SUSTech Southern University of Science and Technology

Machine Learning (CS405) Project Final Presentation

Traffic Signs Detection and Recognition

Bryan (12112230), Fitria (12112351), Lam (12111429), Ryan (12111428)

Fall 2023

SUSTech Southern University of Science and Technology

1 Dataset & Model Preparation

2 Methodology

3 Results

4 Conclusion

SUSTech Southern University of Science and Technology

BACKGROUND

Traffic-sign detection and recognition refer to the technology and processes used to identify and interpret various road signs and signals in the context of intelligent transportation systems (ITS) and autonomous vehicles. This vital field plays a crucial role in enhancing road safety, improving traffic management, and enabling the successful deployment of autonomous vehicles. This project aims to delve into the intricacies of Traffic-Sign Detection and Recognition, ultimately developing and deploying a custom deep neural network to address real-world scenarios.

SUSTech Southern University of Science and Technology

OBJECTIVES

- Based on the research, we will meticulously evaluate the various models and select the most suitable one to implement for this project. The chosen model should exhibit the potential for real-world applicability and robust performance.
- Building upon the selected model, we will actively seek opportunities to enhance its performance. This may involve fine-tuning parameters, implementing innovative algorithms, or exploring novel techniques to optimize Traffic-Sign Detection and Recognition.

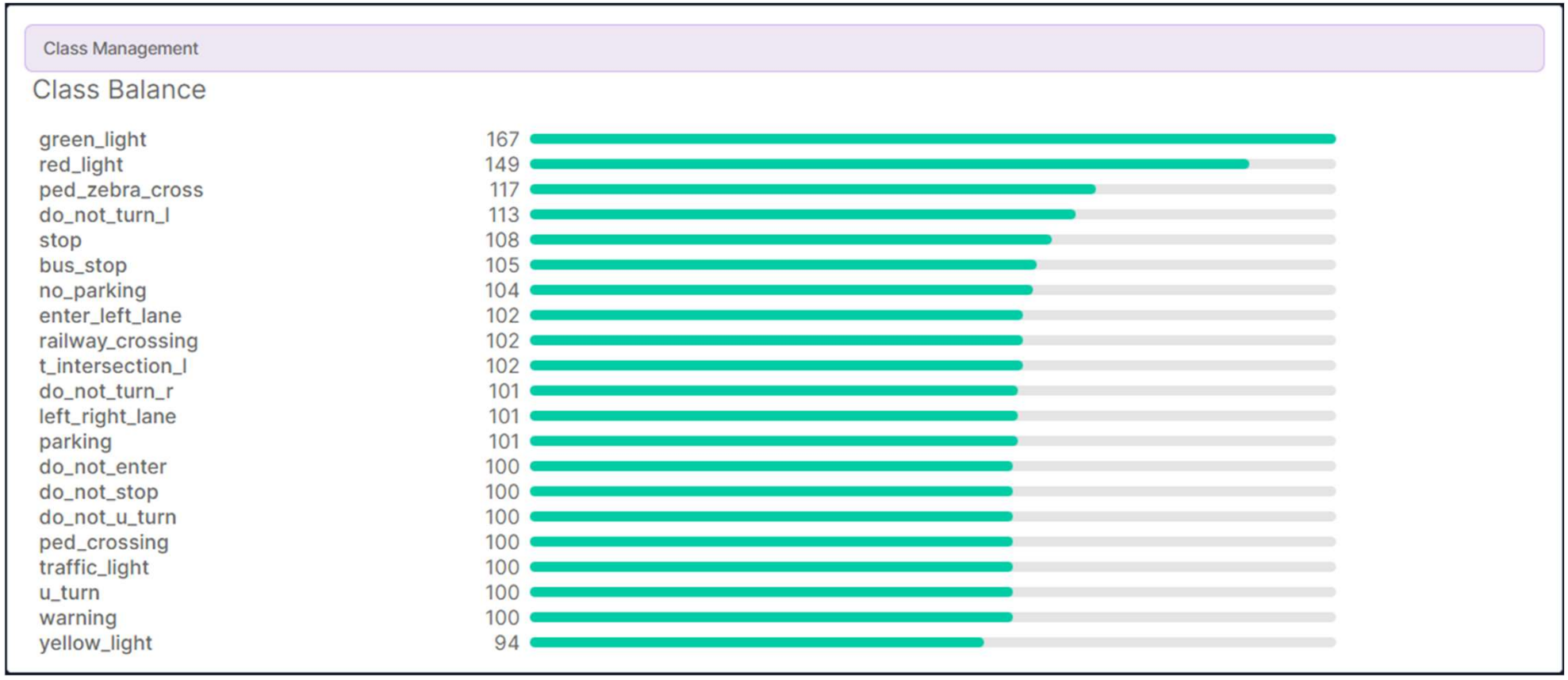
SUSTech Southern University of Science and Technology

01 Datasets Overview

Datasets Overview

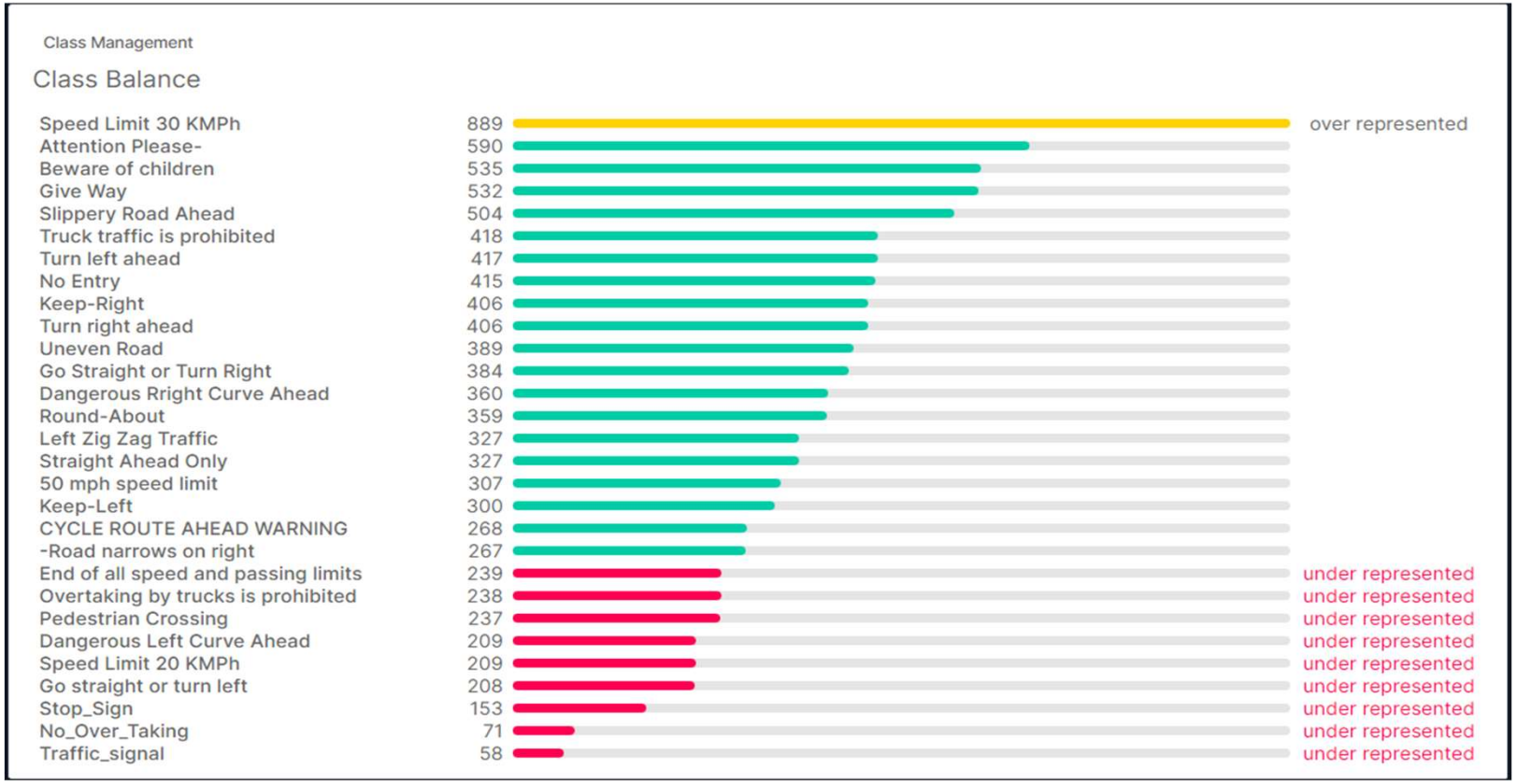
	Dataset 0	Dataset 1	Dataset 2	Custom dataset
Number of samples	2093 images	10000 images	4845 images	3056 images
Train set	66% (1376 images)	71% (7092 images)	85% (4128 images)	69% (2110 images)
Cross-validation set	23% (488 images)	19% (1884 images)	10% (488 images)	31% (946 images)
Test set	11% (229 images)	10% (1024 images)	5% (229 images)	-

Class Distribution – Dataset 0



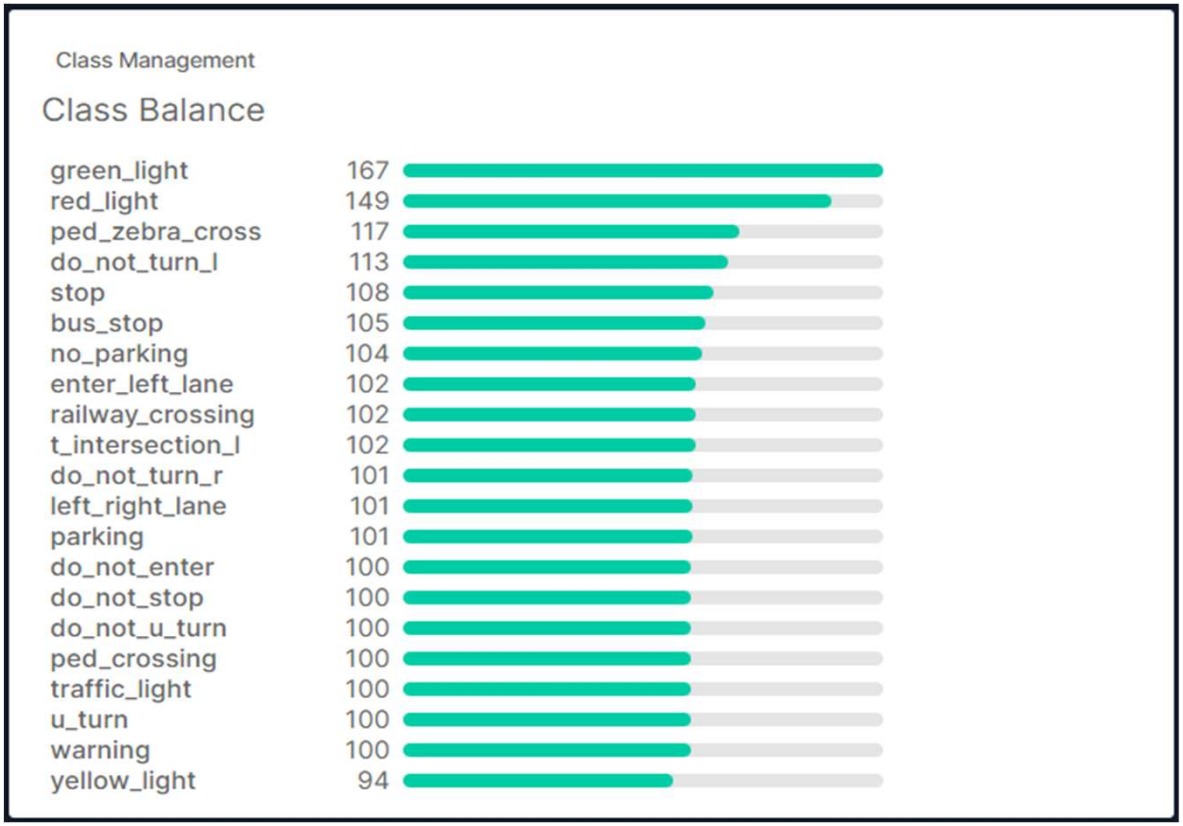
Dataset-0

Class Distribution – Dataset 1



Dataset-1

Class Distribution – Dataset 2

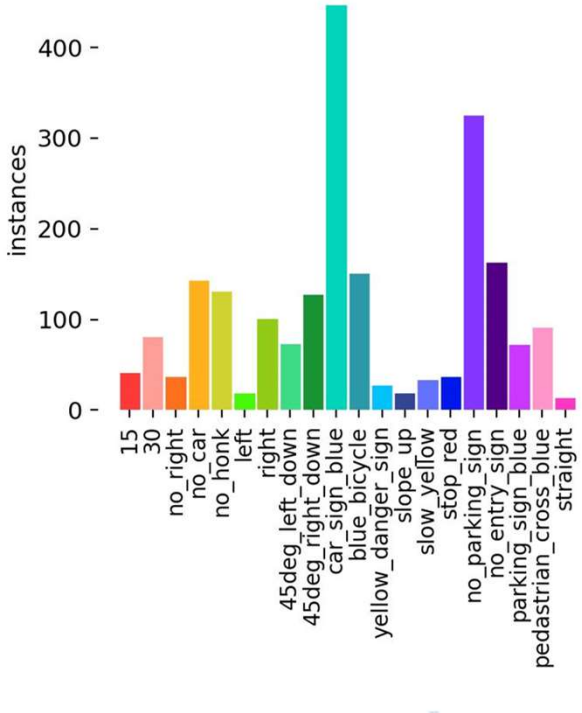


Dataset-2

Custom Dataset

From Different Datasets experiment that yields different results for each signs detection, we experimented in combining different datasets based on sign labels.

- Combined traffic sign datasets from CIFAR, NLPR, and several Roboflow projects
- Delete some not commonly used signs
- Added more traffic signs in a certain class to see the effect of class balance/ imbalance
- Added some data augmentation to certain classes for allowing machine to learn more complex patterns.



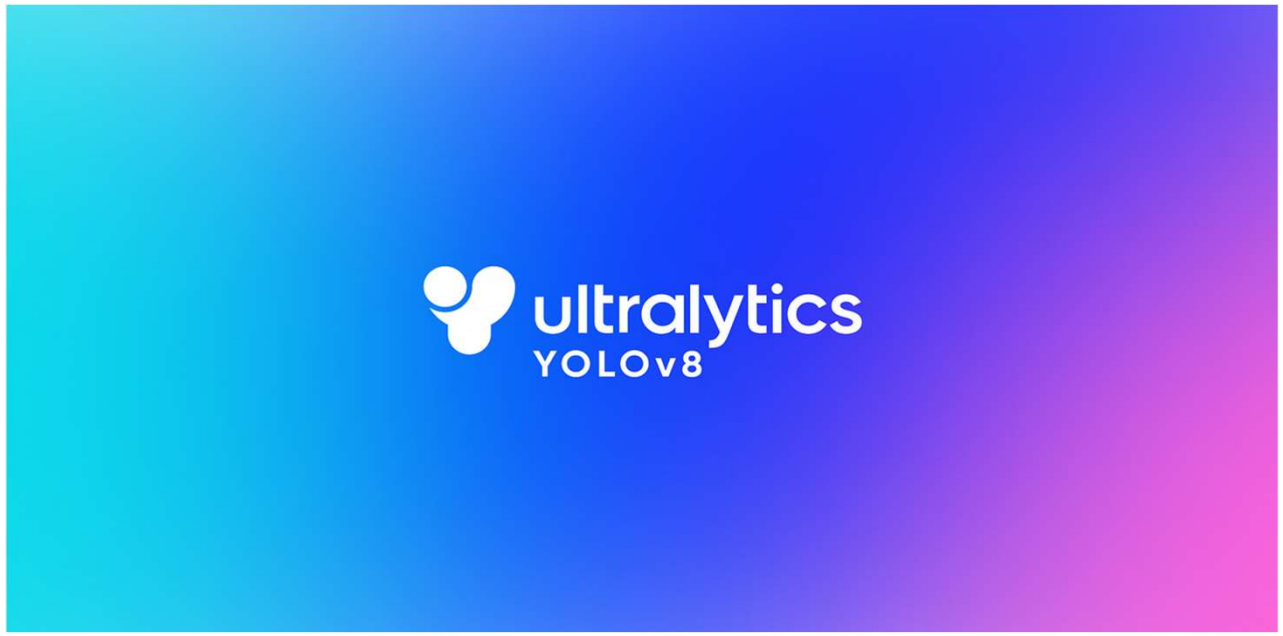
Class distribution

SUSTech Southern University of Science and Technology

02

Experiment

Models



Pretrained Model

- YOLO v8s
- YOLO v8l
- YOLO v8n
- YOLO v8m

SUSTech Southern University of Science and Technology

MODEL TRAINING: 🟦 YOLO v8s

225 layers; 11,143,727 parameters; 11,143,711 gradients; 28.7 GFLOPs

	from	n	params	module	arguments
0	-1	1	928	ultralytics.nn.modules.Conv	[3, 32, 3, 2]
1	-1	1	18560	ultralytics.nn.modules.Conv	[32, 64, 3, 2]
2	-1	1	29056	ultralytics.nn.modules.C2f	[64, 64, 1, True]
3	-1	1	73984	ultralytics.nn.modules.Conv	[64, 128, 3, 2]
4	-1	2	197632	ultralytics.nn.modules.C2f	[128, 128, 2, True]
5	-1	1	295424	ultralytics.nn.modules.Conv	[128, 256, 3, 2]
6	-1	2	788480	ultralytics.nn.modules.C2f	[256, 256, 2, True]
7	-1	1	1180672	ultralytics.nn.modules.Conv	[256, 512, 3, 2]
8	-1	1	1838080	ultralytics.nn.modules.C2f	[512, 512, 1, True]
9	-1	1	656896	ultralytics.nn.modules.SPPF	[512, 512, 5]
10	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
11	[-1, 6]	1	0	ultralytics.nn.modules.Concat	[1]
12	-1	1	591360	ultralytics.nn.modules.C2f	[768, 256, 1]
13	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
14	[-1, 4]	1	0	ultralytics.nn.modules.Concat	[1]
15	-1	1	148224	ultralytics.nn.modules.C2f	[384, 128, 1]
16	-1	1	147712	ultralytics.nn.modules.Conv	[128, 128, 3, 2]
17	[-1, 12]	1	0	ultralytics.nn.modules.Concat	[1]
18	-1	1	493056	ultralytics.nn.modules.C2f	[384, 256, 1]
19	-1	1	590336	ultralytics.nn.modules.Conv	[256, 256, 3, 2]
20	[-1, 9]	1	0	ultralytics.nn.modules.Concat	[1]
21	-1	1	1969152	ultralytics.nn.modules.C2f	[768, 512, 1]
22	[15, 18, 21]	1	2124175	ultralytics.nn.modules.Detect	[21, [128, 256, 512]]

Model summary: 225 layers, 11143727 parameters, 11143711 gradients, 28.7 GFLOPs

SUSTech Southern University of Science and Technology

MODEL TRAINING: YOLOv8m

295 layers; 25,857,478 parameters; 25,857,462 gradients; 79.1 GFLOPs

	from	n	params	module	arguments
0	-1	1	1392	ultralytics.nn.modules.Conv	[3, 48, 3, 2]
1	-1	1	41664	ultralytics.nn.modules.Conv	[48, 96, 3, 2]
2	-1	2	111360	ultralytics.nn.modules.C2f	[96, 96, 2, True]
3	-1	1	166272	ultralytics.nn.modules.Conv	[96, 192, 3, 2]
4	-1	4	813312	ultralytics.nn.modules.C2f	[192, 192, 4, True]
5	-1	1	664320	ultralytics.nn.modules.Conv	[192, 384, 3, 2]
6	-1	4	3248640	ultralytics.nn.modules.C2f	[384, 384, 4, True]
7	-1	1	1991808	ultralytics.nn.modules.Conv	[384, 576, 3, 2]
8	-1	2	3985920	ultralytics.nn.modules.C2f	[576, 576, 2, True]
9	-1	1	831168	ultralytics.nn.modules.SPPF	[576, 576, 5]
10	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
11	[-1, 6]	1	0	ultralytics.nn.modules.Concat	[1]
12	-1	2	1993728	ultralytics.nn.modules.C2f	[960, 384, 2]
13	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
14	[-1, 4]	1	0	ultralytics.nn.modules.Concat	[1]
15	-1	2	517632	ultralytics.nn.modules.C2f	[576, 192, 2]
16	-1	1	332160	ultralytics.nn.modules.Conv	[192, 192, 3, 2]
17	[-1, 12]	1	0	ultralytics.nn.modules.Concat	[1]
18	-1	2	1846272	ultralytics.nn.modules.C2f	[576, 384, 2]
19	-1	1	1327872	ultralytics.nn.modules.Conv	[384, 384, 3, 2]
20	[-1, 9]	1	0	ultralytics.nn.modules.Concat	[1]
21	-1	2	4207104	ultralytics.nn.modules.C2f	[960, 576, 2]
22	[15, 18, 21]	1	3776854	ultralytics.nn.modules.Detect	[2, [192, 384, 576]]

Model summary: 295 layers, 25857478 parameters, 25857462 gradients, 79.1 GFLOPs

SUSTech Southern University of Science and Technology

MODEL TRAINING: 🟩 YOLOv8l

365 layers; 43,631,382 parameters; 43,631,366 gradients; 165.4 GFLOPs

	from	n	params	module	arguments
0	-1	1	1856	ultralytics.nn.modules.Conv	[3, 64, 3, 2]
1	-1	1	73984	ultralytics.nn.modules.Conv	[64, 128, 3, 2]
2	-1	3	279808	ultralytics.nn.modules.C2f	[128, 128, 3, True]
3	-1	1	295424	ultralytics.nn.modules.Conv	[128, 256, 3, 2]
4	-1	6	2101248	ultralytics.nn.modules.C2f	[256, 256, 6, True]
5	-1	1	1180672	ultralytics.nn.modules.Conv	[256, 512, 3, 2]
6	-1	6	8396800	ultralytics.nn.modules.C2f	[512, 512, 6, True]
7	-1	1	2360320	ultralytics.nn.modules.Conv	[512, 512, 3, 2]
8	-1	3	4461568	ultralytics.nn.modules.C2f	[512, 512, 3, True]
9	-1	1	656896	ultralytics.nn.modules.SPPF	[512, 512, 5]
10	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
11	[-1, 6]	1	0	ultralytics.nn.modules.Concat	[1]
12	-1	3	4723712	ultralytics.nn.modules.C2f	[1024, 512, 3]
13	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
14	[-1, 4]	1	0	ultralytics.nn.modules.Concat	[1]
15	-1	3	1247744	ultralytics.nn.modules.C2f	[768, 256, 3]
16	-1	1	590336	ultralytics.nn.modules.Conv	[256, 256, 3, 2]
17	[-1, 12]	1	0	ultralytics.nn.modules.Concat	[1]
18	-1	3	4592640	ultralytics.nn.modules.C2f	[768, 512, 3]
19	-1	1	2360320	ultralytics.nn.modules.Conv	[512, 512, 3, 2]
20	[-1, 9]	1	0	ultralytics.nn.modules.Concat	[1]
21	-1	3	4723712	ultralytics.nn.modules.C2f	[1024, 512, 3]
22	[15, 18, 21]	1	5584342	ultralytics.nn.modules.Detect	[2, [256, 512, 512]]

Model summary: 365 layers, 43631382 parameters, 43631366 gradients, 165.4 GFLOPs

SUSTech Southern University of Science and Technology

Experiment Setup

Experiment	Model	Dataset
Experiment-0	YOLO v8s	Dataset 0
Experiment-1	YOLO v8l	Dataset 0
Experiment-2	YOLO v8l	Dataset 1
Experiment-3	YOLO v8l	Dataset 2
Experiment-4	YOLO v8n	Custom dataset
Experiment-5	YOLO v8s	Custom dataset

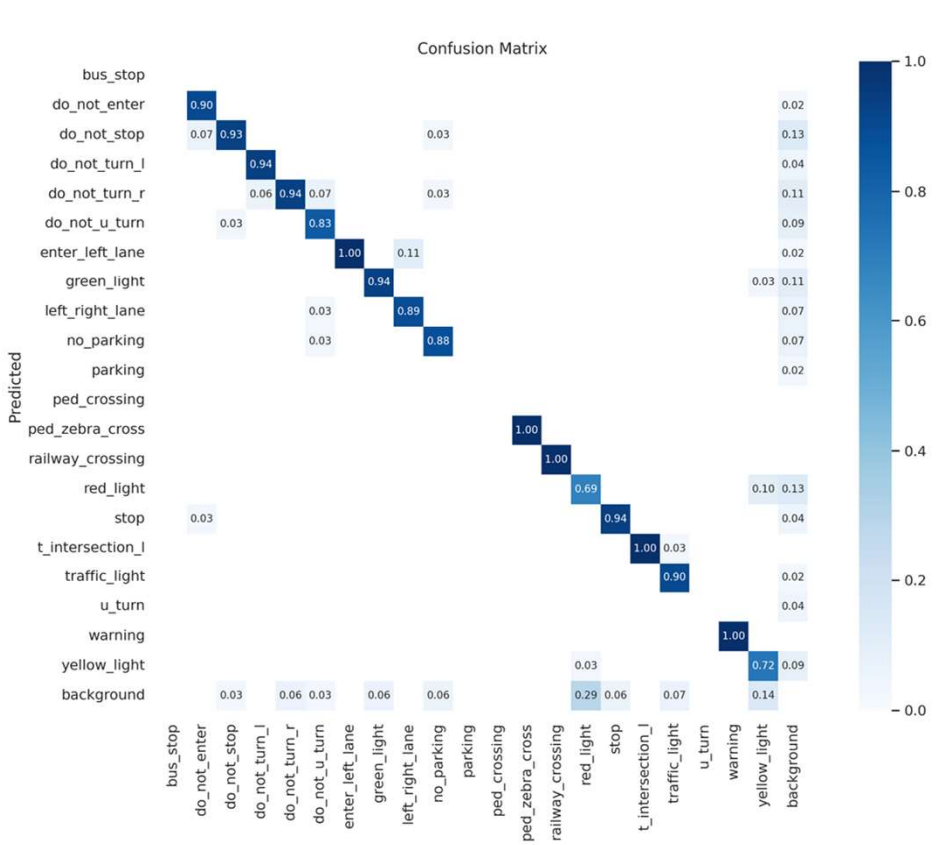
Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

SUSTech Southern University of Science and Technology

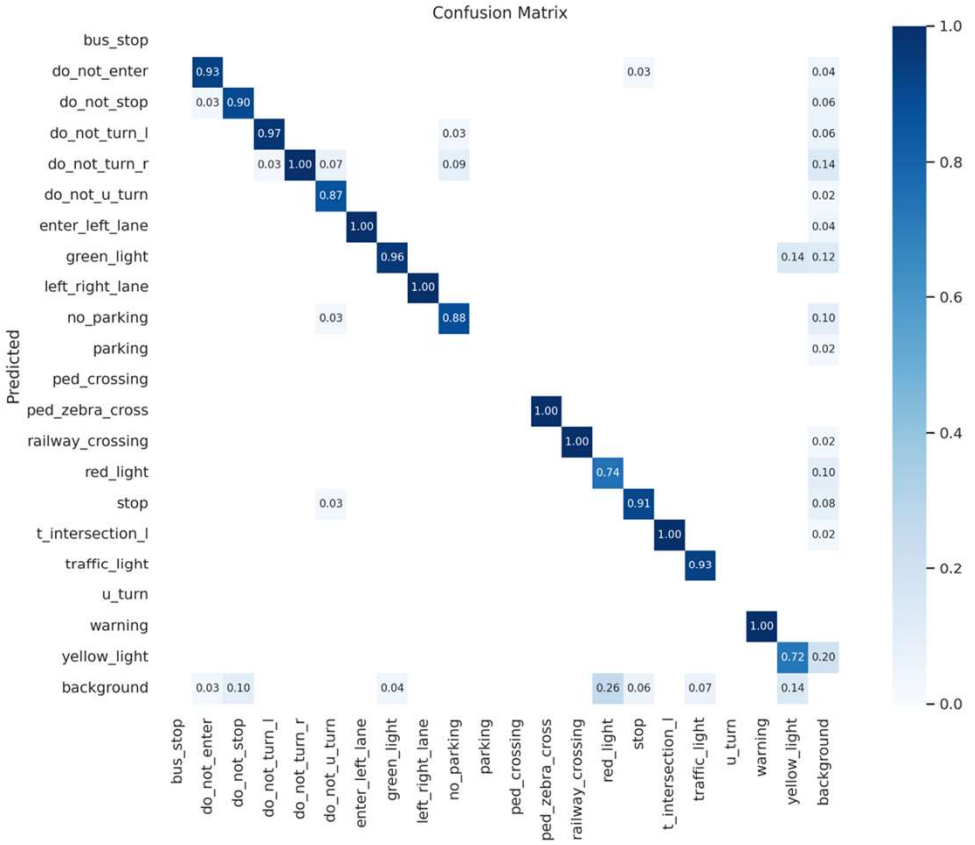
03

Results

Confusion Matrix

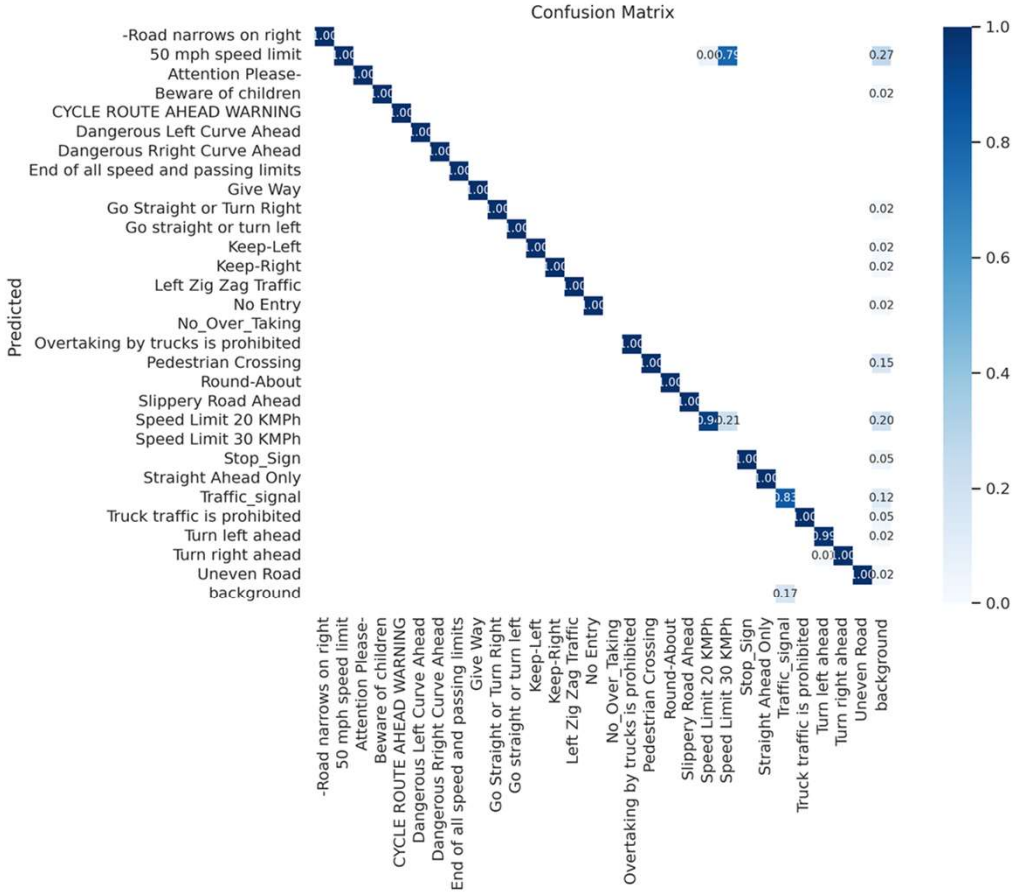


Experiment-0: 🟦 YOLOv8s + Dataset 0

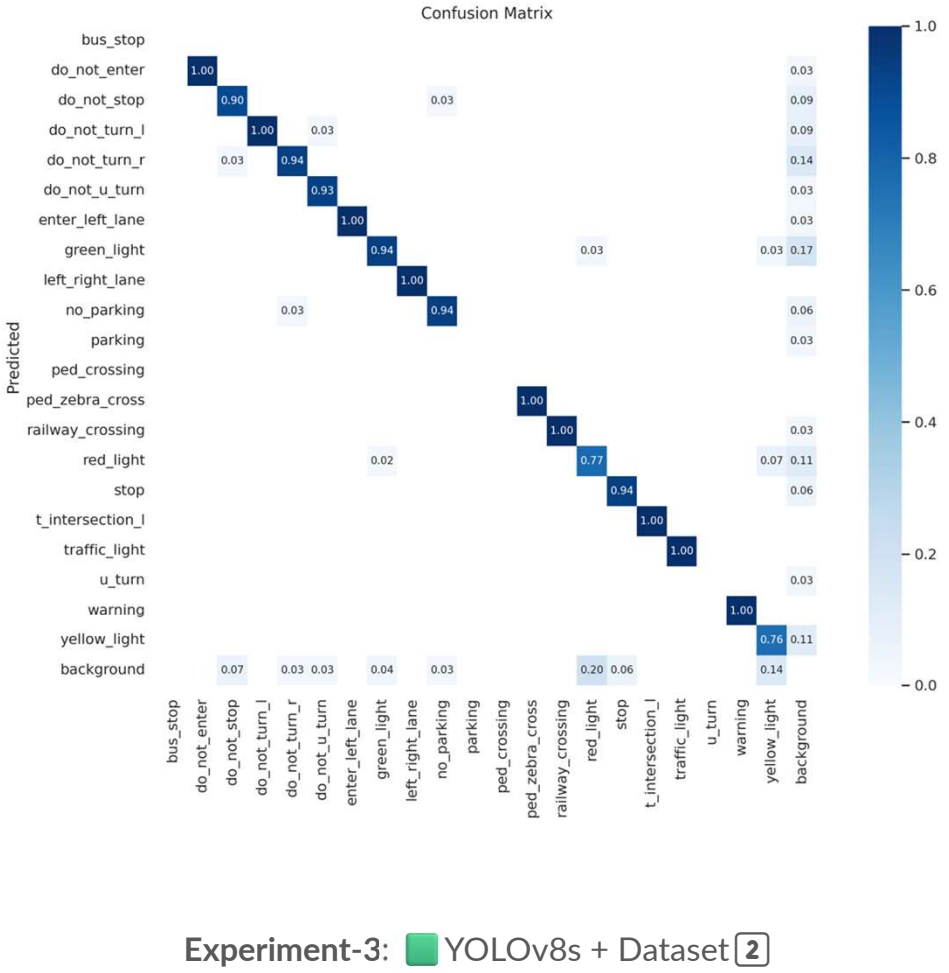


Experiment-1: 🟩 YOLOv8l + Dataset 0

Confusion Matrix

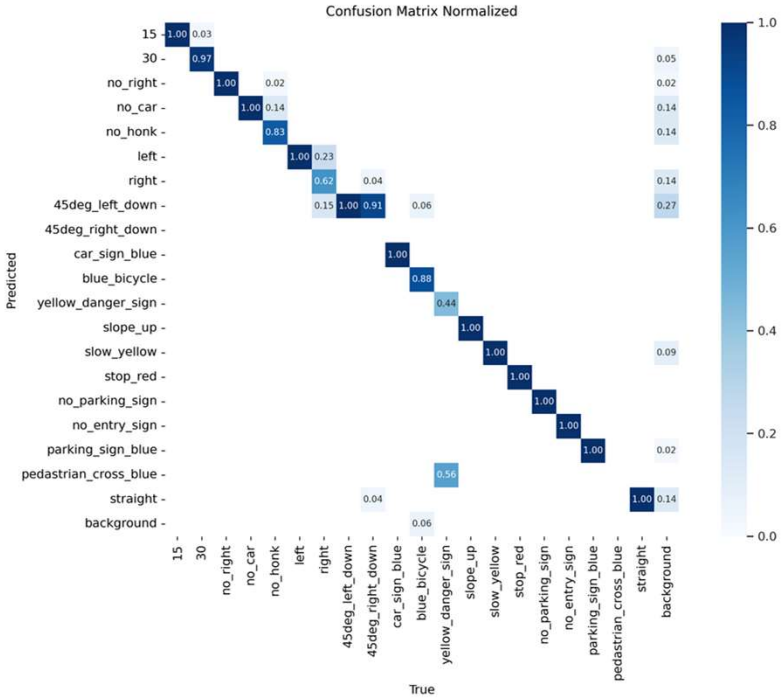


Experiment-2: 🟩 YOLOv8s + Dataset 1

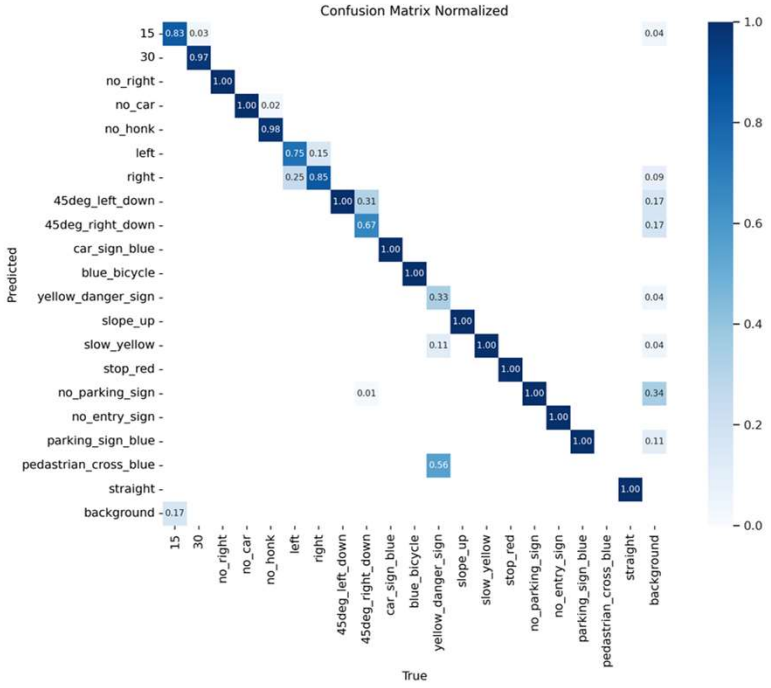


Experiment-3: 🟩 YOLOv8s + Dataset 2

Confusion Matrix

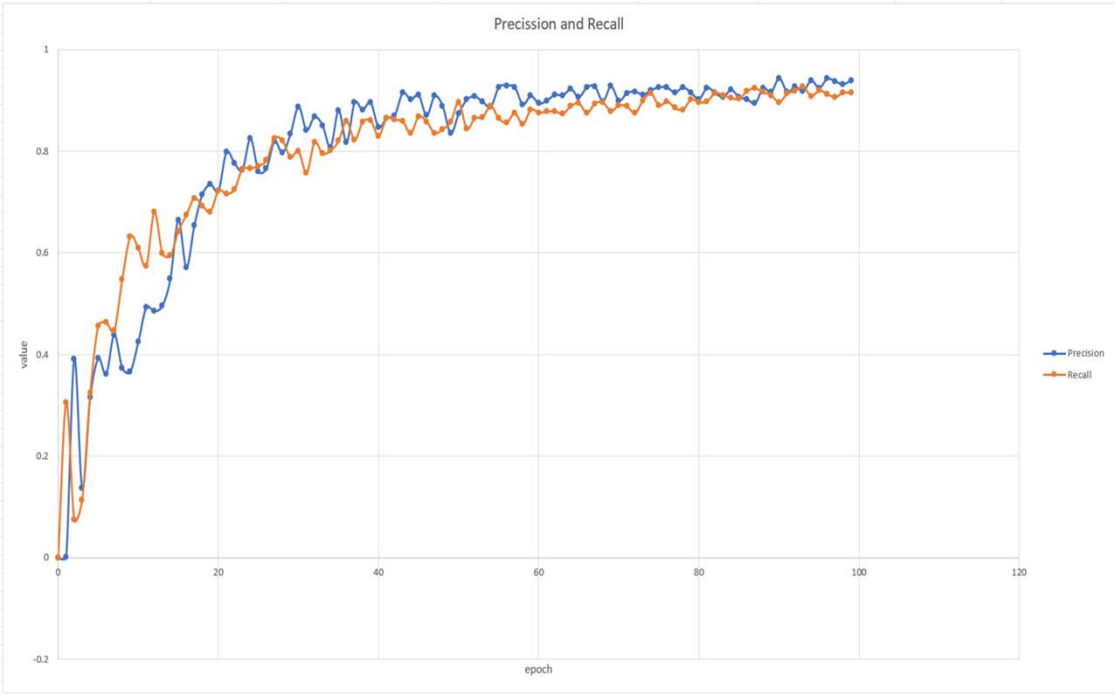


Experiment-4: 🟧 YOLOv8n + Custom Dataset

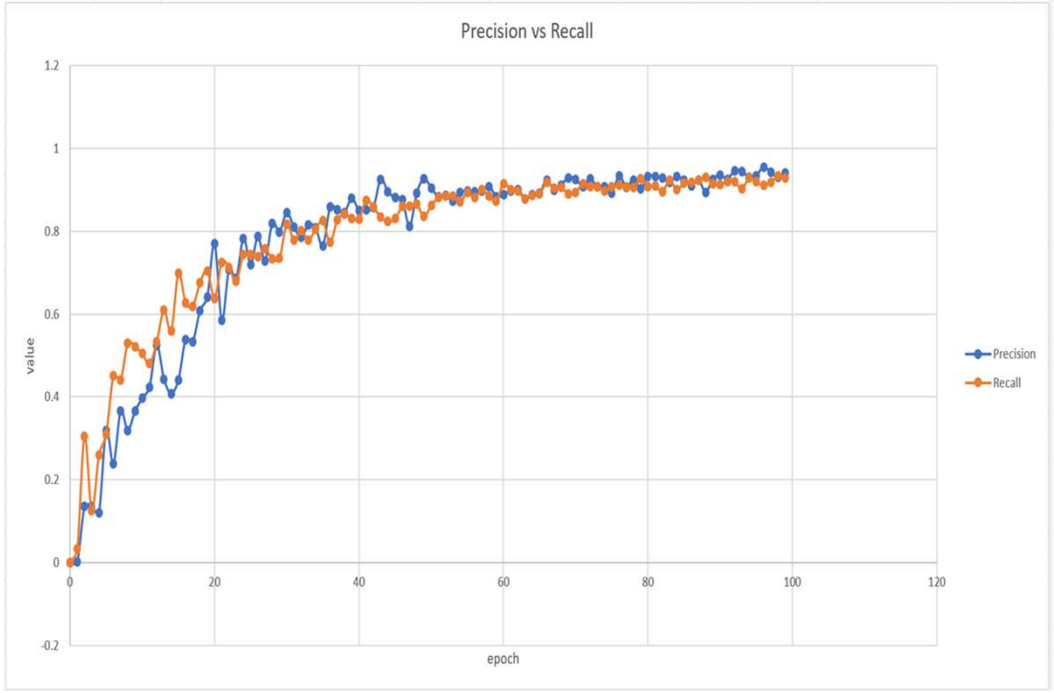


Experiment-5: 🟦 YOLOv8s + Custom Dataset

Precision and Recall

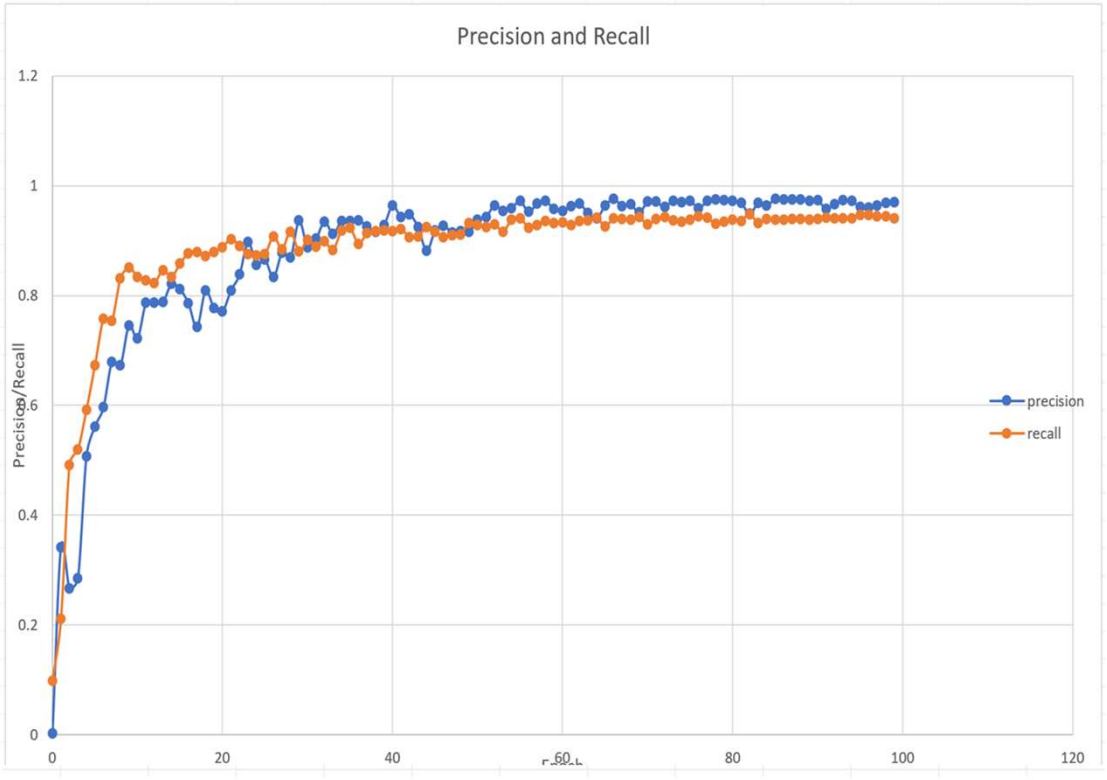


Experiment-0: 🟦 YOLOv8s + Dataset 0

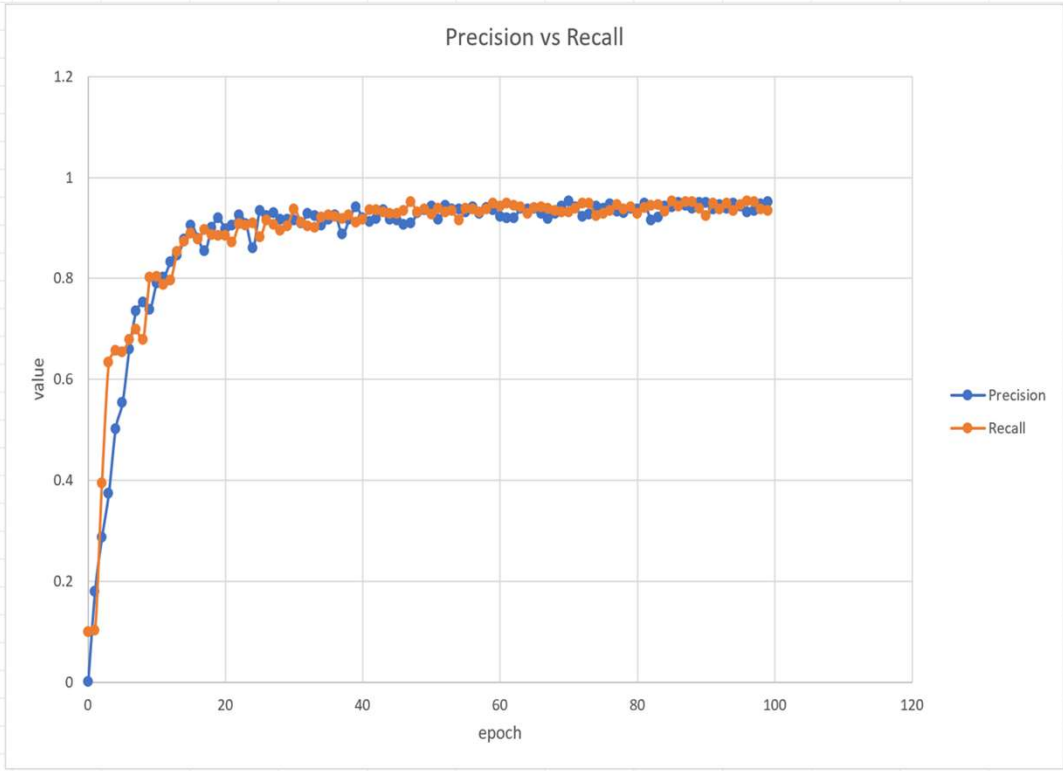


Experiment-1: 🟩 YOLOv8l + Dataset 0

Precision and Recall

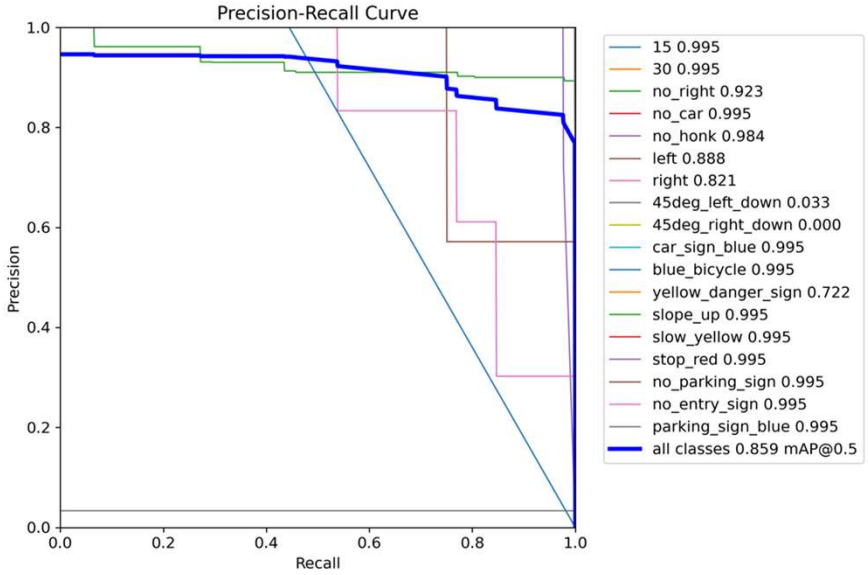


Experiment-2: 🟩 YOLOv8s + Dataset 1

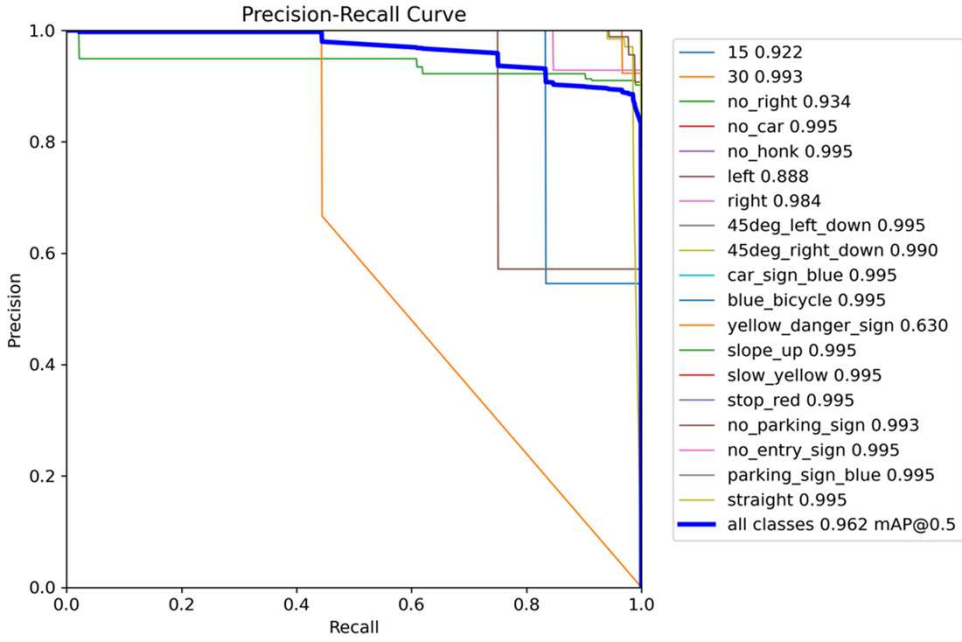


Experiment-3: 🟩 YOLOv8s + Dataset 2

Precision and Recall

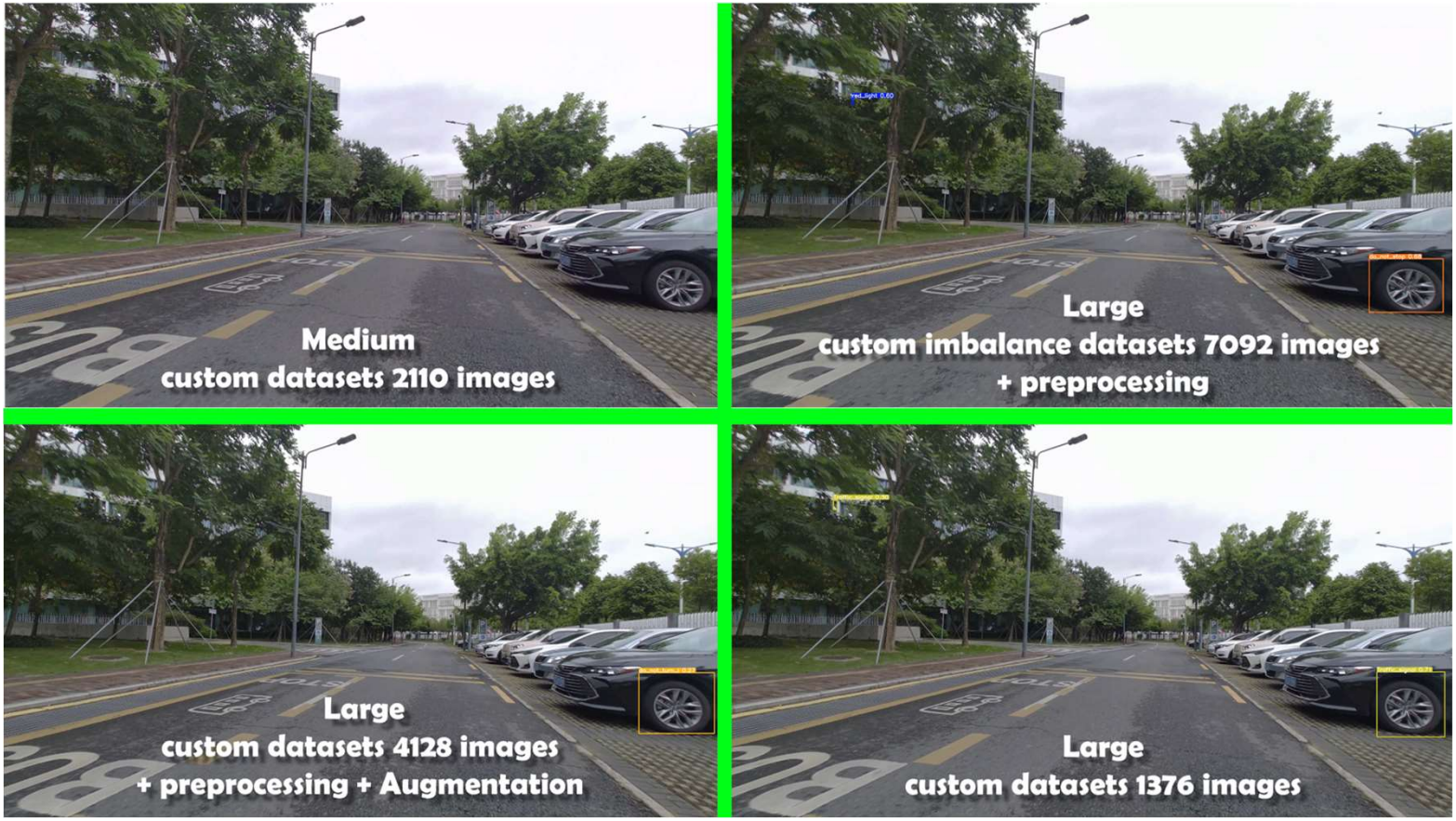


Experiment-4: 🟧 YOLOv8n + Custom Dataset

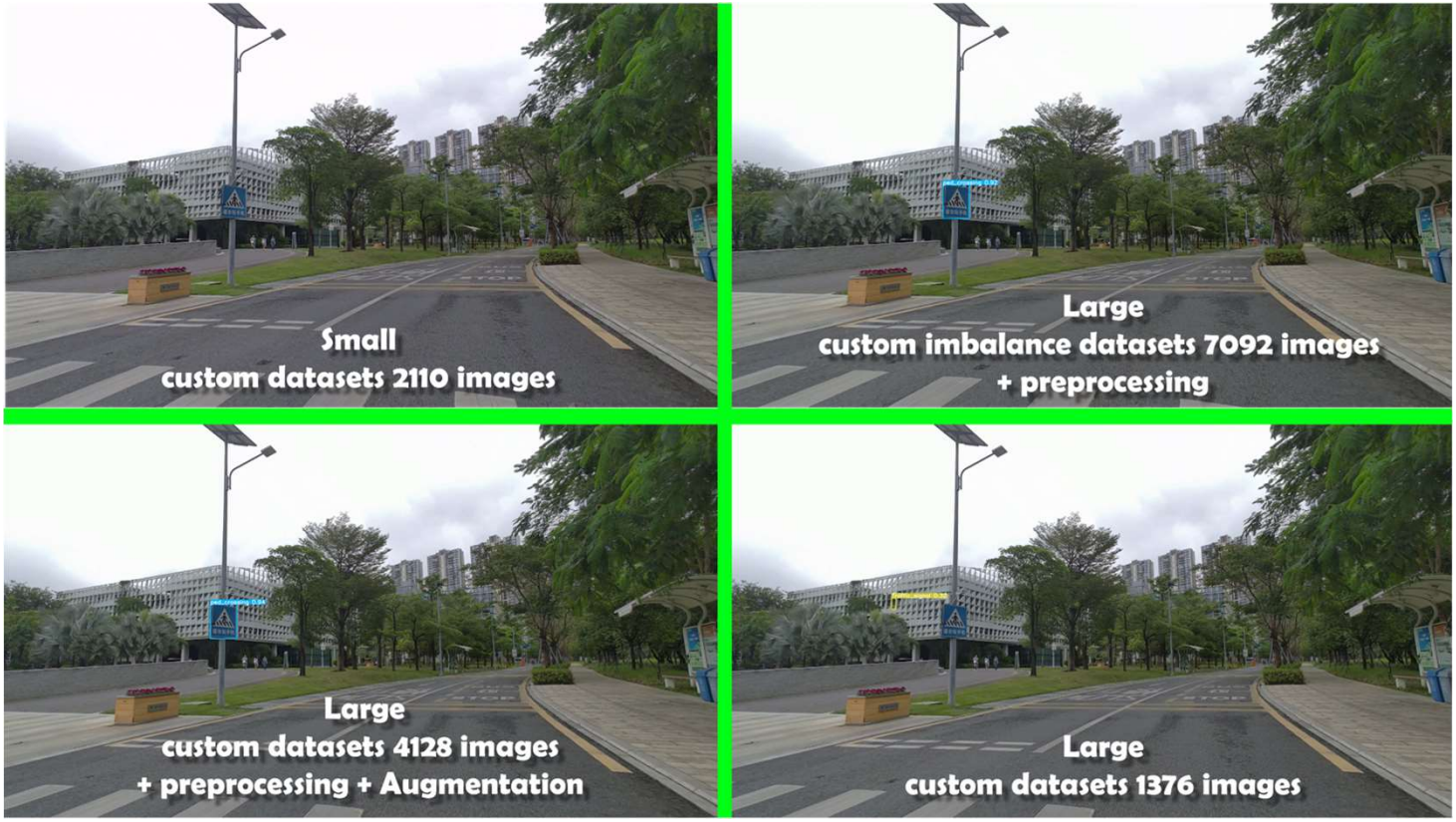


Experiment-5: 🟦 YOLOv8s + Custom Dataset

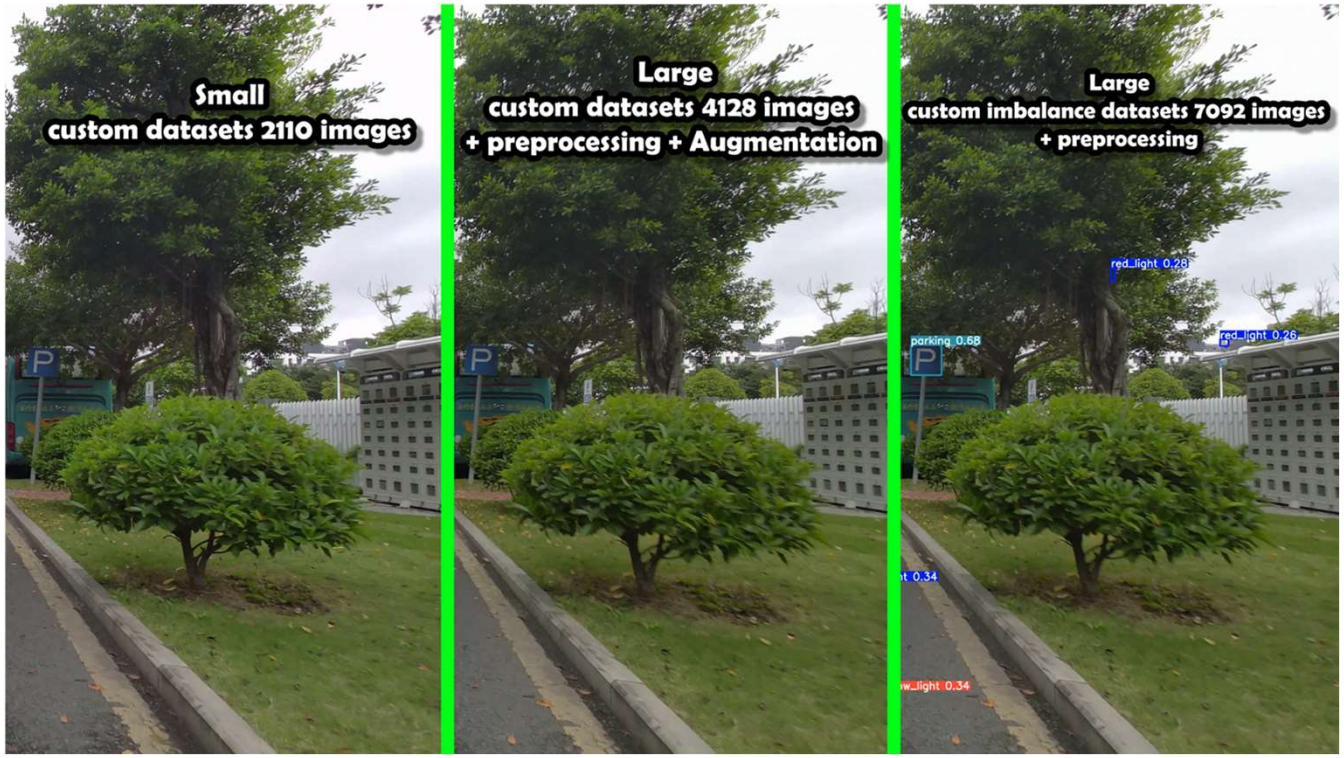
Testing Footage 1



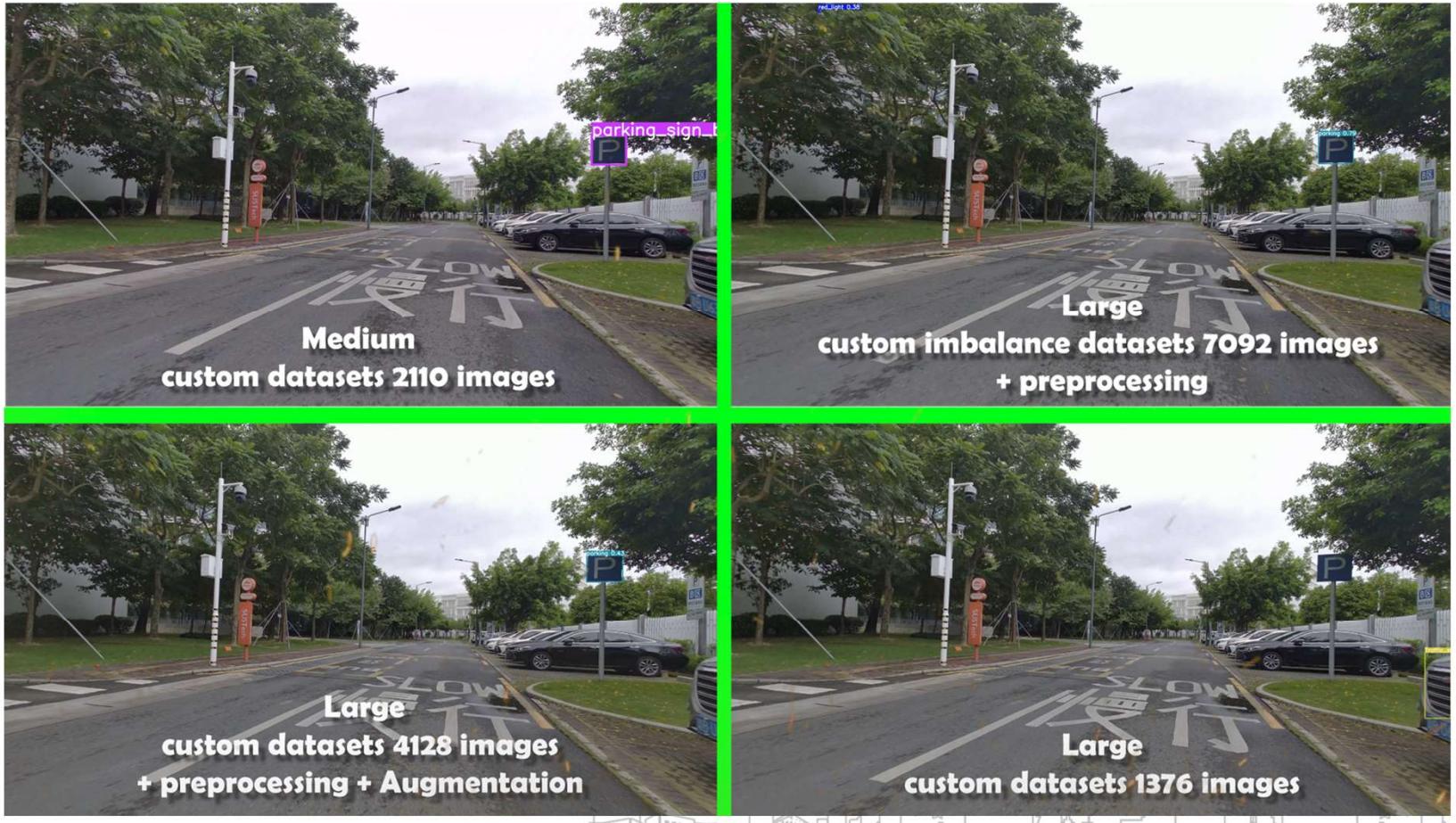
Testing Footage 2



Testing Footage 3

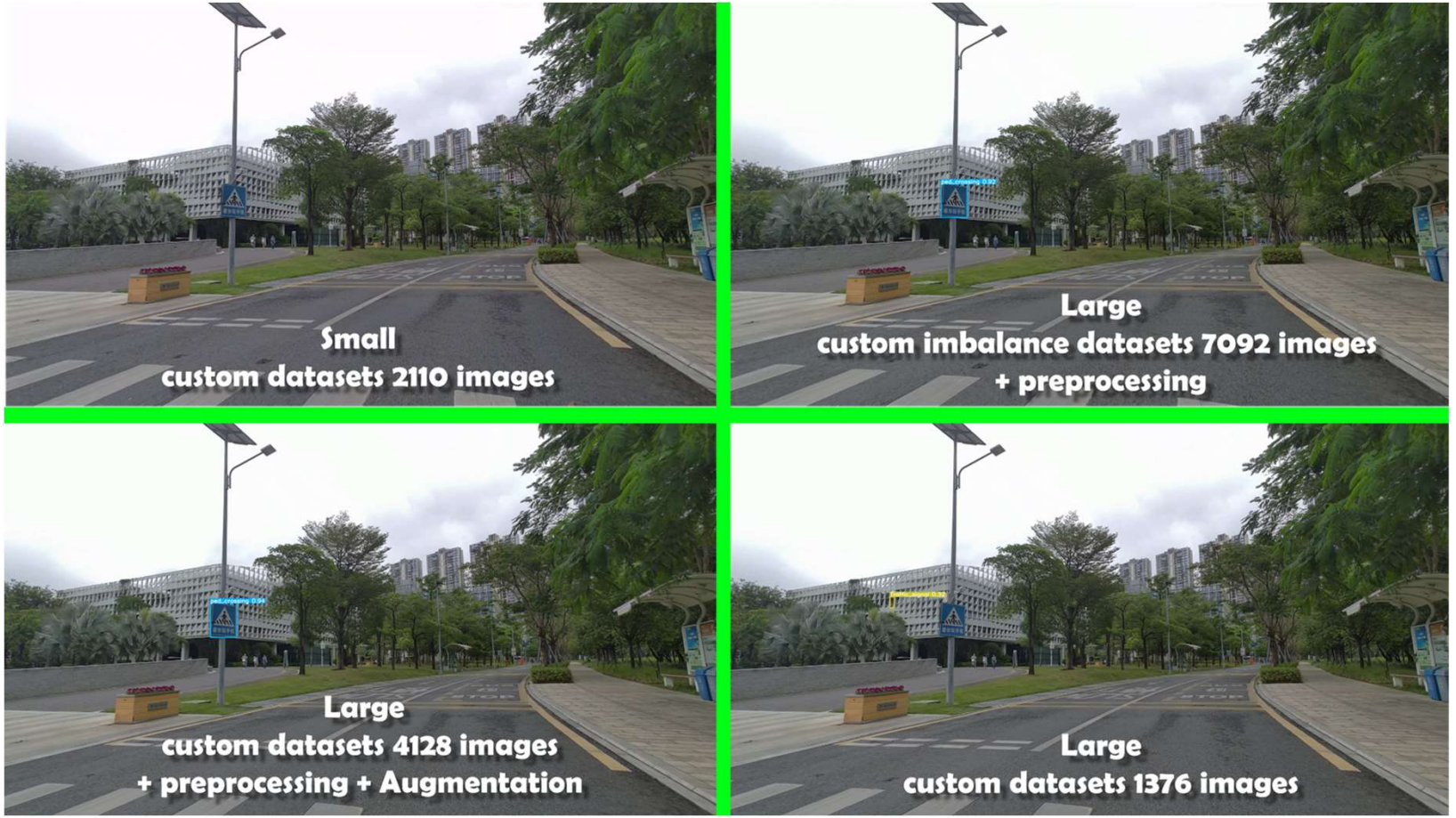


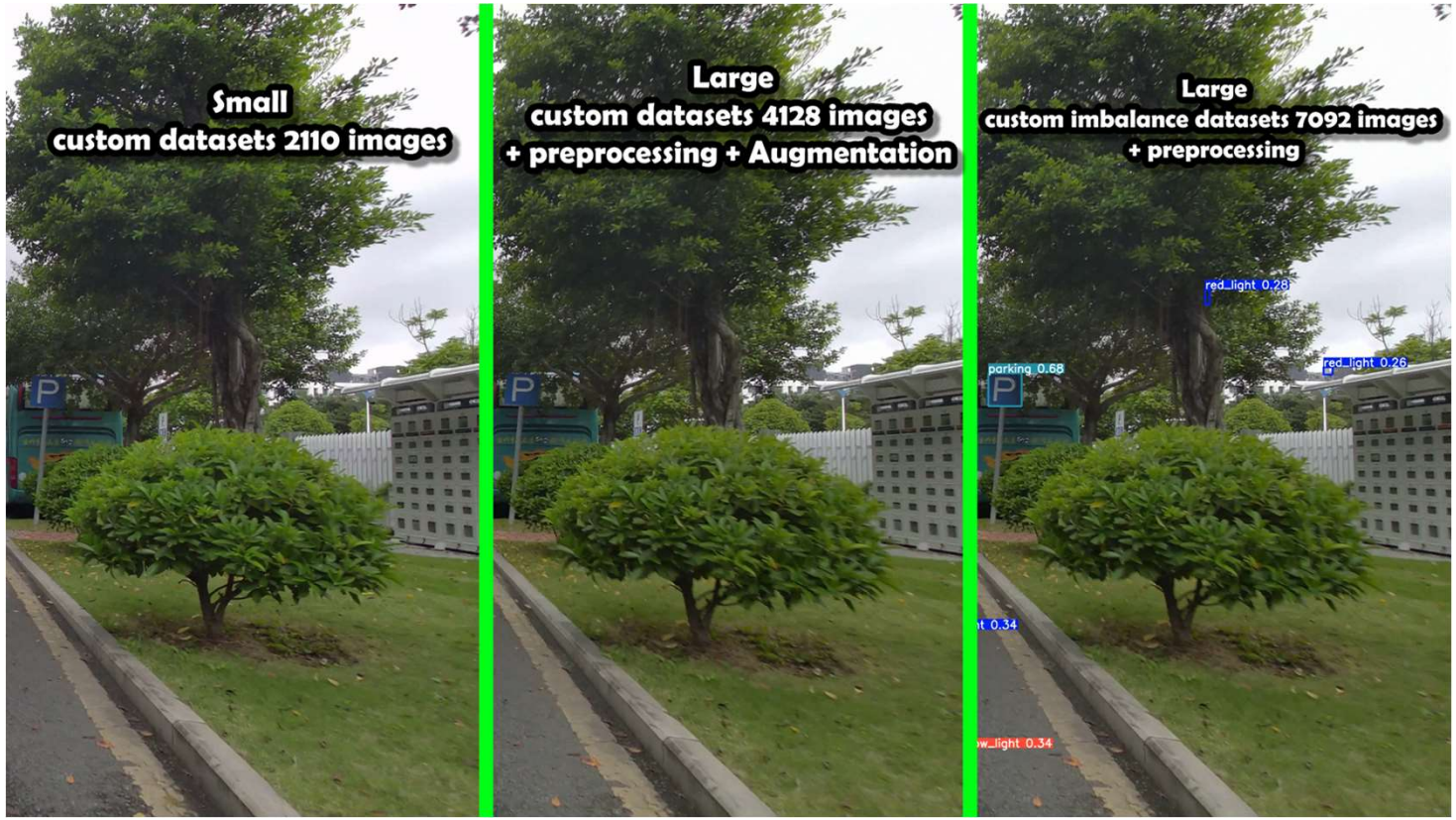
SUSTech Southern University of Science and Technology




Medium
custom datasets 1001 images
Large
custom imbalance datasets 7092 images
+ preprocessing
Large
custom datasets 4128 images
+ preprocessing + Augmentation
Large
custom datasets 1376 images

Further Analysis




Small
custom datasets 2110 images
Large
custom datasets 4128 images
+ preprocessing + Augmentation
Large
custom imbalance datasets 7092 images
+ preprocessing
red_light 0.28
parking 0.68
red_light 0.26

SUSTech Southern University of Science and Technology

04

Conclusion

🟦 **YOLOv8m**

- Trained on Custom Dataset of 1001 images.
- Outperformed smaller models.
- Demonstrated strong accuracy in identifying traffic signs in videos.
- Highlights the effectiveness of larger models for video and image understanding.

🟩 **YOLO8l**

- Initially trained on a custom dataset of 7092 images with preprocessing.
- Exhibited false positives and negatives in video testing, likely due to dataset imbalance favoring majority classes.
- Training on a smaller custom dataset of 4128 images with additional preprocessing and augmentation improved performance.
- Despite improvements, some traffic signs were still misclassified.

🔍 Key Observations

- Larger models like YOLOv8l can struggle with traffic sign detection in videos without adequate preprocessing and augmentation.
- Dataset quality and balance significantly impact model performance.
- Augmentation and preprocessing enhance performance but don’t fully address misclassification issues.

📝 Conclusion

- YOLOv8m performs well overall.
- YOLOv8l shows promise with augmentation and preprocessing but requires further refinement.
- ddressing overfitting and exploring the impact of dataset characteristics can improve model reliability and accuracy.

Task Division

Member	Task
FARIDA FITRIA ZUSNI	Data collection, Documentation
TANG RYAN TZE HOU	Data annotations, Training the Model
NGUYEN THANH LAM	Training the model, evaluating models performance
ANTHONY BRYAN	Data collection and data annotation

SUSTech Southern University of Science and Technology

References

1) Ultralytics. (n.d.). *Ultralytics GitHub repository*. Retrieved from https://github.com/ultralytics/ultralytics?ref=blog.roboflow.com
2) Roboflow. (2023). *What's new in YOLOv8?* Retrieved from https://blog.roboflow.com/whats-new-in-yolov8/#:~:text=YOLOv8%20was%20launched%20on%20January%2010th%2C%202023
3) Ultralytics. (n.d.). *Ultralytics documentation*. Retrieved from https://docs.ultralytics.com/
4) Roboflow. (n.d.). *What is an anchor box?* Retrieved from https://blog.roboflow.com/what-is-an-anchor-box/
5) Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv*. Retrieved from https://arxiv.org/abs/2007.14917

SUSTech Southern University of Science and Technology

THANKS for WATCHING

Fall 2023