Ahmedabad University

CSE541 Computer Vision

Object detection techniques in case of small objects

1. Paridhi Jain AU2120226
2. Kathan Dave AU2140113
3. Saahil Doshi AU2140108
4. Jash Parikh AU2140106

Problem Statement

- Small Object Detection
- Deep Learning Models
- Performance metrics will be mAP@0.5, mAP@0.75, mAP@0.95, and confusion matrix.

Figure 1: Vis Drone Dataset image (Retrieved from Zhu et al.)

Datasets Discussion

Training Dataset

VisDrone-DET dataset

288 video clips formed by 261,908 frames and 10,209 static images.

by the AISKYEYE team at Lab of Machine Learning and Data Mining , Tianjin University, China.

Primary Work to be done on images.

Images and Annotations are provided.

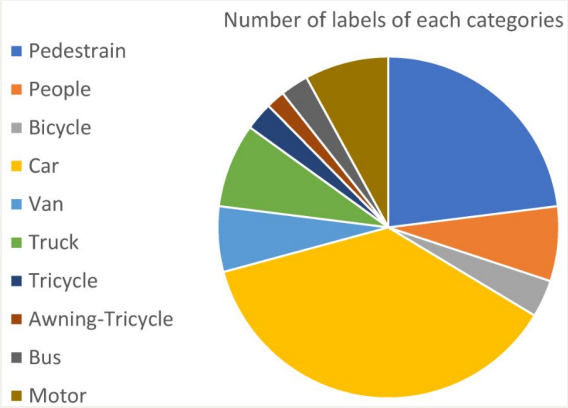

Figure 3: Number of instances of labels for each category

VisDrone-DET dataset

- trainset (1.44 GB): BaiduYun | GoogleDrive
- valset (0.07 GB): BaiduYun | GoogleDrive
- testset-dev (0.28 GB): BaiduYun | GoogleDrive (GT avalialbe)
- testset-challenge (0.28 GB): BaiduYun | GoogleDrive

Figure 2: VisDrone Dataset Details(retrieved from VisDrone Repsoitory)

```
names:
  0: pedestrian
  1: people
  2: bicycle
  3: car
  4: van
  5: truck
  6: tricycle
  7: awning-tricycle
  8: bus
  9: motor
```

Why HIC-YOLO V5?

- A modified version of YOLO-V5
- Real-time detection
- Less Computational overhead.
- Ease of use.
- Scope of Improvement.

Working of YOLO v5

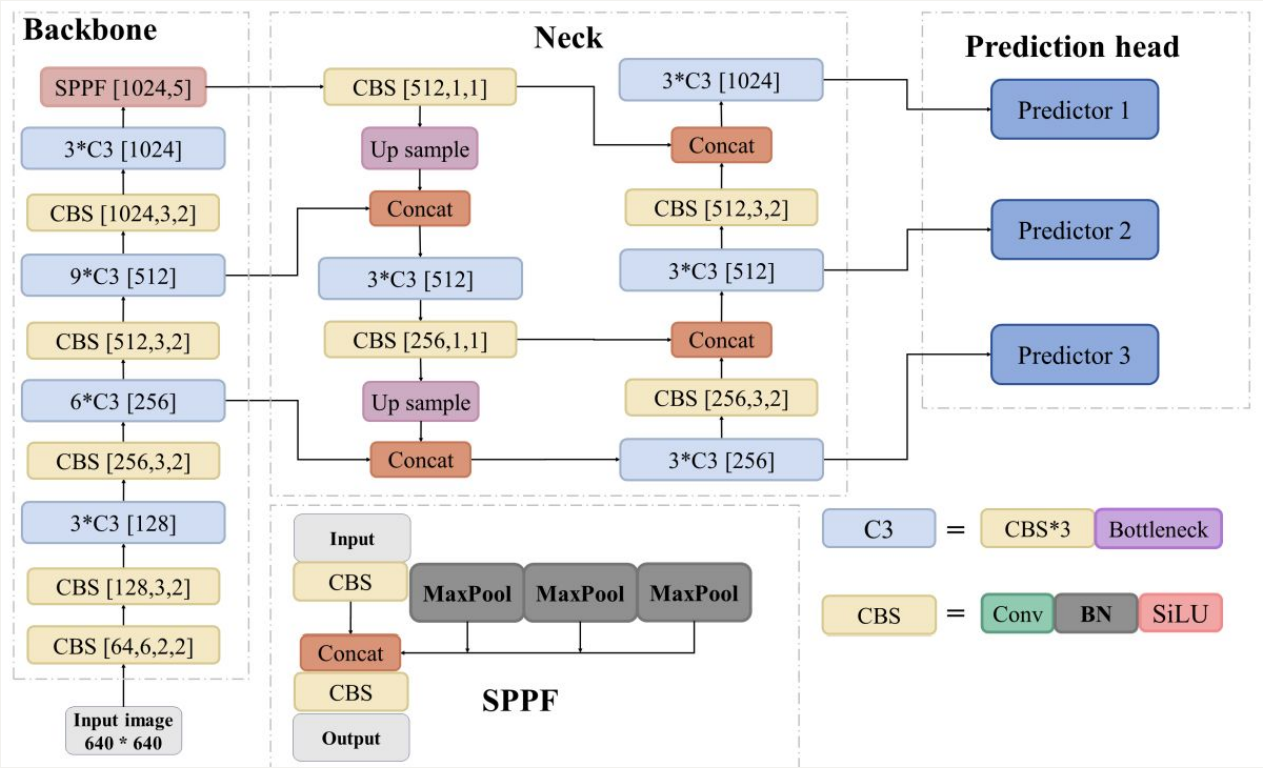

Figure 4: Structure of YOLOv5 (Retrieved from Shiyi Tang et al.,2023)

HIC-YOLOV5(Head, Involution and CBAM-YOLOv5)

- Small Object Prediction Head
- Involution Block Integration.
- CBAM Attention Module
- Selective Feature Extraction
- Improved Object Detection
- Efficient Model Design

HIC-YOLO V5

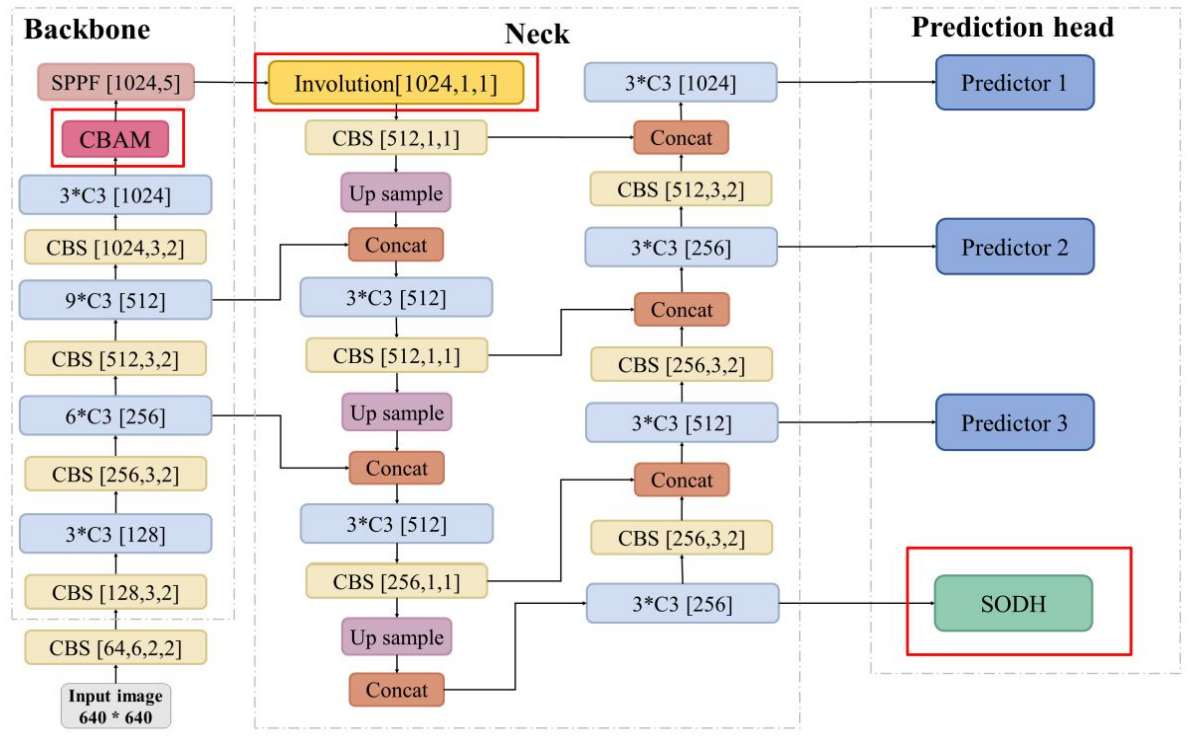

Figure 5: Structure of HIC-YOLOv5 (Retrieved from Shiyi Tang et al., 2023)

CBAM

Convolutional Block Attention Module

- Standard CBAM Placement
- HIC-YOLOv5's CBAM Placement
- Focus on Feature Extraction
- Reduced Computational Cost
- Dual Attention Mechanism

CBAM

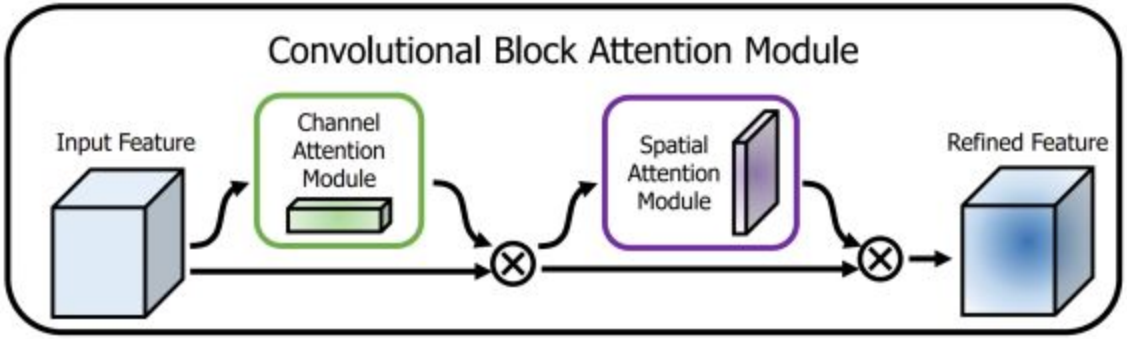

Figure 6: Structure of CBAM(Retrieved from Shiyi Tang et al.,2023)

CFFI

Channel Feature Fusion with Involution (CFFI)

- An inverse of convolution!
- Involution is **Channel-agonist** and **Spatially-specific**.
- would help to preserve the features of small objects.

CFFI

Channel Feature Fusion with Involution (CFFI)

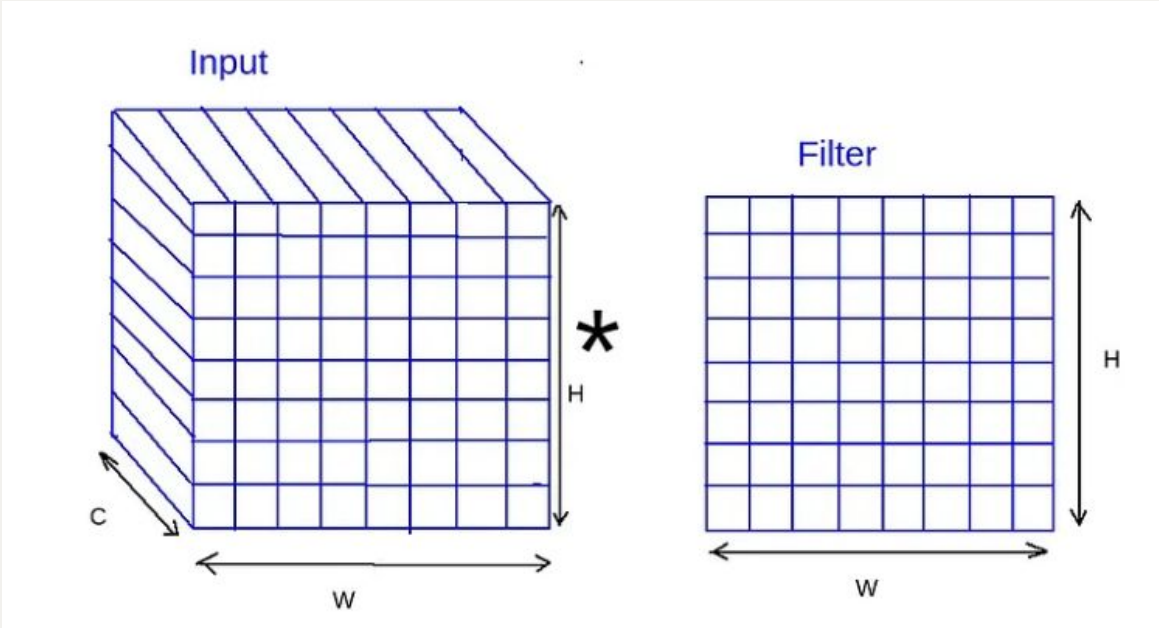

Figure 7: Structure of CFFI(Retrieved from "Involution: Inverting the Inherence of Convolution for Visual Recognition", Medium)

CFFI

Channel Feature Fusion with Involution (CFFI)

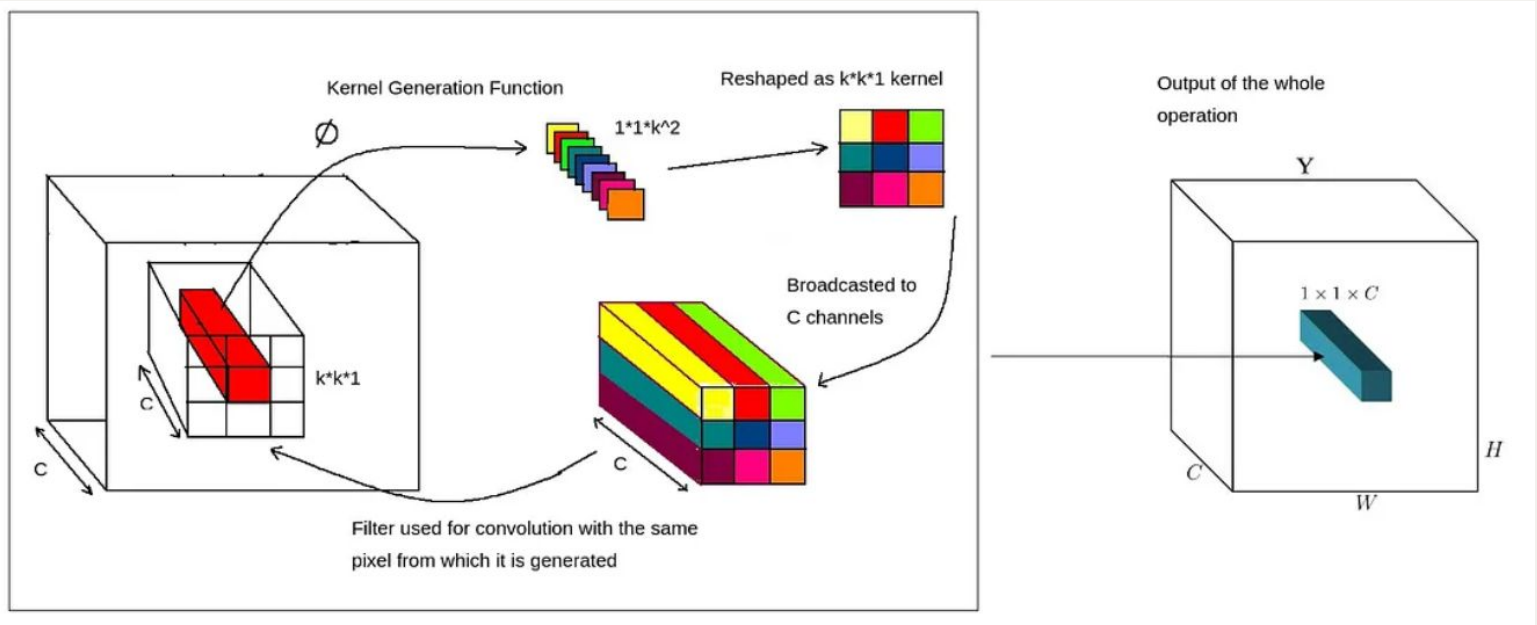

Figure 8: Structure of CFFI(Retrieved from "Involution: Inverting the Inherence of Convolution for Visual Recognition", Medium)

Prediction Heads

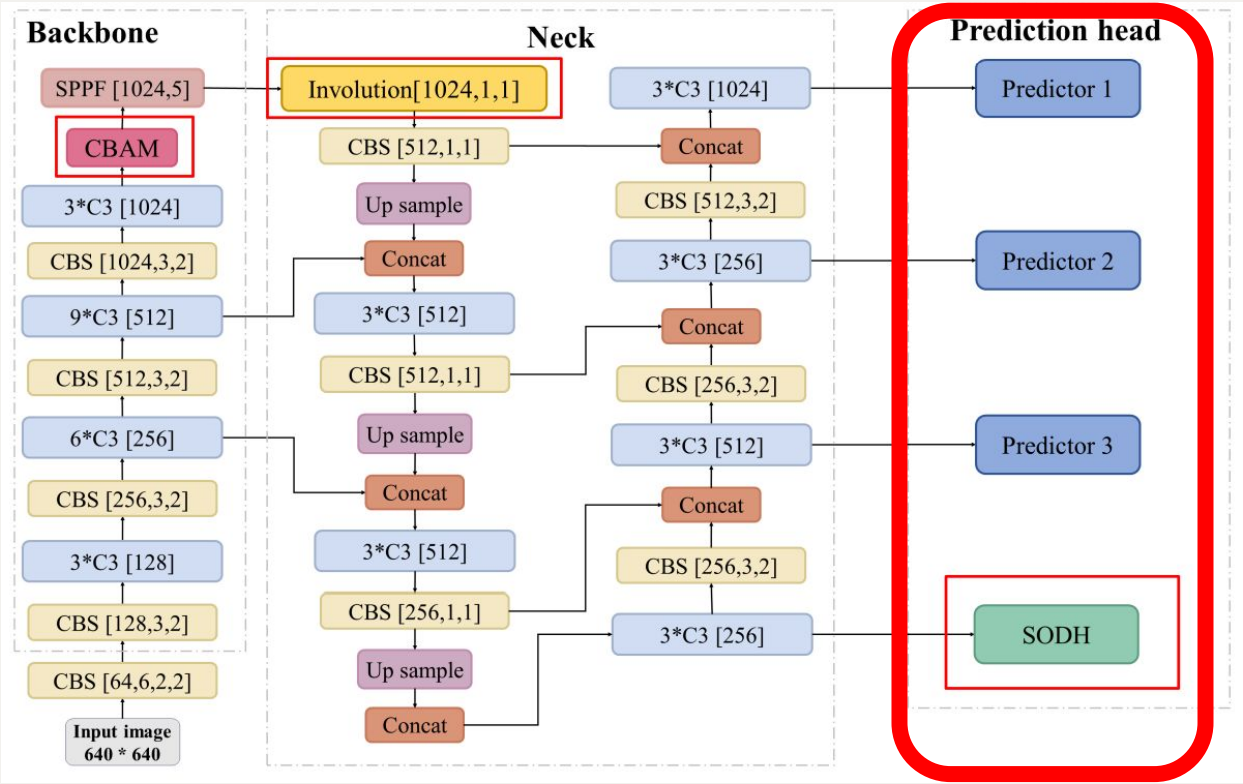

Figure 9: Structure of HIC-YOLOv5 (Retrieved from Shiyi Tang et al., 2023)

Training

- We trained 3 models YOLOv5, YOLOv8 and HIC-YOLOv5.
- Epochs:10
- Train size:- 6000 images
- Val Size:- 500 images
- Test Size:- 1600 images
- Batch Size:- 32

Results

- We trained three models, YOLOv8, YOLOv5 and HIC-YOLOv5

Object Class	YOLOv8	YOLOv5	HIC-yolov5
All	25.4%	29.3%	30.51%
Pedestrian	28.4%	42.7%	46.14%
People	22.7%	36%	27.65%
Bicycle	5.61%	18.7%	2.86%
Car	66.5%	71%	81.97%
Van	31.9%	28.5%	36.84%
Truck	22.8%	25.4%	22.84%
Tricycle	16.4%	0%	12.65%
Awning-tricycle	8.68%	0%	7.31%
Bus	22.2%	31.8%	24.74%
Motor	28.8%	39.1%	42.06%

TABLE I: COMPARISON OF OBJECT DETECTION MODELS ON VAL DATASET - MAP@0.5

Results

YOLOv5 vs HIC-YOLOv5

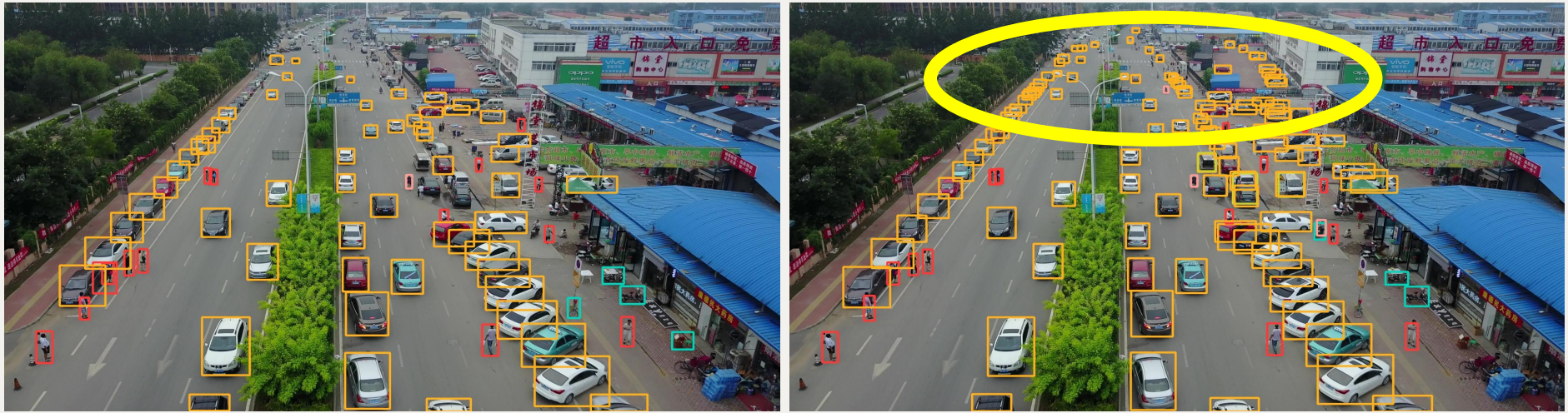

Figure 10: Comparison of Object Detection by YOLOv5 and HIC-YOLOv5

Results

YOLOv5 vs HIC-YOLOv5

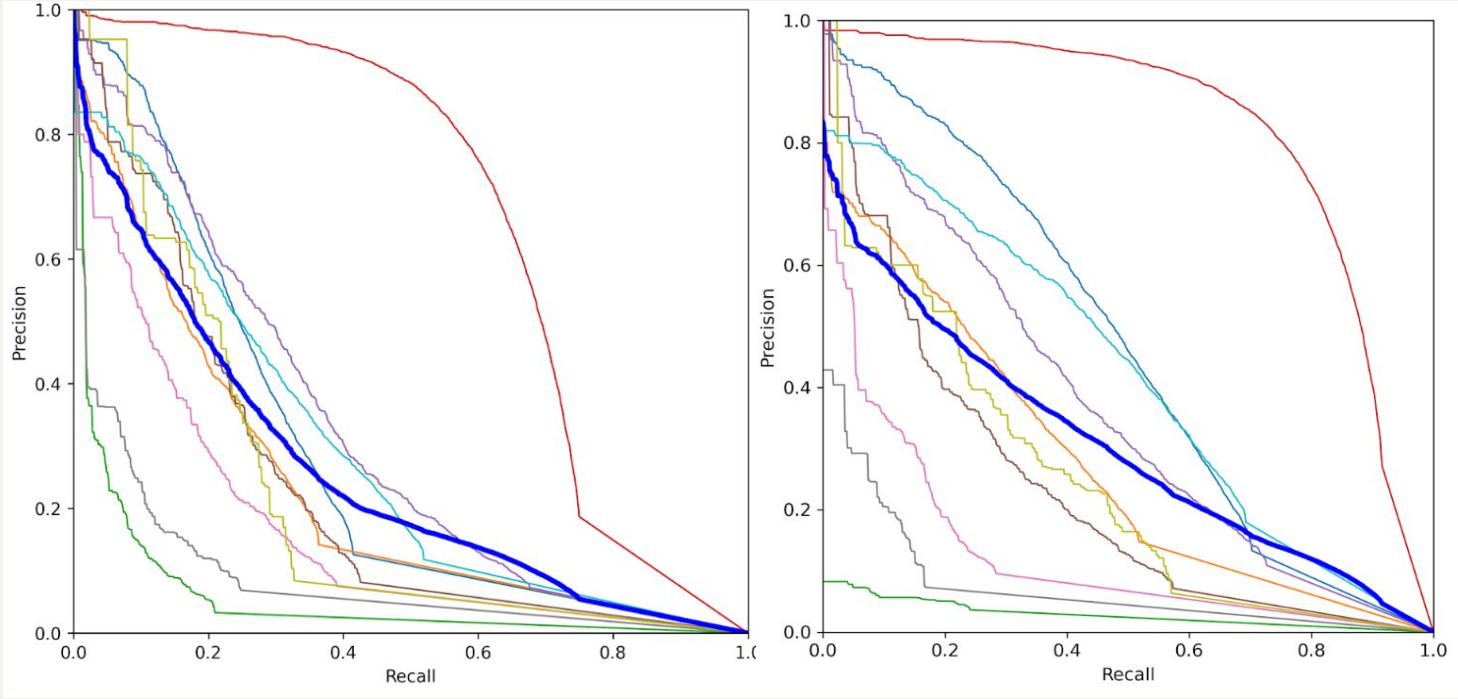

Figure 11: Comparison of Precision vs Recall curve of YOLOv5 and HIC-YOLOv5

Results

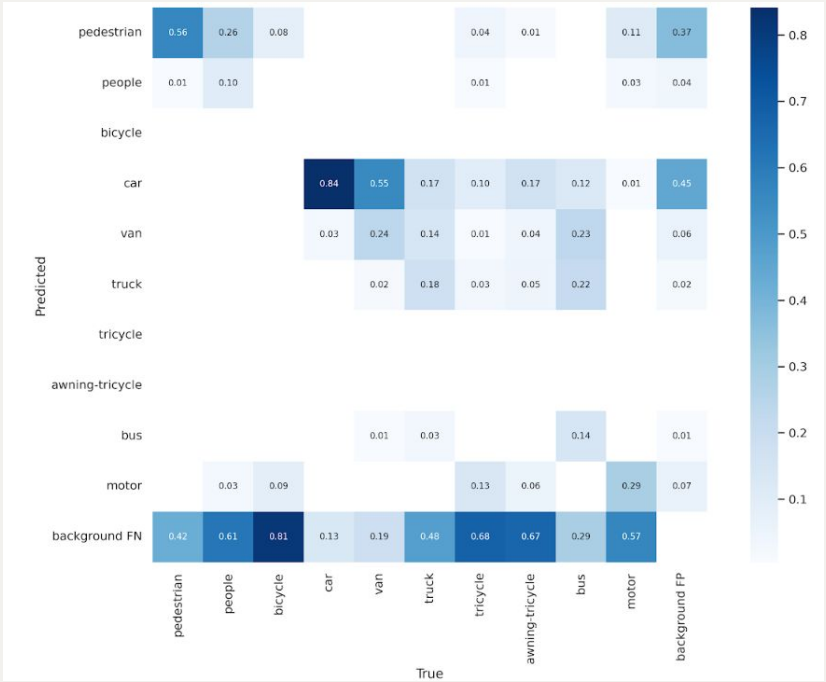

Figure 12: Confusion matrix for HIC-YOLOv5 with the number of epochs during training and validation

Future Work

- Pedestrian Detection.
- Real Time detection.
- More scope for improvement in the architecture
- Better Data Augmentation

References

- Tang, S., Zhang, S., Fang, Y. (2023, September 28). HIC-YOLO V5: Improved YOLO V5 for small object detection. arXiv.org. https://arxiv.org/abs/2309.16393
- Keles¸, M. C., Salmanoglu, B., G¨uzel, M. S., G¨ursoy, B., Bostancı, G. E. (2023). Evaluation of YOLO Models with Sliced Inference for Small Object Detection. arXiv preprint arXiv:2203.04799. Retrieved from https://arxiv.org/pdf/2203.04799
- A. Wang, T. Peng, H. Cao, Y. Xu, X. Wei, and B. Cui, “Tia-yolo v5. An improved yolo v5 network for real-time detection of crop and weed in the field,” Frontiers in Plant Science, vol. 13, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2022.1091655
- Zhu, Pengfei and Wen, Longyin and Du, Dawei and Bian, Xiao and Fan, Heng and Hu, Qinghua and Ling, Haibin(2021). Detection and Tracking Meet Drones Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/TPAMI.2021.3119563.
- Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

THANK YOU!