naviVision

Sai Suhas Vemuri
Pranav Souri Itabada
Pragathi Vemulapalli
Preetham Rao Gottumukula

Intro

Limited spatial awareness can make navigation difficult for blind people.

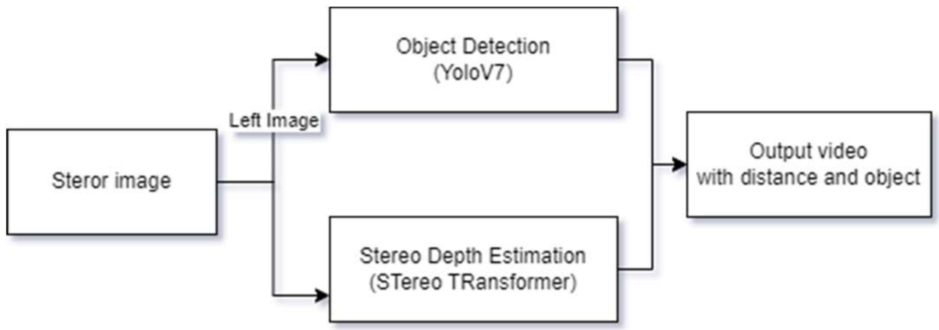
Existing assistive technologies (e.g., guide dogs, white canes) are limited in their ability to detect and inform of hazards ahead.

Advanced methods like LIDAR cameras can be expensive and not easily accessible for all.

Our project aims to provide a cost-effective solution using computer vision techniques, like object detection and stereo depth estimation

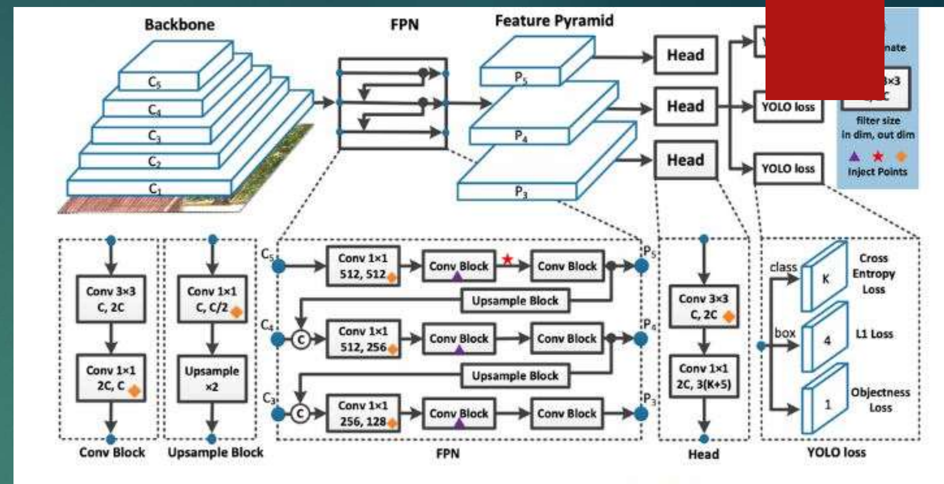
Methodology

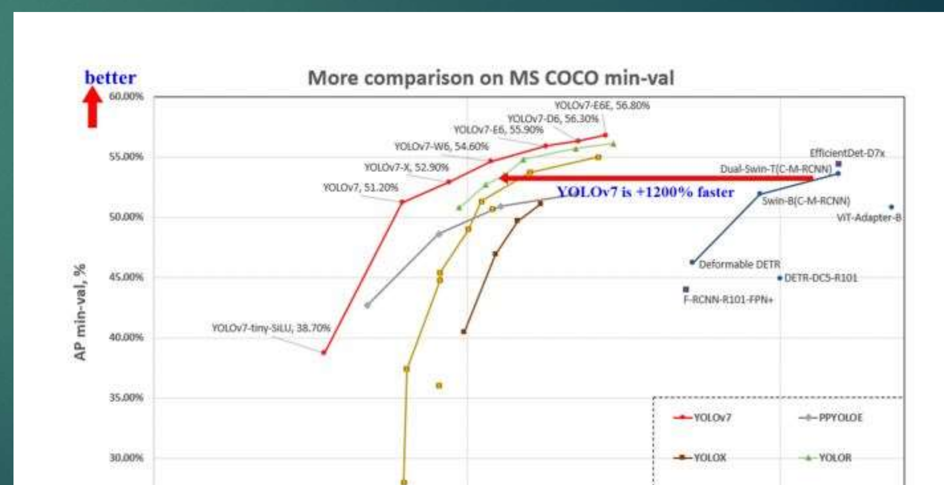
- Our approach is to use an Application-oriented method.
- We use YOLOv7 for object detection and STereo TRansformer for stereo depth estimation.
- We apply YOLOv7 and STereo TRansformer to the images to get the objects in frame and their depths.
- The combination of these techniques provides an effective and cost-efficient solution for navigation assistance for blind people.



Object Detection

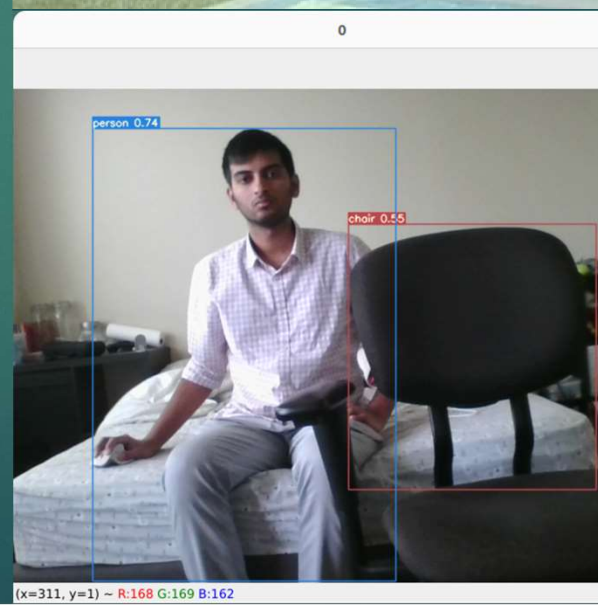
- We constructed a model for object detection and trained it using YOLOv7, an advanced and improved object detection algorithm.
- YOLOv7 is provides a good balance between accuracy and speed

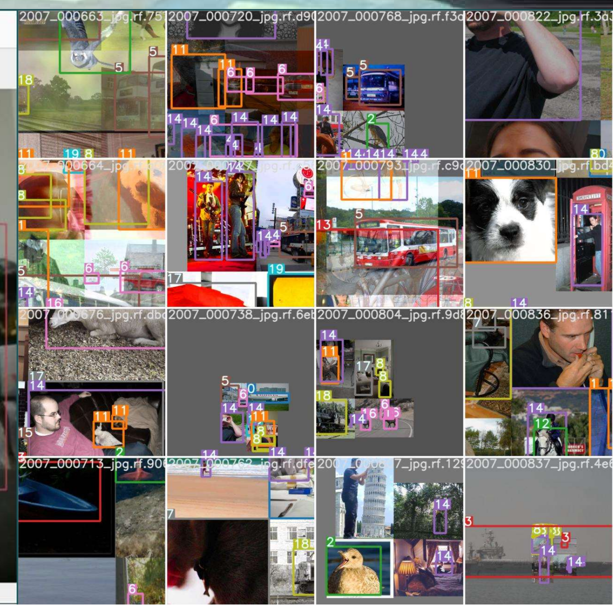




Datasets

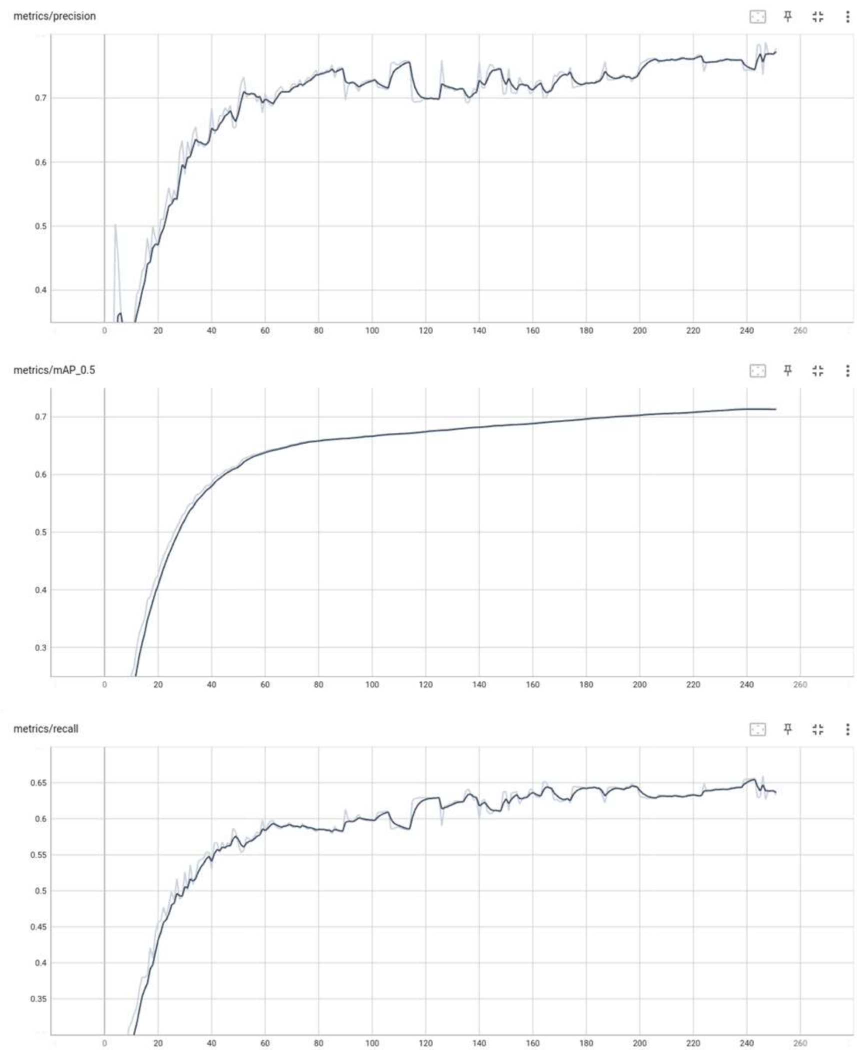
- Initially we used Microsoft COCO dataset for general object detection
- Then using this model as base, we trained on the KITTI object detection dataset for navigation





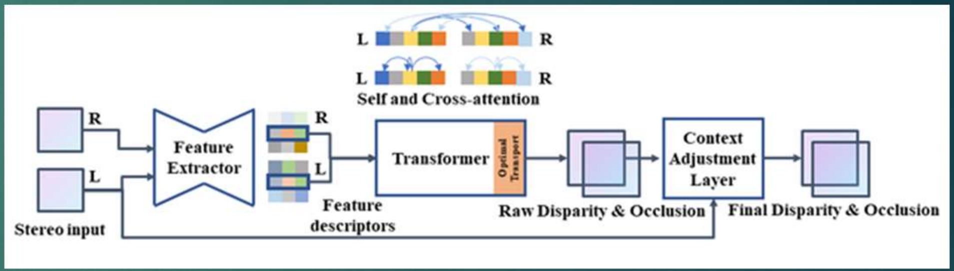
Experiments (Object Detection)

- We used recall, precision and mAp to evaluate our model.
- Once the metrics had plateaued, we stopped training the model.



Stereo Depth Estimation

- Initially we thought of using Bi3d, but since it was not open-source we moved onto STereo Transformer
- STTR has the advantage of occlusion handling, can handle scenes with large depth variations, and generalizes across different domains



Methodology

- We trained the model on Scenflow, KITTI stereo 2015, and driving stereo datasets.
- STTR generates the disparity map, and from that we can get the depth using the formula:

$$depth = \frac{baseline \times focal\ length}{disparity}$$

Experiments (Stereo Depth)

Our Model:

3px Error	EPE	Occ IOU
8.12	1.67	0.91

STTR model:

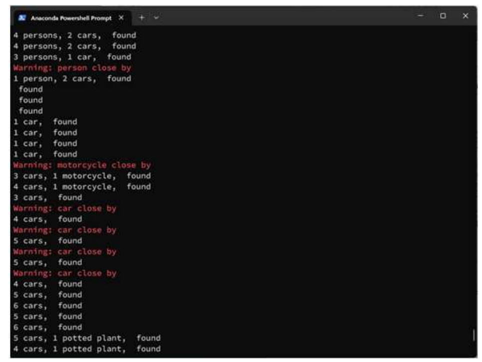
3px Error	EPE	Occ IOU
6.74	1.50	0.98

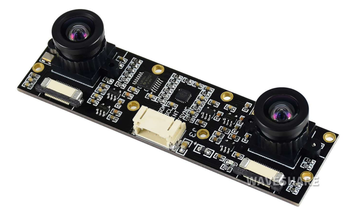
Demo video


car(dist:66.73)
car(dist:67.64)
car(dist:63.02)
car(dist:42.35)
car(dist:36.05)
car(dist:29.88)
car(dist:28.02)

Future Enhancements

- We can mount a stereo camera, and eye tracker to the users head. So whatever they point their eye to can be read out by the application.
- Collect better data specifically for navigation while walking.





Conclusion

- We aimed to create a cheap alternative
- We achieved this using modern technologies in computer vision such as Object detection and Stereo depth estimation.
- Able to fine tune the model and train it with multiple datasets so as to accommodate the user in any condition.
- With the finished model/project, we were able to provide real time navigation.
- Future enhancements would lead to a real world product.