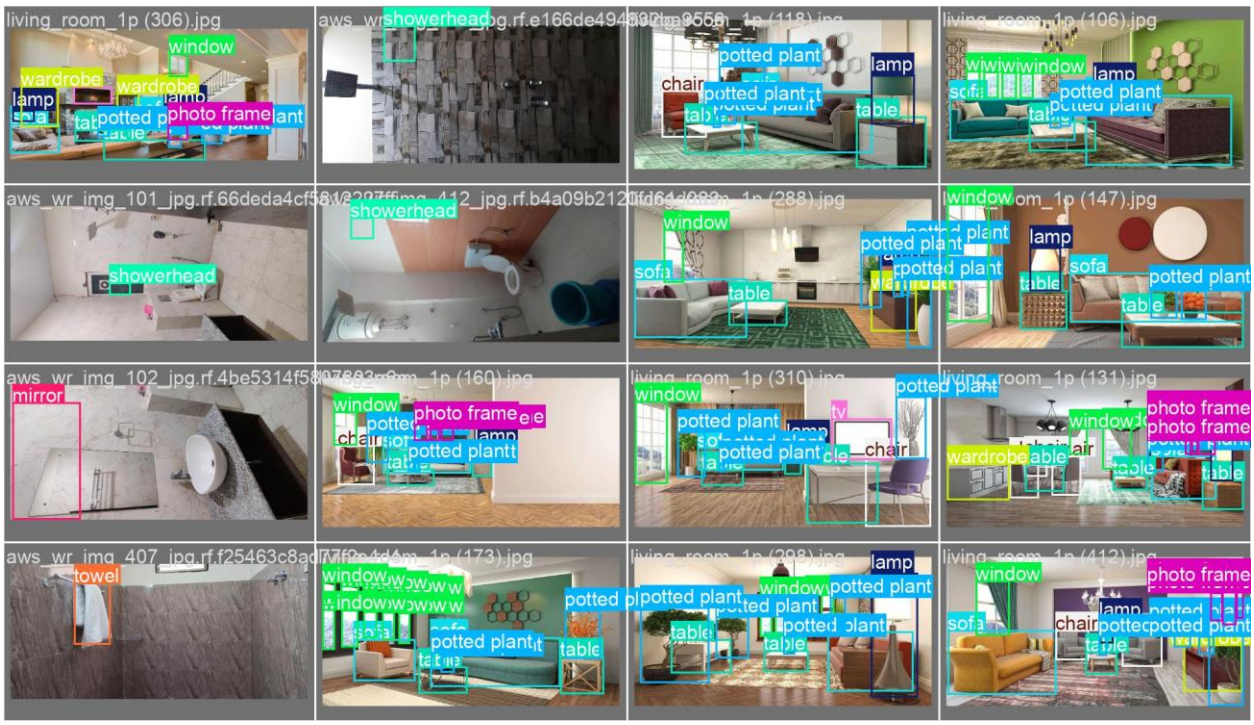

y
x

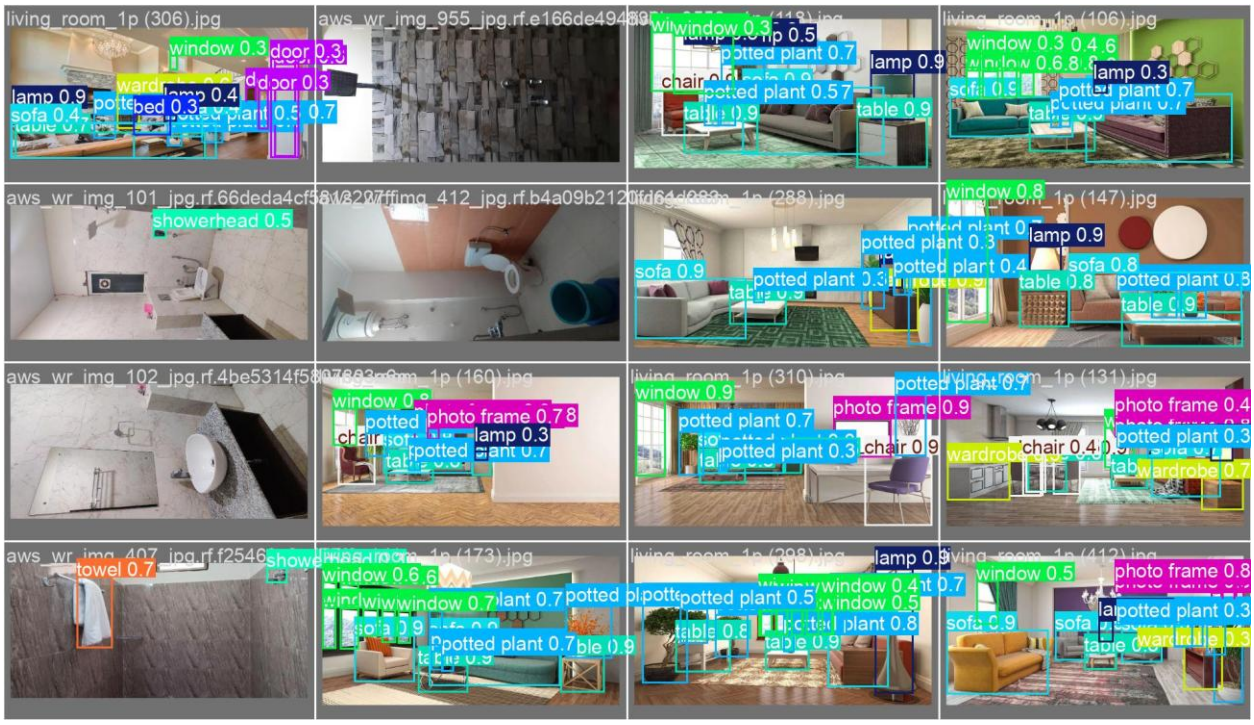

height
width

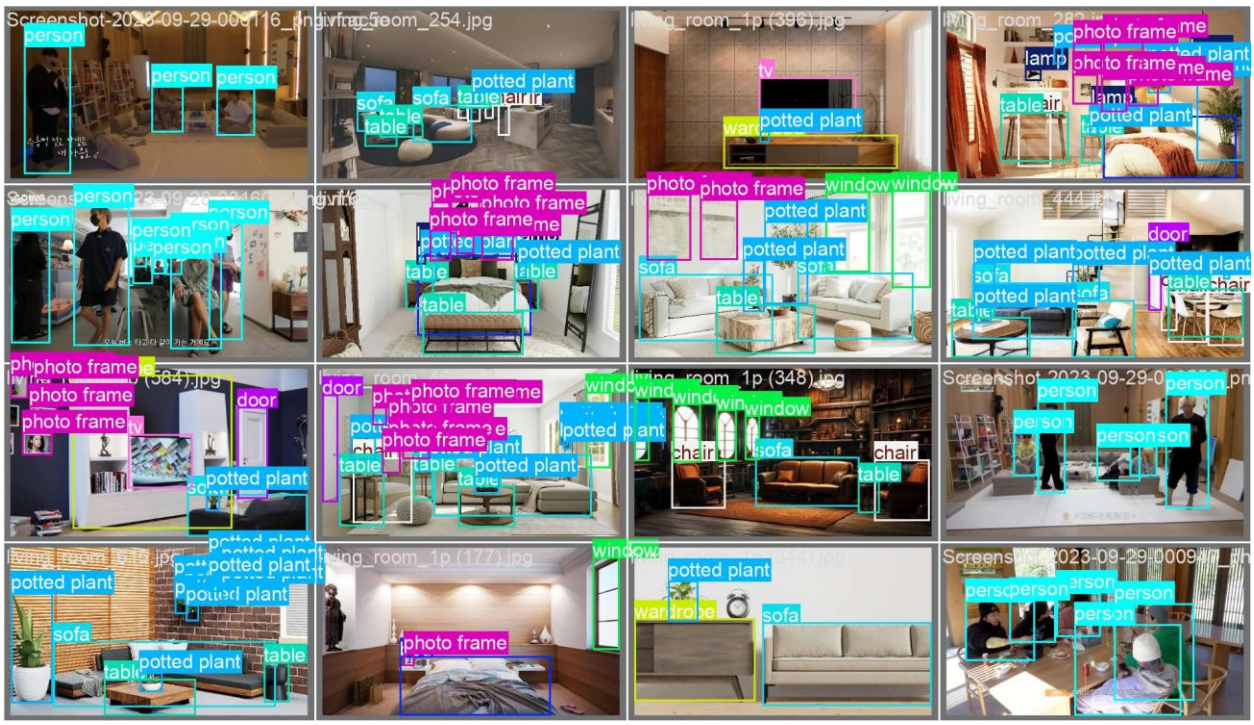

F1-Confidence Curve
F1
Confidence
bed
sofa
chair
table
lamp
tv
laptop
wardrobe
window
door
potted plant
photo frame
person
bathtub
showerhead
sink
toilet
mirror
towel
all classes 0.61 at 0.399


Precision-Confidence Curve
Precision
Confidence
bed
sofa
chair
table
lamp
tv
laptop
wardrobe
window
door
potted plant
photo frame
person
bathtub
showerhead
sink
toilet
mirror
towel
all classes 1.00 at 1.000


Recall-Confidence Curve
Recall
Confidence
bed
sofa
chair
table
lamp
tv
laptop
wardrobe
window
door
potted plant
photo frame
person
bathtub
showerhead
sink
toilet
mirror
towel
all classes 0.82 at 0.000


living_room_1p (306).jpg
window
wardrobe
wardrobe
lamp
photo frame
aws_wr showerhead jpg.rf.e166de494
potted plant
lamp
chair
table
living_room_1p (106).jpg
window
lamp
potted plant
aws_wr_img_101_jpg.rf.66deda4cf
showerhead
showerhead
window
potted plant
sofa
table
window
lamp
potted plant
sofa
table
aws_wr_img_102_jpg.rf.4be5314f
mirror
window
photo frame
potted plant
potted plant
chair
living_room_1p (131).jpg
photo frame
photo frame
wardrobe
aws_wr_img_407_jpg.rf.f25463c8
towel
window
potted plant
potted plant
table
window
potted plant
potted plant
chair
photo frame


living_room_1p (306).jpg
window 0.3
door 0.3
lamp 0.9
lamp 0.4
bed 0.3
aws_wr_img_955_jpg.rf.e166de494
window 0.3
lamp 0.5
potted plant 0.7
lamp 0.9
chair
potted plant 0.5
table 0.9
living_room_1p (106).jpg
window 0.3
lamp 0.3
potted plant 0.7
aws_wr_img_101_jpg.rf.66deda4cf
showerhead 0.5
sofa 0.9
potted plant 0.3
potted plant 0.4
window 0.8
lamp 0.9
sofa 0.8
table 0.8
potted plant 0.3
aws_wr_img_102_jpg.rf.4be5314f
window 0.9
photo frame 0.78
lamp 0.3
window 0.9
potted plant 0.7
photo frame 0.9
chair 0.9
photo frame 0.4
potted plant 0.3
wardrobe 0.7
aws_wr_img_407_jpg.rf.f2546
towel 0.7
window 0.6
window 0.7
potted plant 0.7
potted plant 0.5
window 0.4
window 0.5
potted plant 0.8
window 0.5
photo frame 0.8
potted plant 0.3
sofa 0.9


Screenshot-2023-09-29-000116_jpg.rf.5e
living_room_254.jpg
living_room_1p (396).jpg
living_room_282.jpg
living_room_1p (384).jpg
living_room_1p (348).jpg
Screenshot 2023-09-29-0
living_room_1p (177).jpg
Screenshot 2023-09-29-0009


Screenshot-2023-09-29-000116_jpg.rf.5e
living_room_254.jpg
living_room_1p (396).jpg
living_room_282.jpg
living_room_1p (384).jpg
living_room_1p (348).jpg
Screenshot 2023-09-29-0
living_room_1p (177).jpg
Screenshot 2023-09-29-0009


Screenshot-2023-09-28-231207_
person
person
living_room_1p (523).jpg
photo frame
window
sofa
wardrobe
photo frame
photo frame
lamp
sofa
sofa
living_room_1p (416).jpg
potted plant
sofa
living_room_1p (553).jpg
tv
potted plant
table
aws_wr_img_711_jpg.rf.23b018795
bathtub
photo frame
photo frame
photo frame
potted plant
sofa
tv
wardrobe
living_room_538.jpg
photo frame
table
table
living_room_1p (35).jpg
window
photo frame
photo frame
table
living_room_1p (18).jpg
door
photo frame
photo frame
wardrobe
wardrobe
table
door
door
door
lamp
lamp
lamp
potted plant
sofa
table
aws_wr_img_92
showerhead
mirror
living_room_1p (434).jpg
photo frame
photo frame
potted plant
sofa
living_room_1p (182).jpg
photo frame
chair
chair
table
photo frame
photo frame
potted plant
potted plant
showerhead
mirror
Screenshot-2023-09-28-231207_
person 0.8
person 0.3
living_room_1p (523).jpg
window 0.5
door 0.3
sofa 0.5
wardrobe 0.4
door 0.4
photo frame 0.4
sofa 0.3
table 0.4
living_room_1p (4
potted plant 0.8
tv 0.8
living_room_1p (553).jpg
lamp 0.6
tv 0.9
chair 0.3
wardrobe 0.3
aws_wr_img_711_jpg.rf.23b018795
bathtub 0.8
photo frame 0.8
photo frame 0.3
potted plant 0.8
sofa 0.8
tv 1.0
wardrobe 0.9
living_room_538.jpg
living_room_1p (35).jpg
photo frame 0.6
potted plant 0.5
table 0.8
wardrobe 0.9
wardrobe 0.3
door 0.8
tv 0.9
table 0.8
window 0.5
window 0.6
lamp 0.7
lamp 0.7
table 0.7
potted plant 0.5
chair 0.5
showerhead 0.3
irror 0.8
living_room_1p (434).jpg
tv 0.9
tv 0.3
window 0.3
potted plant 0.7
sofa 0.8
living_room_1p (182).jpg
wardrobe 0.4
table 0.3
sofa 0.3
lamp 0.3
chair 0.4
table 0.6
photo frame 0.8
plant 0.6
showerhead 0.4
showerhead 0.6
mirror 0.8

Improvements on the model

We can see in the visualizations above especially the second images that there is a class imbalance in the dataset now what we can do we can search for datasets that contain those classes(bed, person, sink, toilet, and laptop) so that they have more appearances in training improving our metrics(precision, recall, mapping) especially for those minority classes.

Outcomes & Learnings

- Successfully built and evaluated a 19-class indoor object detection model
- Achieved high precision and recall while maintaining a lightweight architecture
- Integrated multiple datasets into a cohesive training set
- Addressed key deployment issues such as file paths, model size, and prediction storage

---

Final Note

This indoor object detection model is production-ready for integration into our SafeWalk app. It provides a solid base for obstacle detection and will support further enhancements such as real-time navigation, multimodal input, and robust edge deployment.