A Comparative Analysis of Deep Learning Based YOLO Models in Object Detection for Autonomous Vehicles and the Development of a Unified Model

Sean Kill, **MSCS Candidate, CoAS**; **Travis Bowman**, **BSRE+BSCS Candidate, CoE**; **Aaron Wisneski**, **BSCS Candidate, CoAS**;
Devson Butani, **MSCS Candidate, CoAS**; **Ryan Kaddis**, **BSCS Candidate, CoAS**; **CJ Chung**, **PhD, Professor, CoAS**

College of Arts & Science, Lawrence Technological University

Be curious. Make magic.

ABSTRACT

The Intelligent Ground Vehicle Competition (IGVC), is a yearly competition that give students a multi disciplinary challenge in developing self driving cars. Four recent IGVC problems include: stop sign and fake stop sign detection, simulated pothole detection, object in road detection, and pedestrian detection. All of these problems are solvable with traditional computer vision methods to perceive an element and react to it accordingly. This research revisits these past problems and solves them with deep learning models. YOLO, short for You Only Look Once, is a deep learning model that can inference and predict the location of different objects from photos in real time. New advancements like the release of YOLOv9 have improved deep learning in computer vision tasks. Stop sign detection, and tire detection were previously accomplished with YOLOv8-m models. The past data sets were used for the newly created models. Two new data sets were created for simulated pothole detection and pedestrian detection. Three YOLO models were created for each of the problems using versions 7, 8s, 8m, and 9c. After these models were merged into one unified model that was trained on versions 8s, 8m and 9 to identify all of the classes. This provides the IGVC team with a single model that should reduce the stress on the GPU and increase performance. The performance of these models is compared over precision, recall, confusion matrices, and F1 scores, with the unified model performing the best compared to each individual model. The speed of running the models was also used in evaluating performances. This research provides the IGVC team with a comprehensive list of different YOLO models and their relative performances. The results led to the conclusion that the unified model was the optimal model for all of the object detection. The best unified model uses version 8m with the merging of select classes.

DATA SETS

Data Set List
Originally, 4 datasets were used to create a model for each IGVC problem. The stop sign dataset included roughly 4000 images of the different stop sign classes. This data was collected by past IGVC members, and reused for this project
The tire data set contained roughly 2000 images, and was also taken from previous year's IGVC work.
The simulated pothole dataset was created for this project. The data was collected with a Gopro camera that was held at a height as close as possible to the camera on the vehicle. It contains roughly 2000 images.
The pedestrian dataset contained roughly 2000 images. It was also collected specifically for this project with a Gopro camera.

Data Unification
After these datasets were all created, they were combined into one dataset with all 8 classes. This dataset contains approximately 9600 images. With data augmentation this jumps up to roughly 15000 images.

Class Unification
One final dataset was created that merged some of the classes to improve performance. This dataset has roughly the same number of images, but merges the stop sign, vandalized sign, and obstructed stop sign into one class. It also removes the fake stop sign class and treats it like a background image.

Count of Images per Class

Class 1:	Tire, 1831
Class 2:	Pothole, 1253
Class 3:	Pedestrian with vest, 1302
Class 4:	Pedestrian without vest, 932
Class 5:	Stop sign, 2709
Class 6:	Stop sign obstructed, 1034
Class 7:	Stop sign vandalized, 758
Class 8:	Fake stop sign, 338
Background images:	Null**,** 416

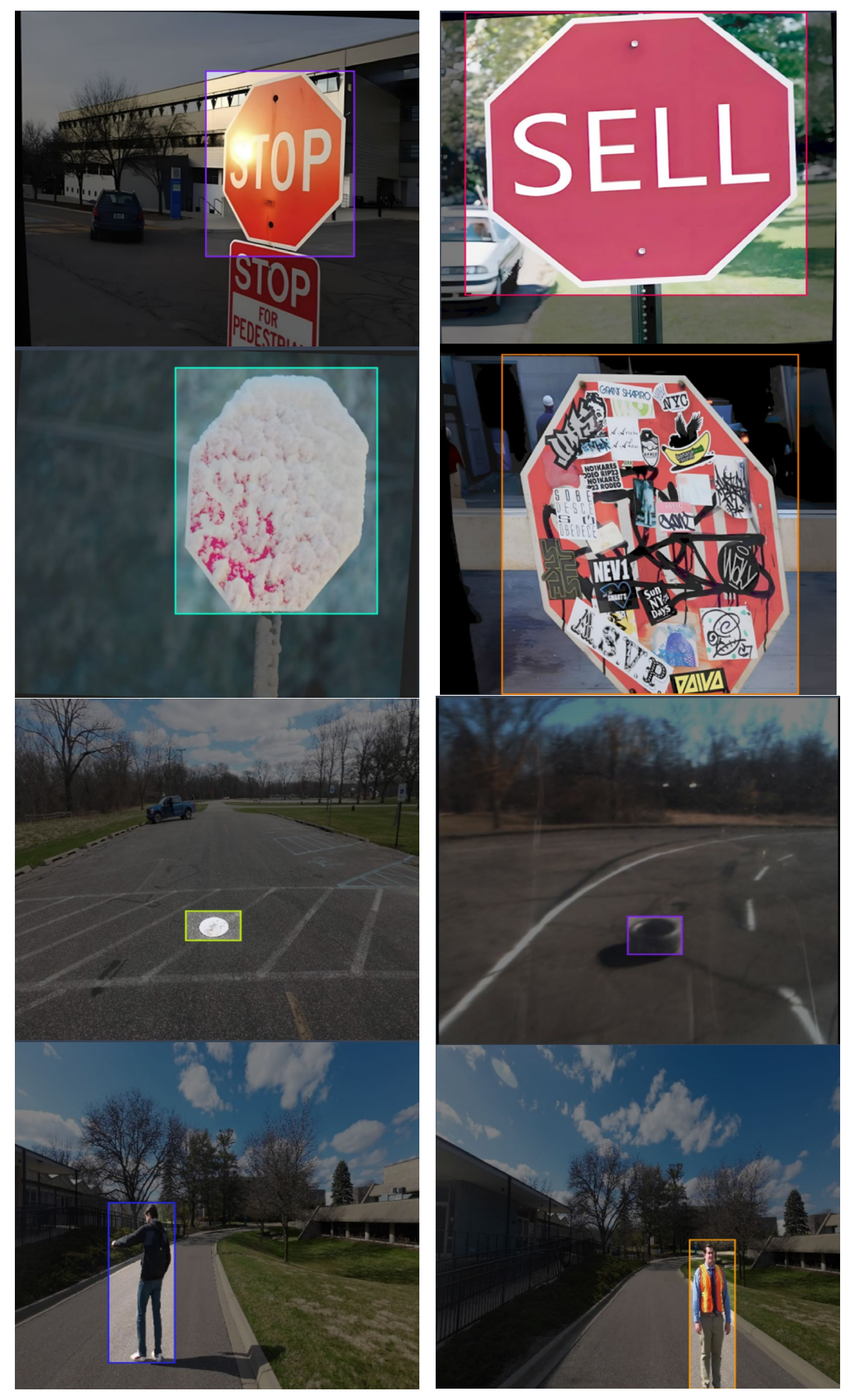

Figure 1: Example of Classes in Datasets

MODEL EXPERIMENTS

Data Processing:
The data processing included a number of steps for each of the models. All of our images were resized to 640x640 images for each model.

Training Metrics
The models were trained from Google Colab PRO accounts with the T4 GPU, and a GPU server with a Nvidia A100 GPU with 80GB VRAM. Last years models were trained off of RTX 3080 with only 10GB of VRAM.

Separate Models
Separate models were created for each of the four problems. Each separate model was trained with YOLO version 7, 8m, and 9c. The precision recall curve, F1 score, and confusion matrix were used to evaluate these models.
For the stop sign models version 8 performed the best with a map score of 0.904. Version 9 was close behind at 0.895.
The tire model's best precision accuracy score for all classes was 0.986 from version 8. Version 9 was narrowly in second place at 0.968.
The pedestrian models performed very close to each other. Version 8 scored 0.988, and version 9 scored 0.987.
The pothole model's best score was 0.975 from version 8. Version 9 was the second best at 0.968.
These models helped evaluate the proficiency of YOLO versions, and the data sets. Version 8 performed the best in each case, but version 9 was close behind. Version 7 was much further behind in each case.

Unified Models
After these models were developed, a unified model was created in versions 8 and 9. The results of the unified model were much better than the individual versions.

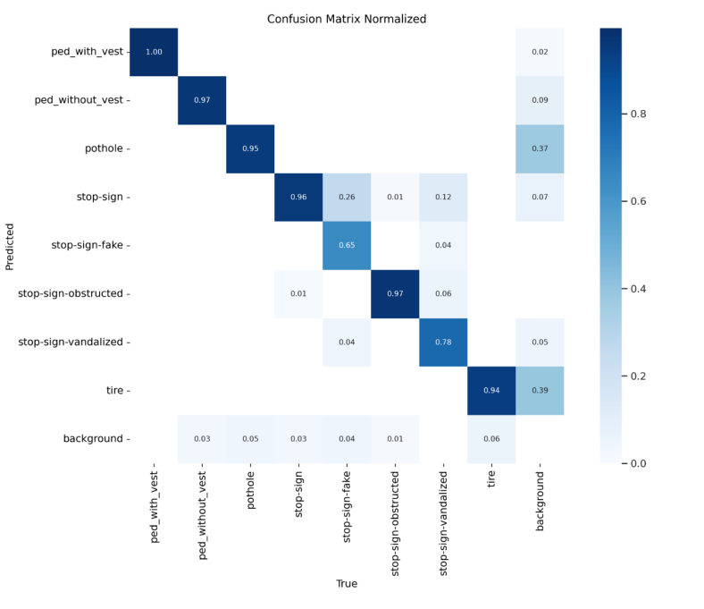

Figure 2: Unified Confusion Matrix 8m

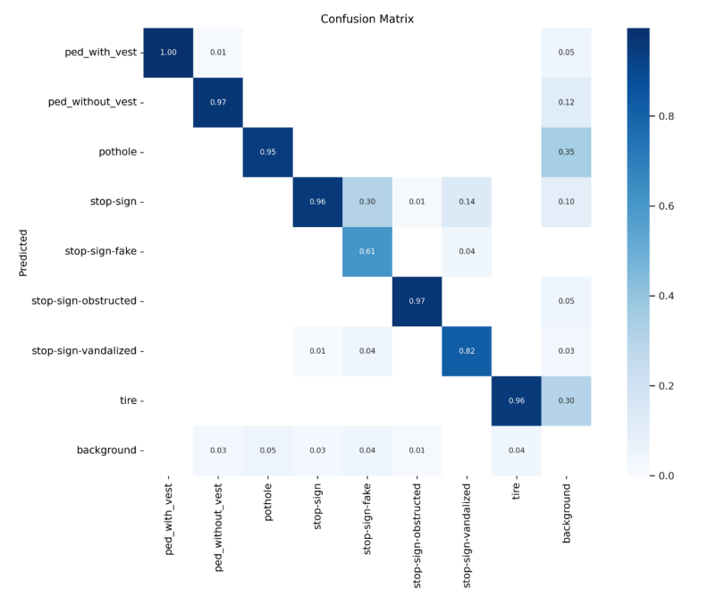

Figure 3: Unified Confusion Matrix 9c

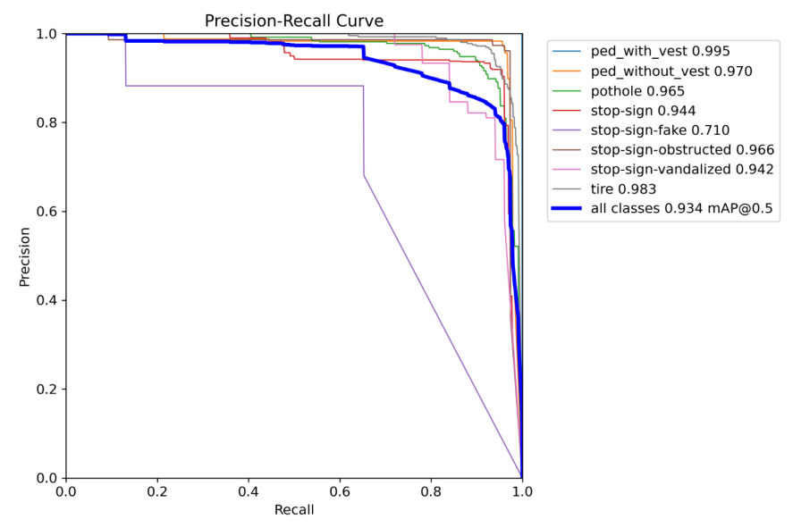

Figure 4: Unified Precision Recall 8m

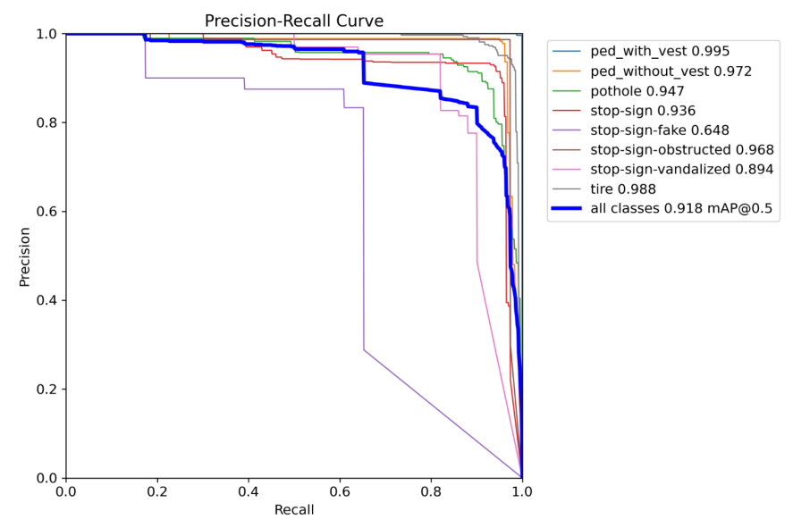

Figure 5: Unified Precision Recall 9c

Version 8 performs slightly better than version 9. The fake tire score also brings down the score by a significant amount. In an attempt to increase the models ability to determine real from fake stop signs, the merged class unified model was created.

YOLOv8 vs YOLOv9

Model Comparison

YOLOv8 and YOLOv9 utilize different architectures; version 9 implements hallmarks in reducing data loss. Both models performed at a very similar level despite these differences. As of right now, version 8 performs significantly better in terms of inference and prediction because version 9 has not been implemented into the ultralytics package, so inference is slow and the results are not as robust. The time to inference was tested across the separate and unified models. In every case version 8 performed significantly better. Because of this YOLO version 8 has been chosen as the model to move forward with for implementation. The future development of version 9 may make it viable as an option in the future, but it is not at that stage yet.

YOLOv8 and YOLOv9 Results Comparison:

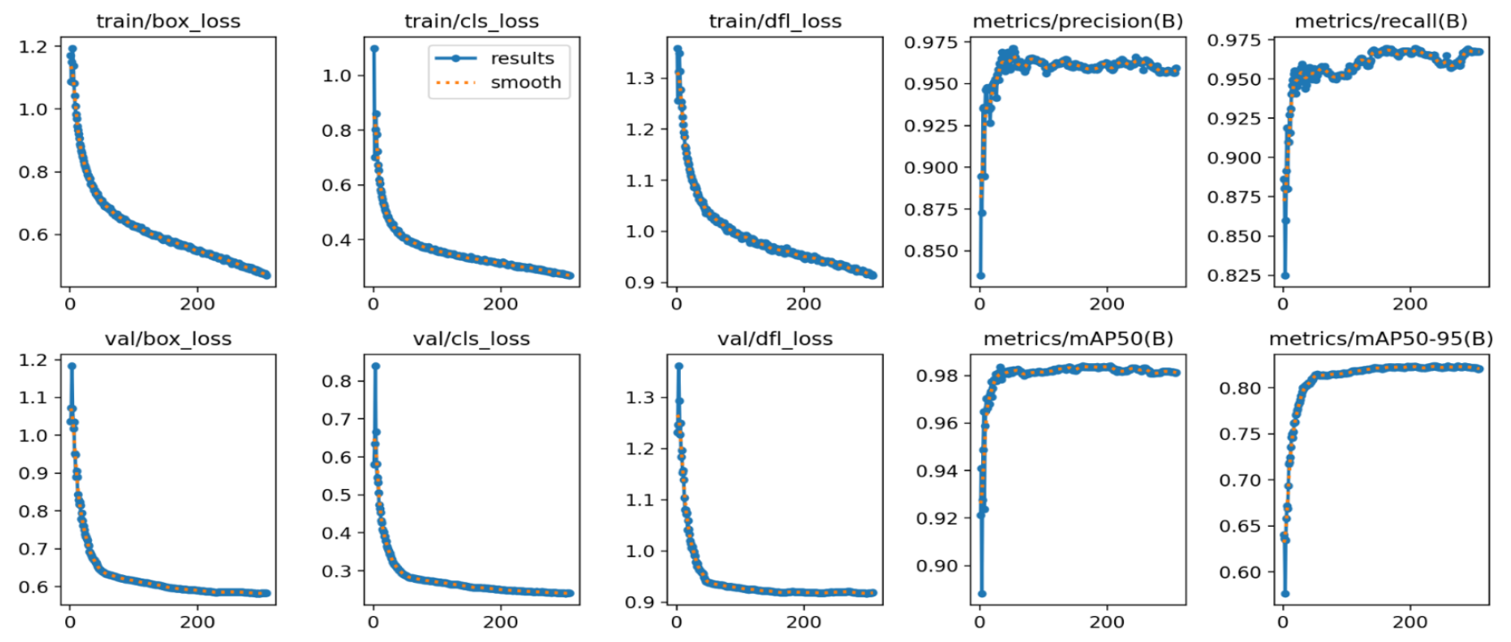

Figure 6: Train and Validation Results YOLOv8 Unified Model

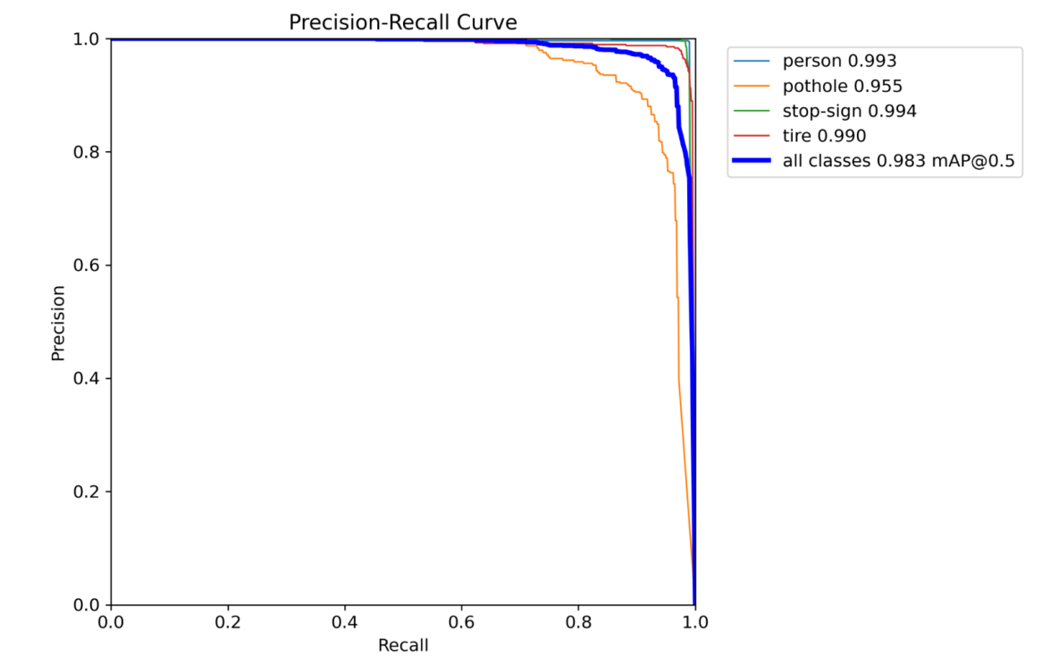

Figure 7: YOLOv8 Precision-Recall Curve Unified Model

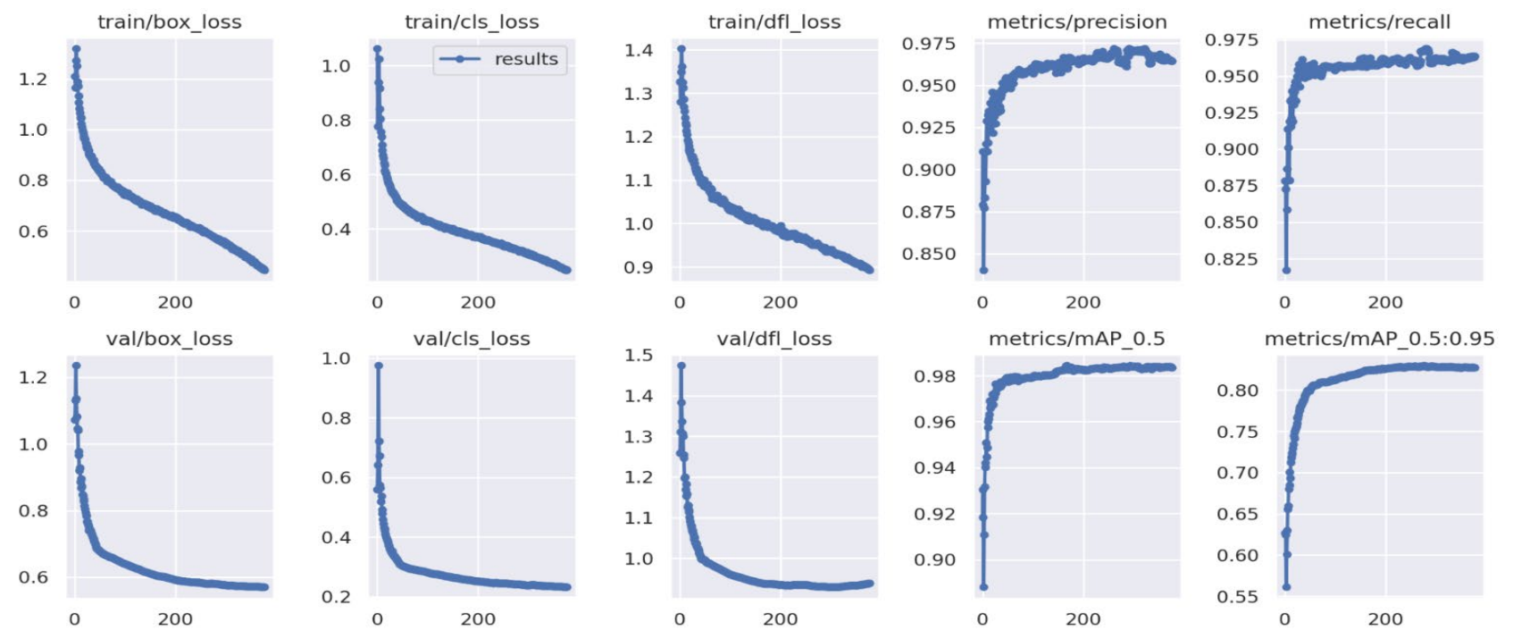

Figure 8: Train and Validation Results YOLOv9 Unified Model

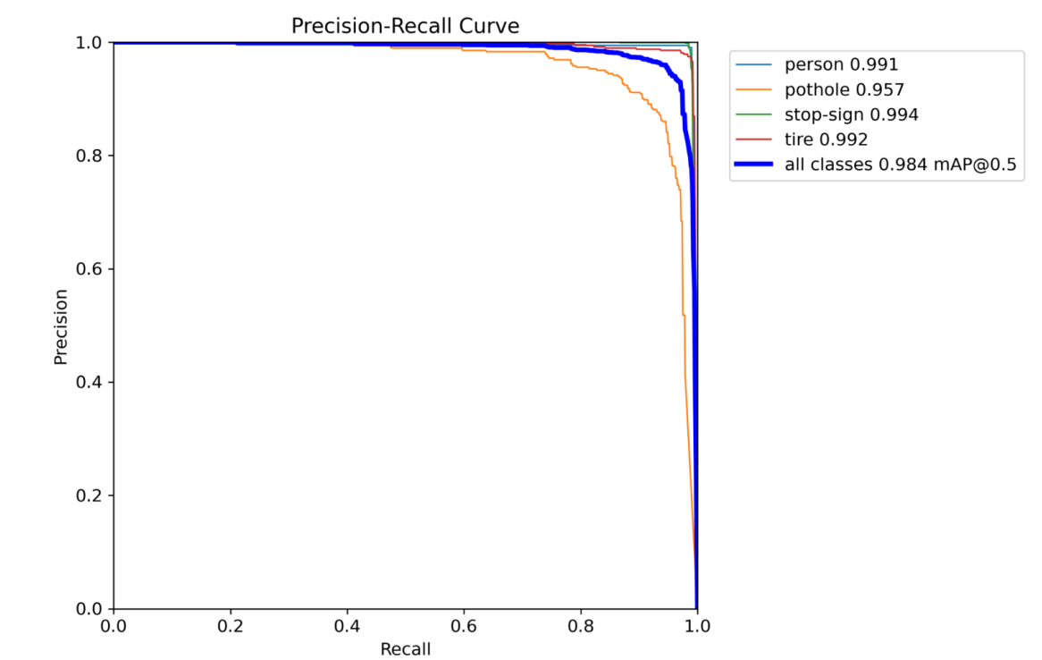

Figure 9: YOLOv9 Precision-Recall Curve Unified Model

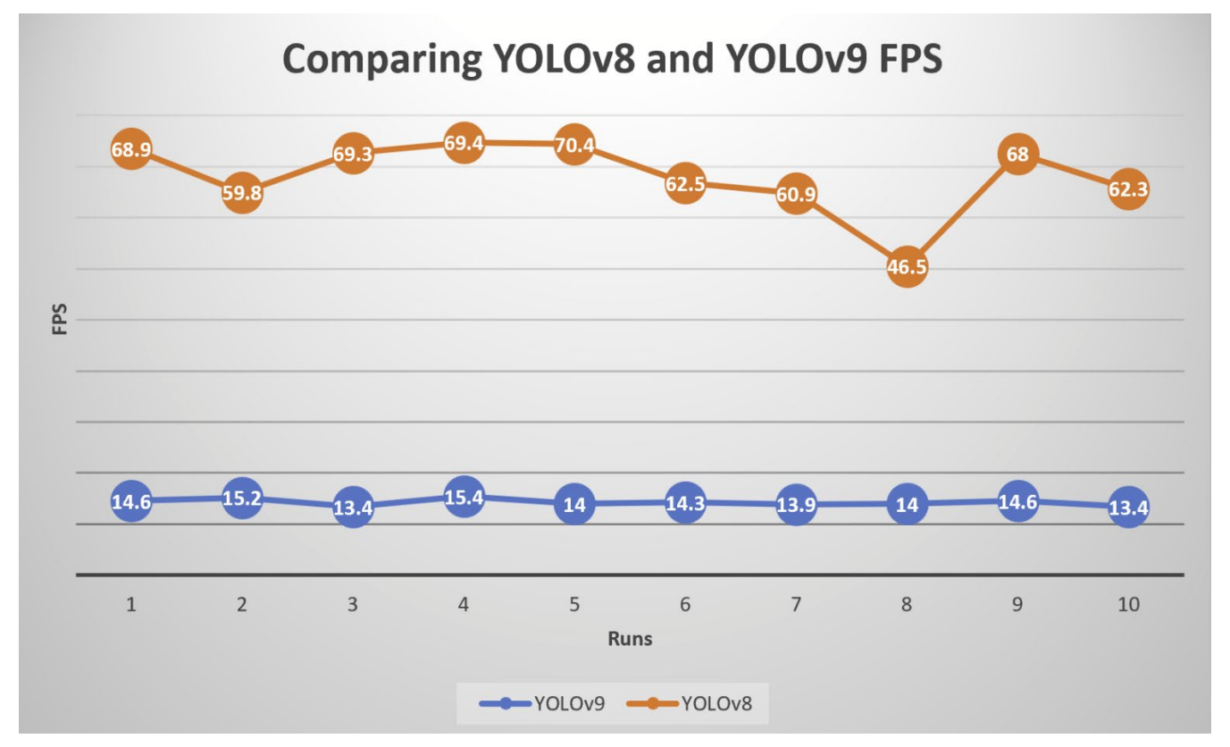

Figure 10: This graph shows the FPS of a 4K 120 FPS video feed through each model YOLOv8 and YOLOv9. YOLOv8 utilized the predict function, and YOLOv9 used the inference function since it cannot predict yet. Predict returns the bounding box of the class, and inference returns what classes are in the photo

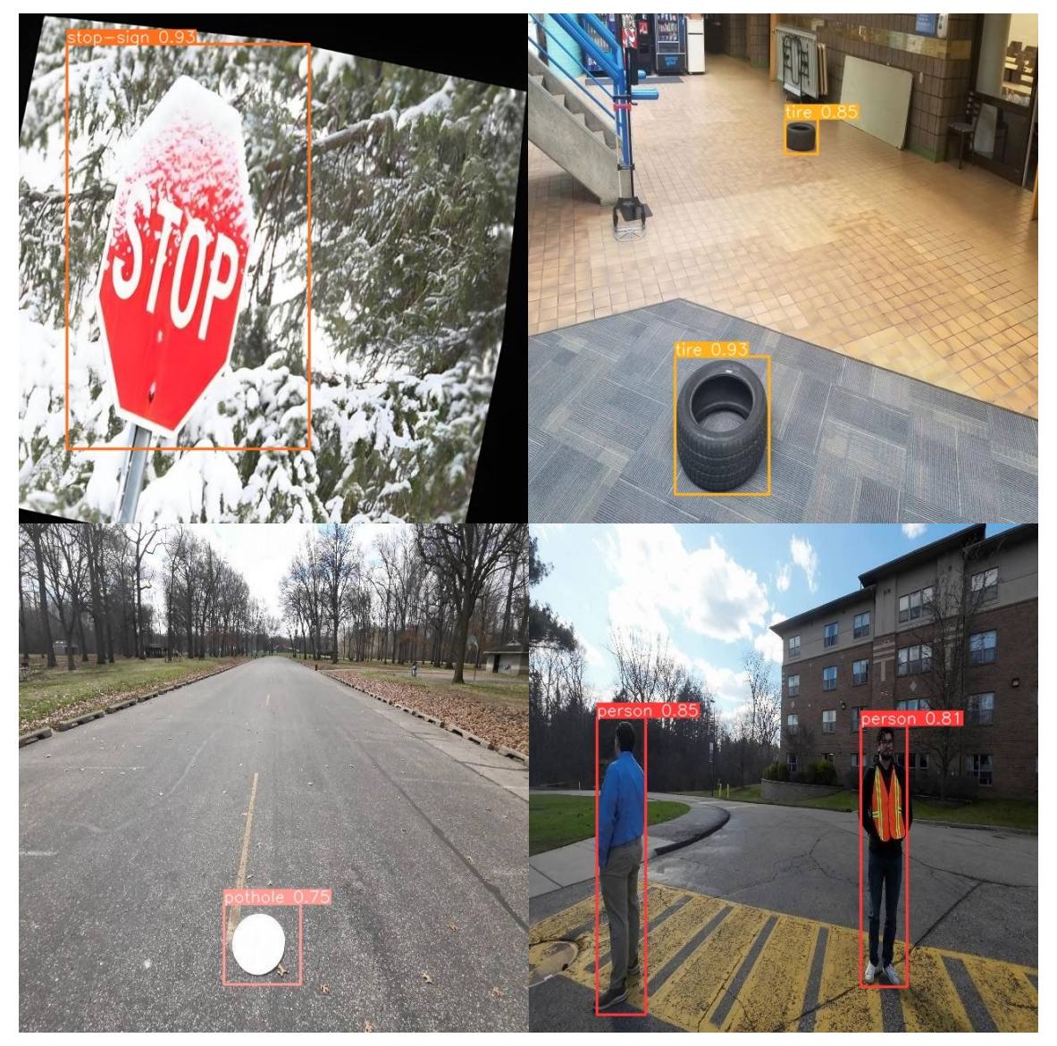

Figure 11: YOLOv8 Unified Model Prediction Results

DEVELOPMENT OF A UNIFIED MODEL

Reasoning for a Unified Model:
The development of unified model would be a more efficient way to detect all of the objects for the purposes of the IGVC competition. Only having to load one model's weights should put less stress on the GPU during real-time inferencing. Merging all of the datasets also gives the unified dataset more diversity and balance, which was an issue with some of our previous datasets for the individual models. The idea was that a more balanced dataset would lead to a better performing model.

Merging the Datasets:
The merging of our previous datasets was necessary for the training of a unified model. This process involved exporting all of our individual model datasets from roboflow, then combining them into one unified dataset. The labels for the unified datasets included person, stop sign, pothole, tire, and null. The pedestrian-without-vest and pedestrian-with-vest classes were both reclassified into the person class. The initial purpose of the two classes was to balance the dataset for the individual pedestrian model. This was no longer necessary with the merging of the other datasets, which would naturally create more balance among the classes. The stop-sign-fake class was added to the null class, as it was not actually necessary to detect the fake stop sign for the competition, just to not detect it as a real stop sign. The other three stop sign classes were merged into the stop-sign class, because whether an image was classified as a stop-sign, stop-sign-vandalized, or stop-sign-obstructed classification, it was still a real stop sign for which detection was necessary.

Model Performance Compared with Previous Implementations:
The results from the unified model trained with YOLOv8 compared with the previous stop sign and tire detection models::

UNIFIED MODEL

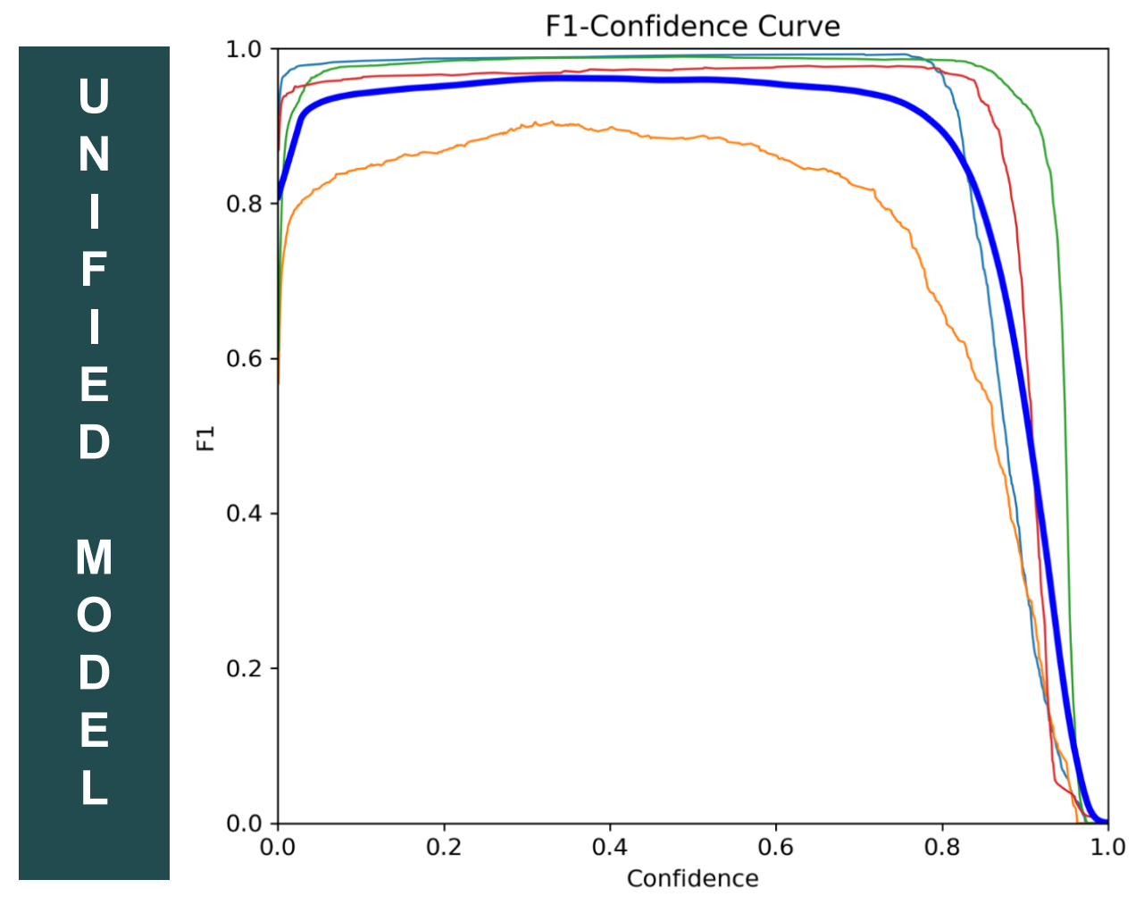

Figure 12: F1-Confidence Curve for YOLOv8 Unified Model

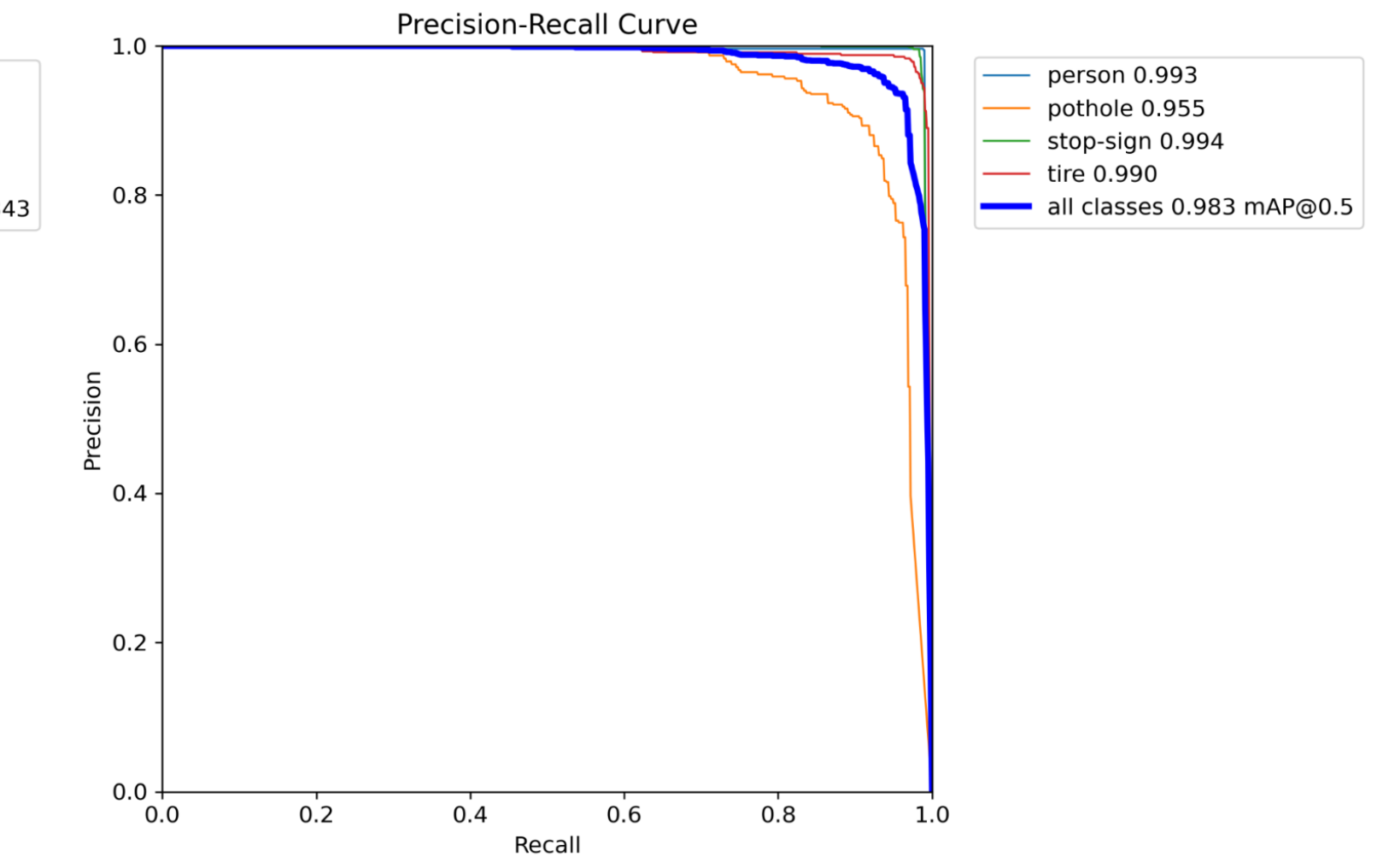

Figure 13: Precision-Recall Curve for YOLOv8 Unified Model

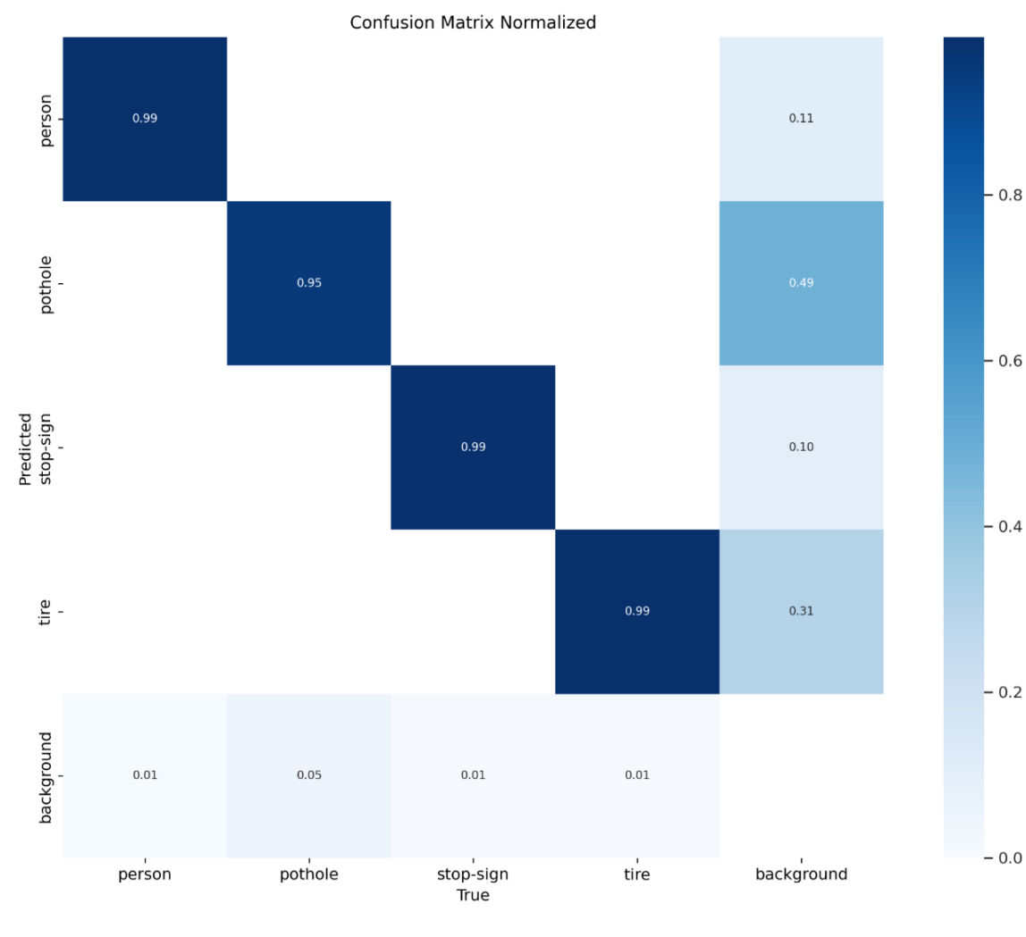

Figure 14: Normalized Confusion Matrix for YOLOv8 Unified Model

PREVIOUS STOP SIGN

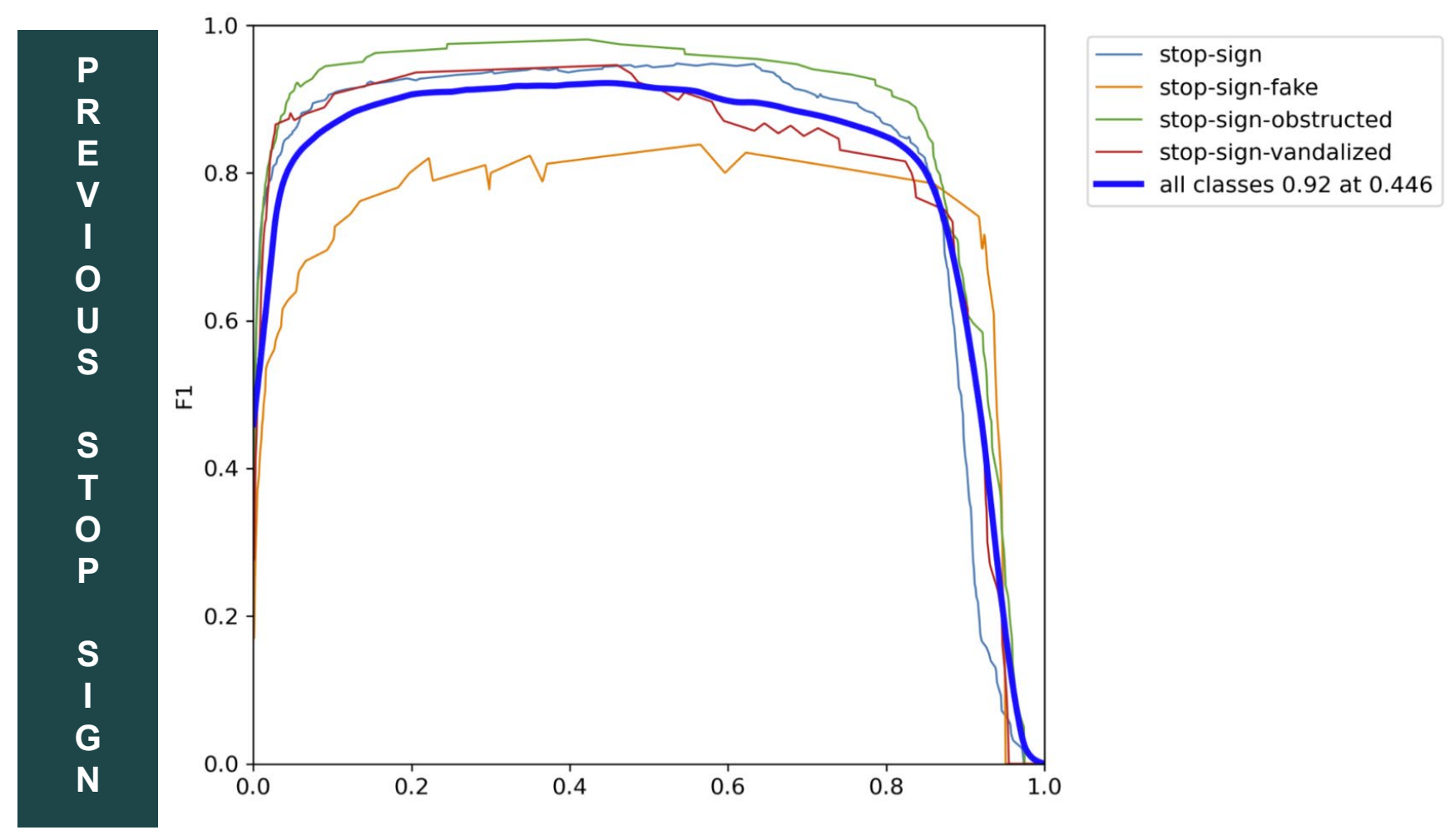

Figure 15: Previous Stop Sign YOLOv8 Model F1-Confidence Curve

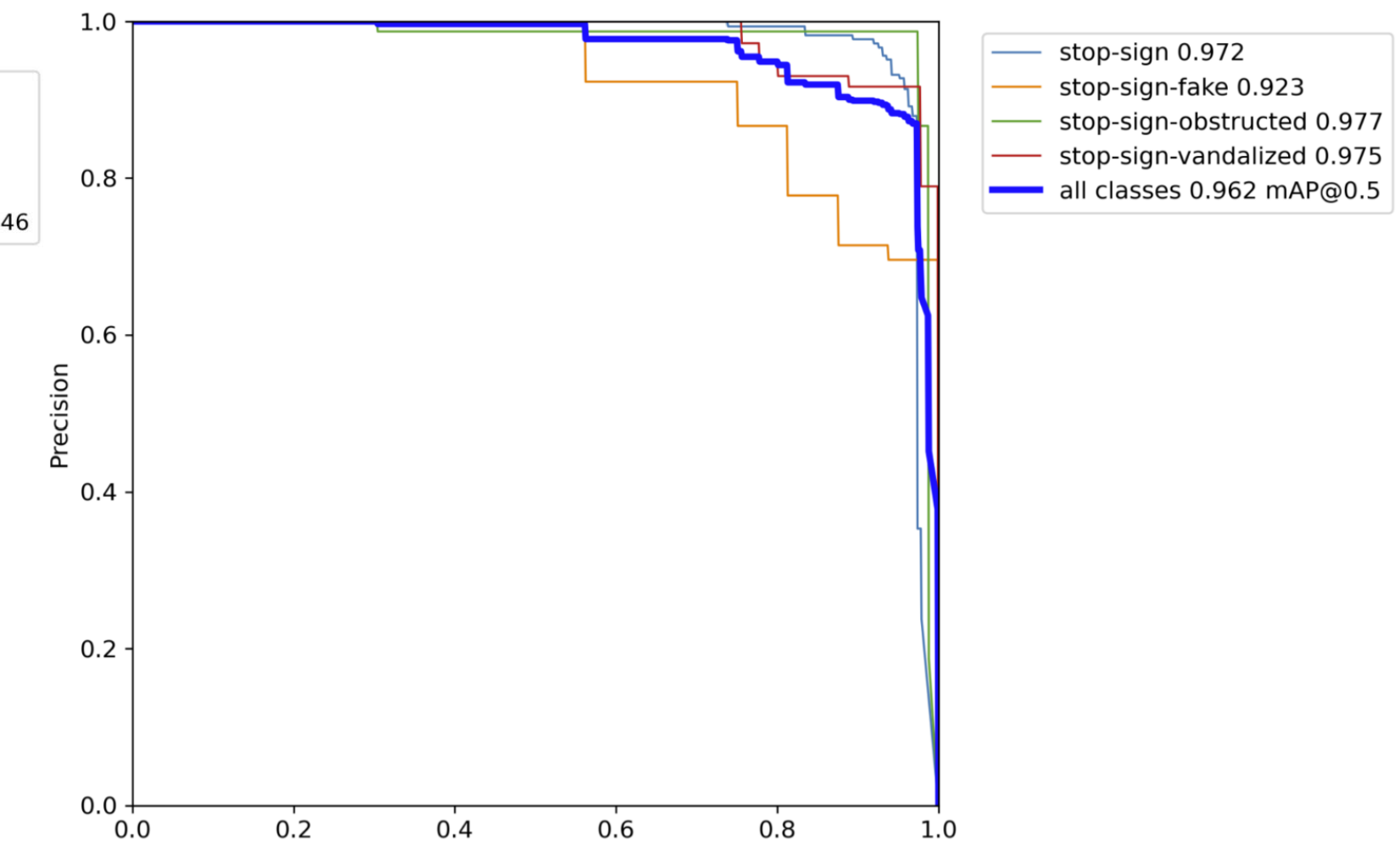

Figure 16: Previous Stop Sign YOLOv8 Model Precision-Recall Curve

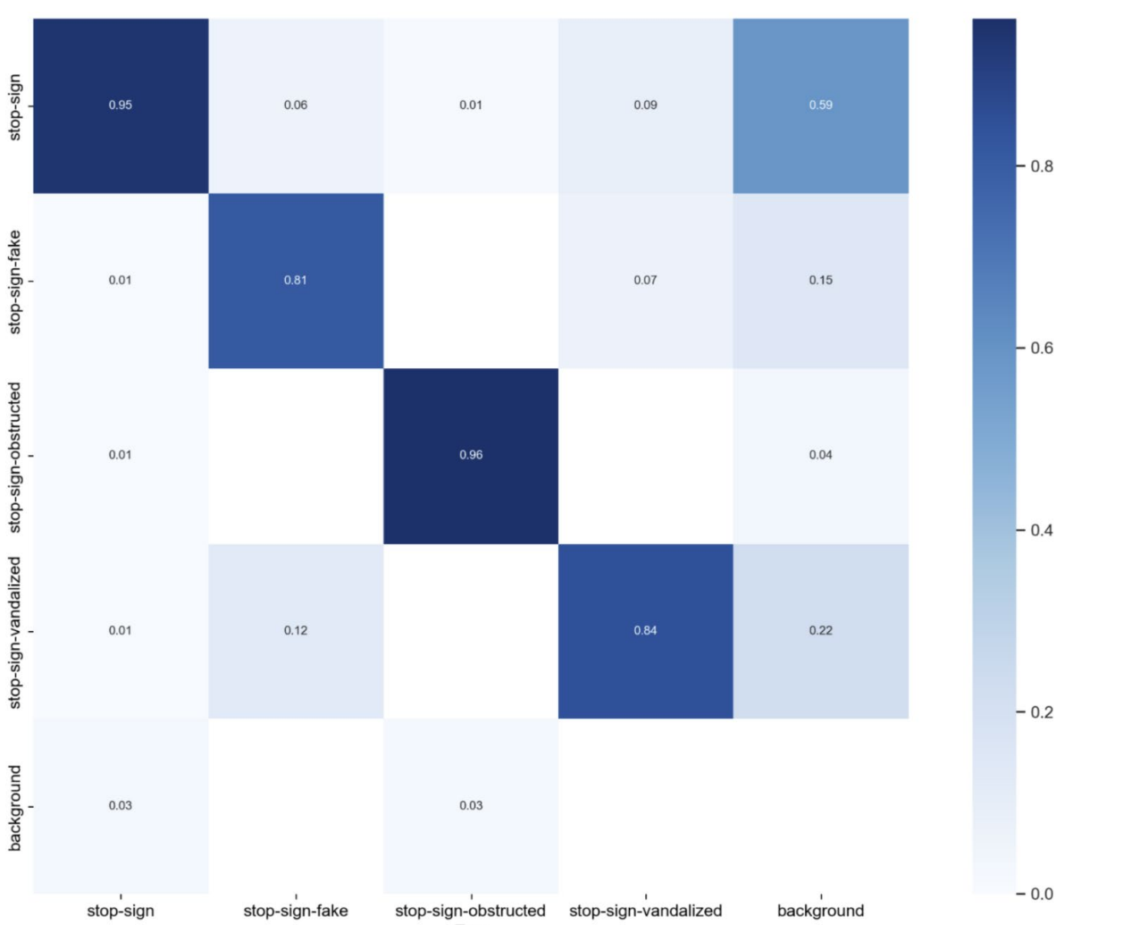

Figure 17: Normalized Confusion Matrix for Previous Stop Sign YOLOv8 Model

PREVIOUS TIRE

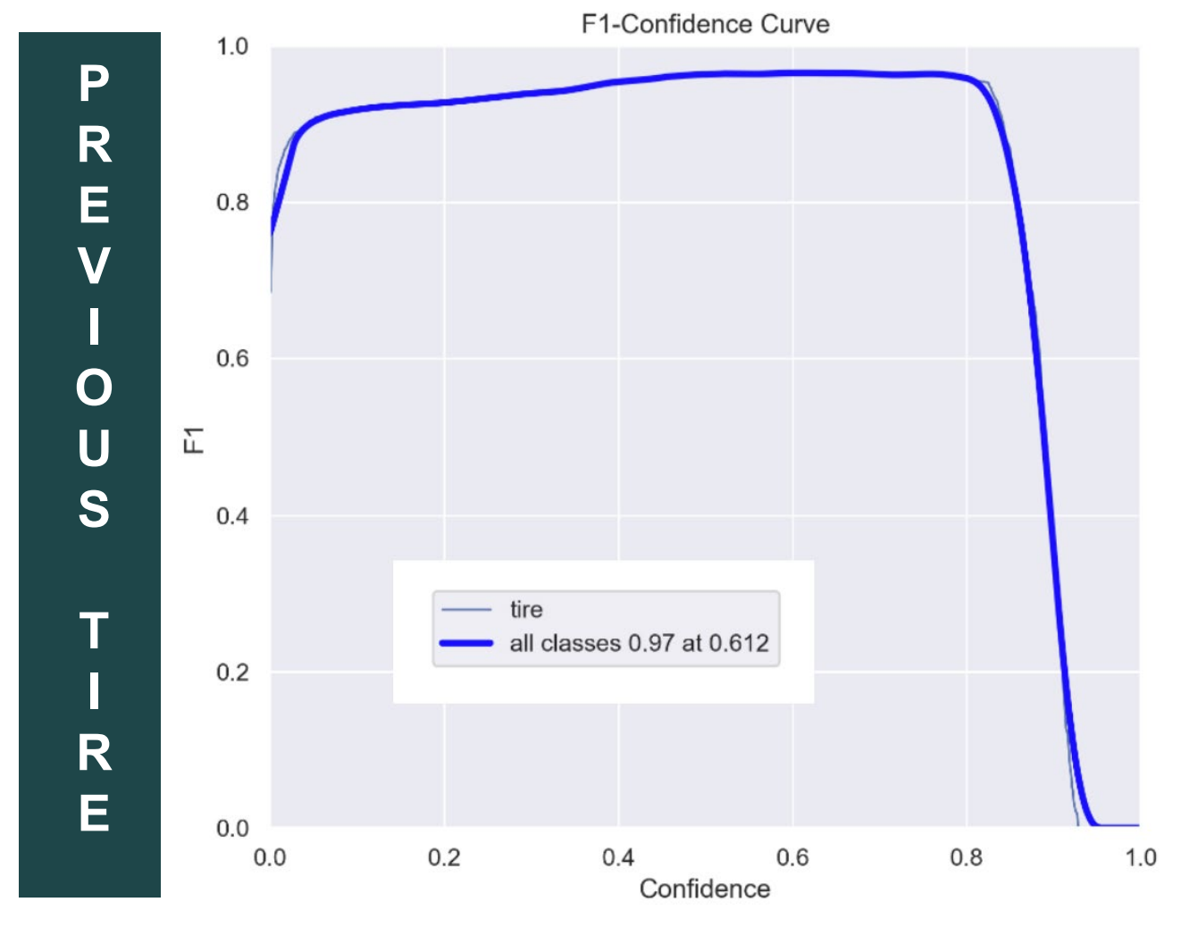

Figure 18: Previous Tire YOLOv8 Model F1-Confidence Curve

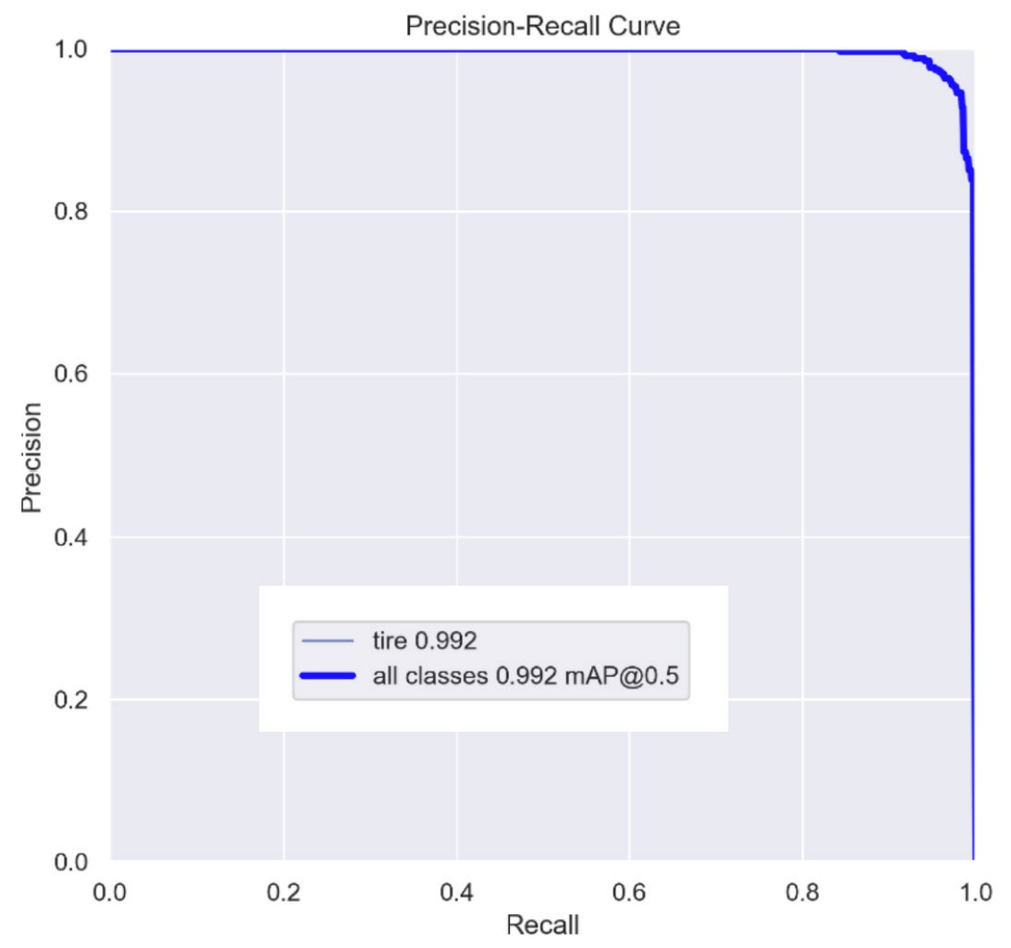

Figure 19: Previous Tire YOLOv8 Model Precision-Recall Curve

Comparison Conclusions:
The unified model performs comparably to the previous models, better in the stop sign case.In the process of converting the stop-sign-fake images into null images along with combining the other three stop-sign classes, the accuracy of the unified model for stop signs was greatly improved from our own individual models. The accuracy of the person class is improved over the accuracy of the pedestrian-with-vest and pedestrian-without-vest classes from the previous iterations of the unified model. In conclusion the unified model seems to be the best model to use for the object detection needs of the IGVC team.

FUTURE WORK

Due to various shortcomings, and the current development of YOLO models, there are several places where this work can be improved upon and continued in the future.

Diverse Data Collection: Collecting more fake stop sign data is important in improving the accuracy of finding fake stop signs. It is a very challenging class to identify, and needs more data. It is very underrepresented in the current data set being used.

Development of YOLOv9
YOLO version 9 is not fully developed at this time. It cannot be fully utilized, and assessed yet. YOLOv9 is supposed to be very different structurally from past versions. It performs roughly the same in accuracy to version 8, but with future developments it may surpass version 8 in accuracy, among other metrics.

Deployment and Integration onto the ACTOR Vehicles
Currently, the unified model has not been deployed onto the ACTOR vehicles. In the near future, they will be integrated onto the car and the performance will be evaluated.

Testing of Other Metrics
Currently, the accuracy of the model is the main criteria being evaluated. In the future the VRAM required to operate the model will be evaluated.

REFERENCES

[1] V. Singal, "YOLOv9 – Object Detection – Part 1," *AI Trends*, Feb. 25, 2024.https://medium.com/ai-trends/yolov9-object-detection-with-programmable-gradient-information-pgi-and-generalized-efficient-4fa3352409cc (accessed Apr. 14, 2024).
[2] Ultralytics, "YOLOv9," *docs.ultralytics.com*. https://docs.ultralytics.com/models/yolov9/
[3] "Sign in to Roboflow," *app.roboflow.com*. https://app.roboflow.com/ltuws (accessed Apr. 14, 2024).
https://www.robofest.net/cj/StopSignDetection.pdf
https://www.robofest.net/AutoEV/TireDetection.pdf
http://www.igvc.org/