Image Recognition to Detect Different Vehicle Types in Riga

Eden Elisha Thomas, Dmitrijs Gizdevans, Gonzalo Gamez

Objective

This project focuses on training a model for accurate vehicle identification and classification within Riga, Latvia. The model is trained to recognize and categorize the following vehicle types

V (Passenger Vehicles): This category includes standard cars used for personal transportation.

C (Cargo Vehicles): This encompasses all commercial vehicles, from smaller vans (C1) to large trucks with trailers (C4).

S (Buses): This category specifically targets public transport buses.

Team Coordination and Delegation

Project Timeline

Week	Task/Deliverable	Responsible
1	Project requirements	Dmitrijs, Eden, Gonzalo
2	Data collection, data exploration	Dmitrijs, Eden
3	Data annotation, data formatting	Dmitrijs, Eden, Gonzalo
4	Model training and evaluation	Dmitrijs, Eden
5	Model improvement and tuning	Dmitrijs, Eden
6	Model testing with new data	Dmitrijs, Eden
7	Final Model adjustments and validation	Dmitrijs, Eden
8	Final presentation	Dmitrijs, Eden, Gonzalo

Table 1. Project Timeline

Team Roles

Name	Title	Roles
Dmitrijs	Data Engineer	Business Understanding, Data Preparation, Modelling, Test & Validate
Eden	Data Analyst	Business Understanding, Data Understanding, Modelling, Communication of Insights
Gonzalo	Project Manager	Management, Support

Table 2. Team Roles

	BUSINESS UNDERSTANDING	DATA UNDERSTANDING	DATA PREPARATION	MODELLING	TEST & VALIDATE	DEPLOYMENT	COMMUNICATION OF INSIGHTS	ONGOING ASSESMENT
DATA SCIENTIST	■	■	■	■	■		■	
DATA ANALYST	■	■					■	■
DATA ENGINEER		■	■		■	■		

Table 3. Data Science Competency Skills (Data To Decisions CRC, 2017)

Team Coordination and Delegation

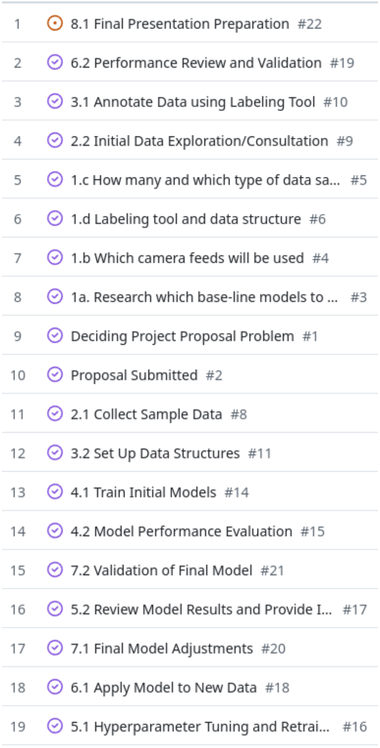


Fig 1. Backlog

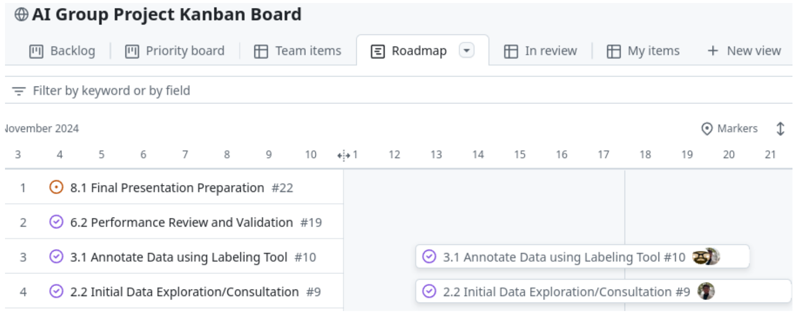


Fig 2. Kanban Board Roadmap

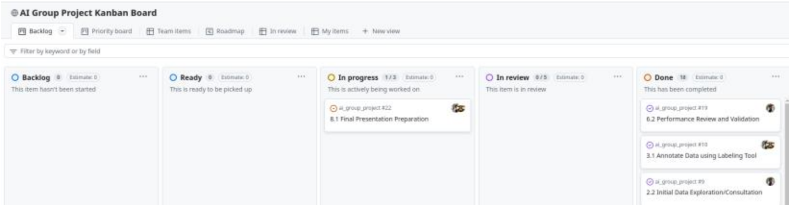
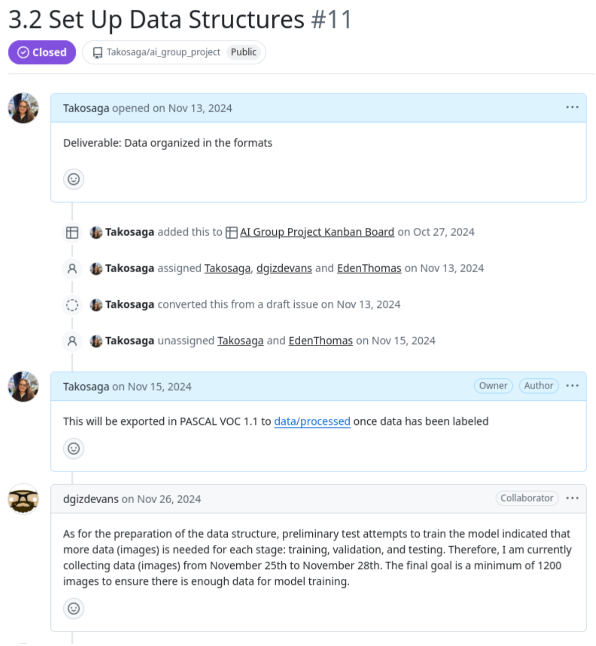


Fig 3. Kanban Board

Team Coordination and Delegation

Fig 4. Google Meets logo

Fig 5. Google Docs with other logo



3.2 Set Up Data Structures #11

Closed | Takosaga/ai_group_project | Public

Takosaga opened on Nov 13, 2024

Deliverable: Data organized in the formats

Takosaga added this to AI Group Project Kanban Board on Oct 27, 2024

Takosaga assigned Takosaga, dgizdevans and EdenThomas on Nov 13, 2024

Takosaga converted this from a draft issue on Nov 13, 2024

Takosaga unassigned Takosaga and EdenThomas on Nov 15, 2024

Takosaga on Nov 15, 2024 | Owner | Author

This will be exported in PASCAL VOC 1.1 to data/processed once data has been labeled

dgizdevans on Nov 26, 2024 | Collaborator

As for the preparation of the data structure, preliminary test attempts to train the model indicated that more data (images) is needed for each stage: training, validation, and testing. Therefore, I am currently collecting data (images) from November 25th to November 28th. The final goal is a minimum of 1200 images to ensure there is enough data for model training.

Fig 6. Task write ups

Fig 7. Whatsapp logo

Research

- **Two-Stage Detectors:** Region proposal + classification/refinement (e.g., Faster R-CNN). Higher accuracy, higher complexity.
- **One-Stage Detectors:** Direct bounding box/class prediction (e.g., YOLO, SSD). Efficient, real-time capable.
- **Transformer-based Detectors:** Emerging architecture (e.g., Swin Transformer). State-of-the-art accuracy.

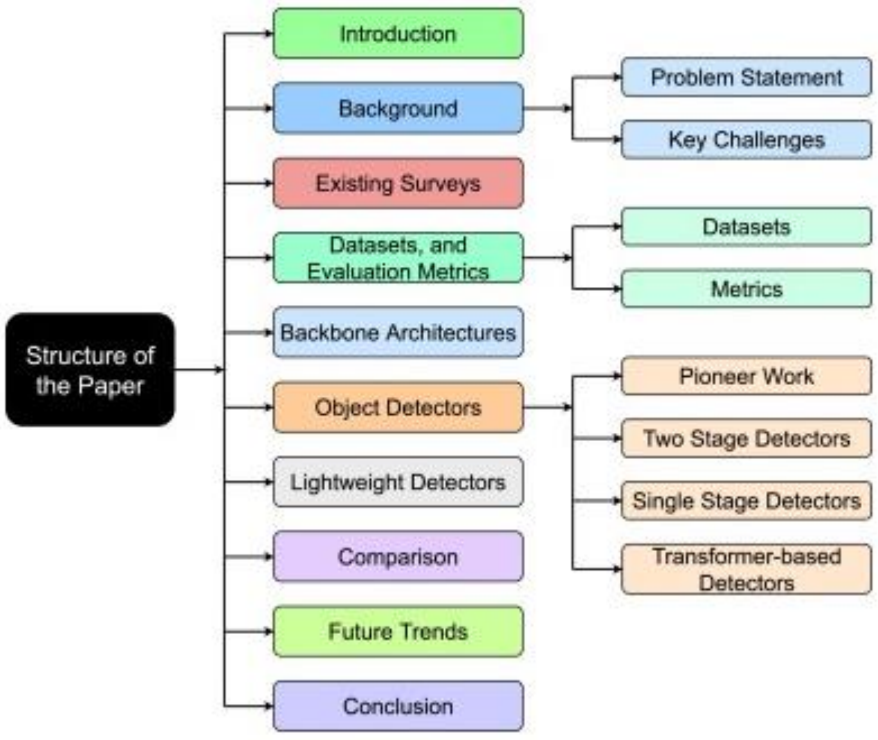


Fig 8. Survey of object detection by Zaidi et al. (2022)

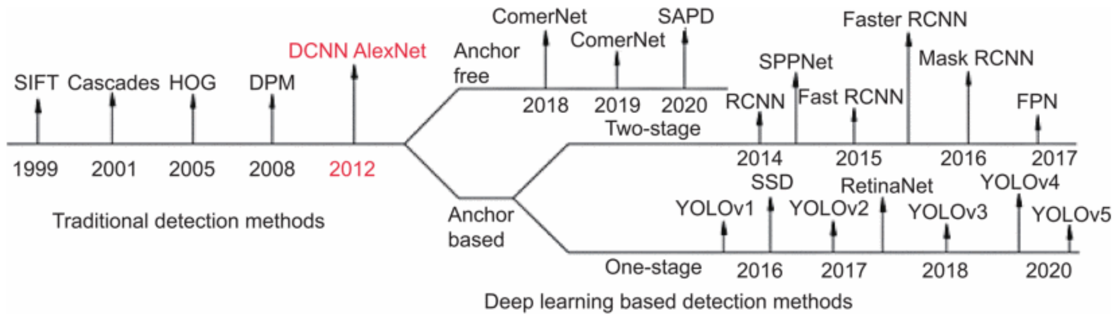


Fig 9. Development of Vehicle Detection(Ma and Xue, 2024)

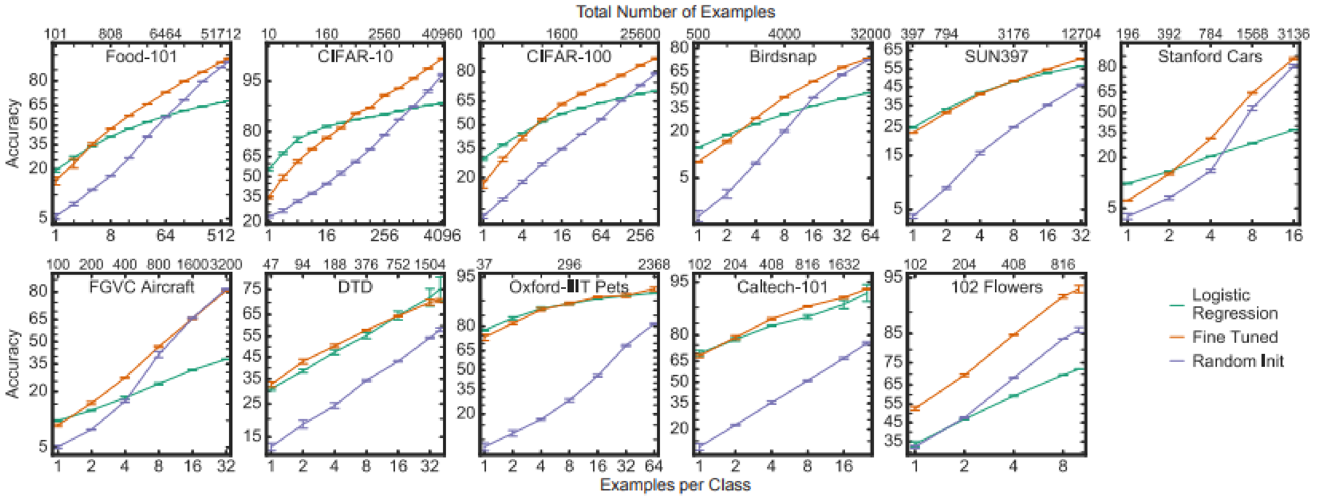


Fig 10. Accuracy for total number of examples and examples per class (Kornblith, Shlens and Le, 2019)

Research

Data Collection

Data Sources:

- Public video streams from webcams capturing traffic scenes.
- Extracted individual frames from streams to generate image datasets.

Challenges in Data Collection:

- Providers often restricted access to video streams.
- Required understanding of streaming technologies to bypass limitations.

Solution:

- Customized HTTP headers to emulate approved devices, enabling continuous access to streams.
- Extracted high-quality frames from public video streams using the **FFmpeg** library

Outcome:

- Successfully gathered a large, diverse dataset of vehicle images from live webcam feeds.
- Data collection process is detailed in the **webcam_capture.ipynb** notebook.

Data Storage:

- Collected images securely stored in **Google Cloud Storage (GCP buckets)** for scalability and integration into the ML pipeline.

Data Collection

Annotation Process

- Manual annotation was performed on the collected images using **CVAT (Computer Vision Annotation Tool)** to label the data accurately.

Labeled Classes

- Passenger Vehicles (V)
- Cargo Vehicles (C)
- Buses (S)

Dataset Splitting : Data was manually divided into three sets:

- Training Set (60%)
- Validation Set (20%)
- Testing Set (20%)

Challenges

- Labeling was manually done and distributed among group members, which was time-intensive.
- Ensuring consistency across annotations required thorough coordination and quality checks.

Model Training and Evaluation

Model Selection and Training

- **YOLOv8** was selected for its efficiency in object detection tasks.
- Pre-trained weights were used as a starting point, and the model was fine-tuned with our labeled data.

Training Parameters

- Epochs: 50
- Image Size: 640x640
- Batch Size: 16

Optimizer: AdamW

Evaluation Metrics

- Precision, Recall, mAP@0.5, and mAP@0.5:0.95 were used to measure performance on validation and test datasets.

Challenges

- Ensuring the model generalized well to unseen data.
- Debugging and resolving configuration issues during training.

Infrastructure

- **Google Cloud Platform** was utilized for saving and sharing training artifacts and results.

Data Augmentation and Enhanced Training

Objective:

- Enhance model performance by introducing variability and diversity into the training dataset.

Key Techniques:

- Applied transformations such as rotations, flips, brightness adjustments, and scaling using the **Albumentations** library.
- Integrated augmentation pipeline directly into the training data preparation workflow.

Process Overview:

- Original training dataset duplicated for augmentation to preserve baseline data integrity.
- Augmented images used alongside original data for retraining the YOLOv8 model.

Challenges:

- Managing increased storage and processing requirements for augmented datasets.
- Ensuring augmented images maintained realism and meaningfulness for model training.

Results:

- Augmented model demonstrated improved confidence and reduced variability in predictions across test sets.
- Metrics showed an approximate **2.2% improvement in average confidence** and a **5.5% reduction in standard deviation**.

Testing on Unlabeled Data

Objective:

- Evaluate the trained YOLOv8 model on unlabeled datasets to assess real-world performance.

Process Overview:

- Five independent unlabeled test sets were created and processed using the trained model.
- Predictions included bounding boxes and class labels (Passenger Vehicles, Cargo Vehicles, Buses).
- Metrics calculated:
  - Average Confidence
  - Median Confidence
  - Standard Deviation
  - Class Distributions

Challenges:

- Handling large-scale datasets and ensuring predictions aligned with real-world expectations.
- Ensuring consistent processing for fair comparisons between baseline and augmented models.

Results:

- Detailed metrics collected for each test set.
- Augmented model showed improved confidence and reduced variability compared to the baseline model.

Challenges and Lessons Learned

- **Key Challenges:**
  - **Data Collection:**
    - Limited access to video streams required bypassing restrictions using tools like ffmpeg and HTTP header modifications.
  - **Data Annotation:**
    - Manual labeling of a large dataset was time-intensive and required team coordination.
    - Annotation tasks were divided among team members but still consumed significant resources.
- **Model Training:**
  - Training the YOLOv8 model with augmented data required additional computation and storage resources.
  - Debugging issues like failed predictions and model inconsistencies.
- **Lessons Learned:**
  - Automation is critical for scaling tasks like annotation and augmentation to save time and resources.
  - Effective use of cloud platforms like Google Cloud significantly accelerates workflows.
  - Iterative testing and validation improve model reliability and performance.
  - The importance of documenting processes for reproducibility and collaboration.
  - Modern technologies and libraries allow for rapid development of AI solutions based on ready-made frameworks.
  - Preliminary planning and understanding of general principles greatly accelerate project workflows.

Evaluation of Models

- Evaluation of Base Model on Test Dataset
- Evaluation of Augmented Model on Test Dataset
- Comparison

- Evaluation of Base Model on Unlabeled Dataset
- Evaluation of Augmented Model on Unlabeled Dataset
- Comparison

Evaluation on Test Dataset

- Key Metrics: Precision, Recall, mAP@0.5, mAP@0.5:0.95, IoU, conf
- Test dataset contained labeled images for three classes: Passenger Vehicles (V), Cargo Vehicles (C), and Buses (S).
- Model evaluation conducted using the .val() function of YOLOv5.
- Evaluated the model across varying IoU thresholds (0.3 to 0.9).
- Confidence thresholds varied to assess precision and recall tradeoffs.
- Outputs included:
  - Class-wise metrics saved in CSV files (metrics_v.csv, metrics_s.csv, metrics_c.csv).
  - Overall metrics (metrics_all.csv).
  - Visualizations: Confusion Matrix, Precision-Recall, F1-Confidence curves etc.

Evaluation of Base Model on Test Dataset

Overview Metrics

- Precision: 75.27% (±2.21%)
- Recall: 78.98% (±2.93%)
- mAP@0.5: 82.12% (±1.91%)
- mAP@0.5:0.95: 71.28% (±0.85%)

Class Specific Metrics on the right

Category	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Passenger Vehicles (V)	79.19% (±3.93%)	84.54% (±2.74%)	87.98% (±2.13%)	72.57% (±1.06%)
Cargo Vehicles (C)	61.67% (±1.36%)	66.19% (±5.42%)	66.12% (±2.56%)	58.83% (±1.86%)
Buses (S)	84.96% (±1.98%)	86.19% (±0.83%)	92.27% (±1.31%)	82.45% (±0.62%)

Table 4. Evaluation of Base Model

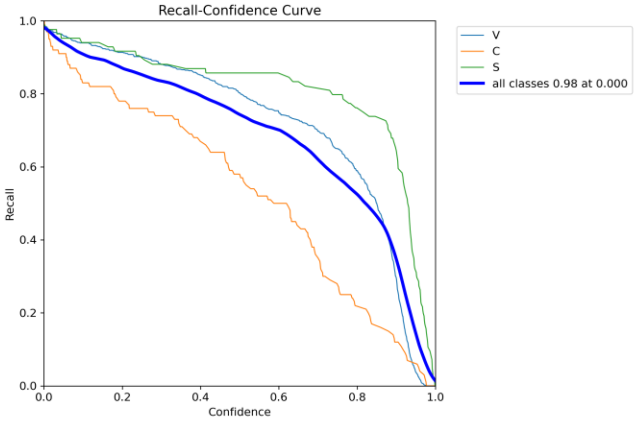


Fig 11. Recall-Confidence Curve Base Model

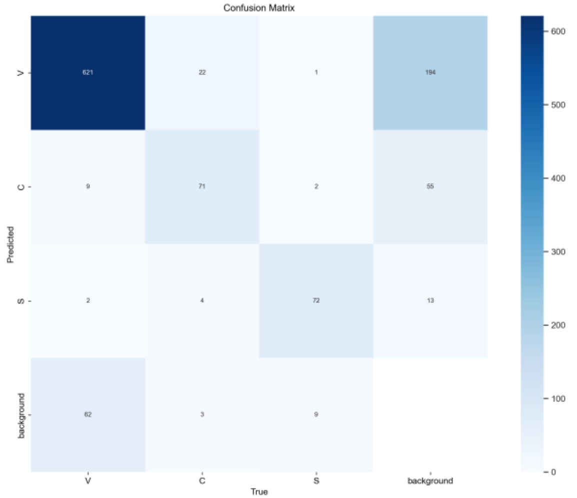


Fig 13. Heatmap Base Model

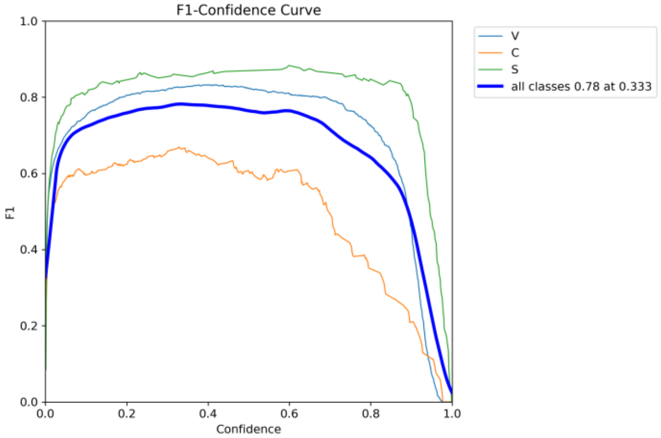


Fig 14. F1-Confidence Curve Base Model

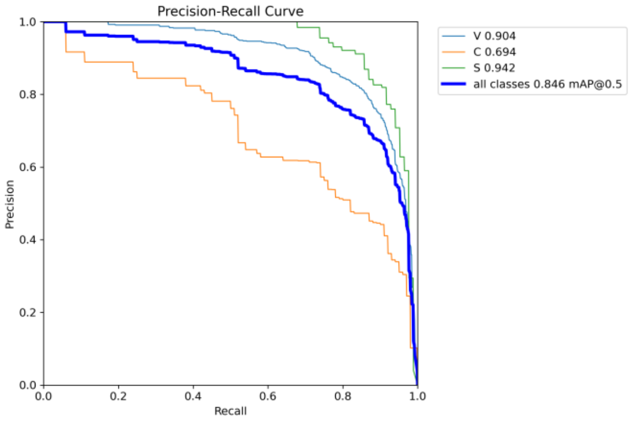


Fig 12. Precision-Recall Curve Base Model

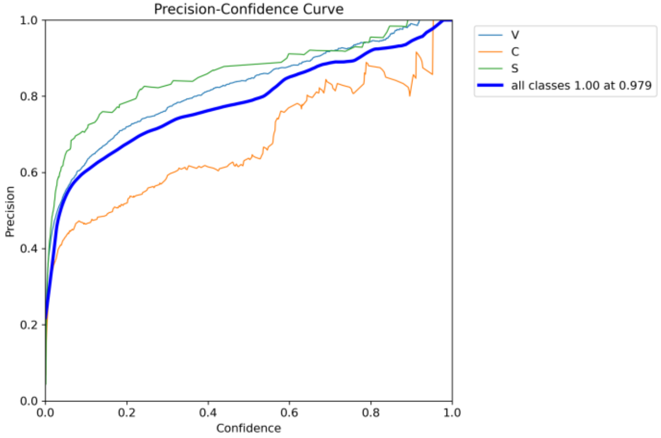


Fig 15. Precision-Confidence Curve Base Model

Evaluation of Augmented Model on Test Dataset

Overview Metrics

- Precision: 79.12% (±3.91%)
- Recall: 79.01% (±5.88%)
- mAP@0.5: 83.79% (±2.21%)
- mAP@0.5:0.95: 72.25% (±1.17%)

Class Specific Metrics on the right

Category	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Passenger Vehicles (V)	82.99% (±2.61%)	81.64% (±6.08%)	88.31% (±2.56%)	73.03% (±1.26%)
Cargo Vehicles (C)	64.45% (±6.98%)	70.88% (±7.35%)	71.75% (±2.43%)	62.31% (±1.79%)
Buses (S)	89.92% (±2.91%)	84.50% (±4.23%)	91.31% (±1.74%)	81.42% (±1.01%)

Table 5. Evaluation of Augmented Model

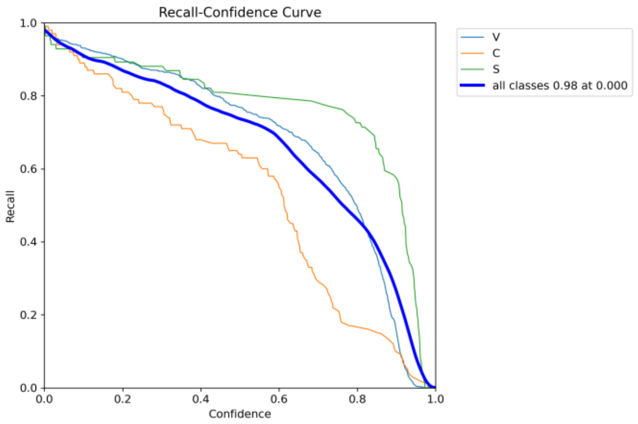


Fig 16. Recall-Confidence Curve Augmented Model

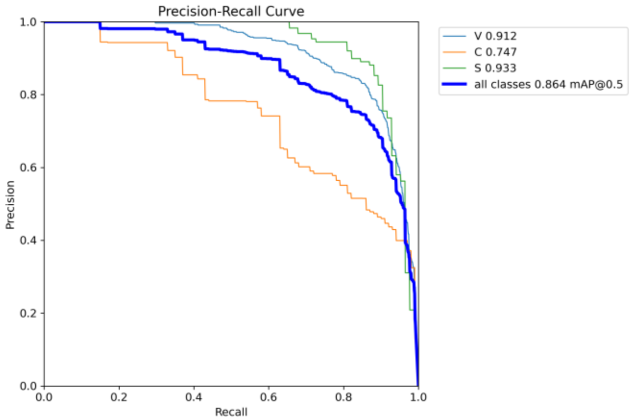


Fig 17. Precision-Recall Curve Augmented Model

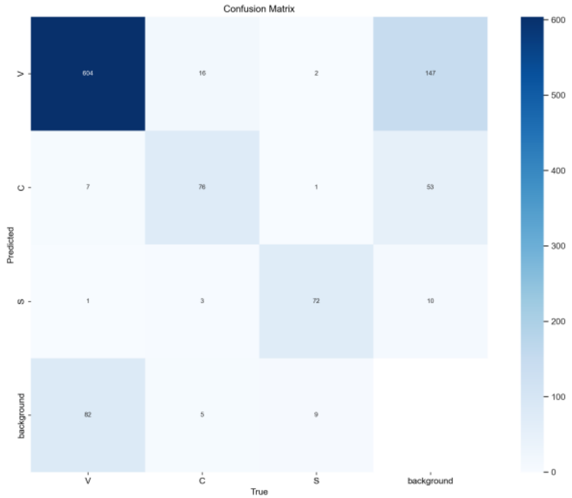


Fig 18. Heatmap Augmented Model

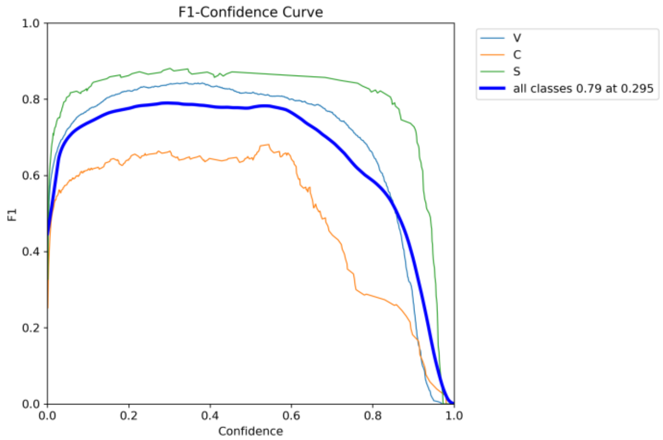


Fig 19. F1-Confidence Curve Augmented Model

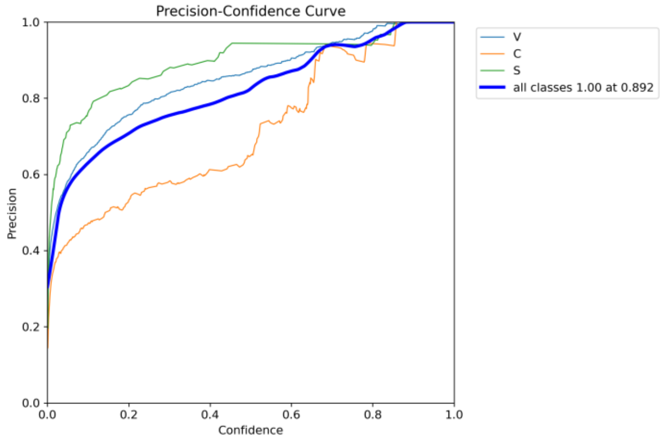


Fig 20. Precision-Confidence Curve Augmented Model

Comparison of Models

- Python scripts aggregated metrics for precision, recall, and mAP values.
- Comparison plots generated using matplotlib and pandas.

Comparison Summary Table

	p	r	map50	map50-95	Model	Class
0	0.752702	0.789753	0.821244	0.712850	Base Model	metrics_all
1	0.616660	0.661923	0.661203	0.588287	Base Model	metrics_c
2	0.849563	0.861895	0.922753	0.824546	Base Model	metrics_s
3	0.791881	0.845442	0.879776	0.725718	Base Model	metrics_v
4	0.791235	0.790089	0.837903	0.722568	Augmented Model	metrics_all
5	0.644559	0.708890	0.717524	0.623144	Augmented Model	metrics_c
6	0.899237	0.845001	0.913085	0.814242	Augmented Model	metrics_s
7	0.829908	0.816376	0.883102	0.730319	Augmented Model	metrics_v

Table 6. Comparision Summary Table

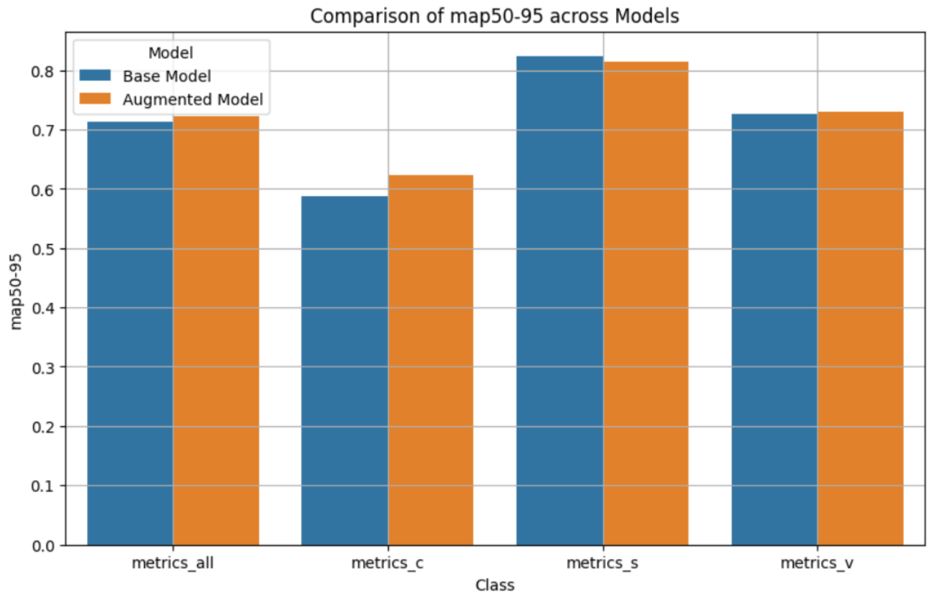


Fig 21. Comparison of map50-95

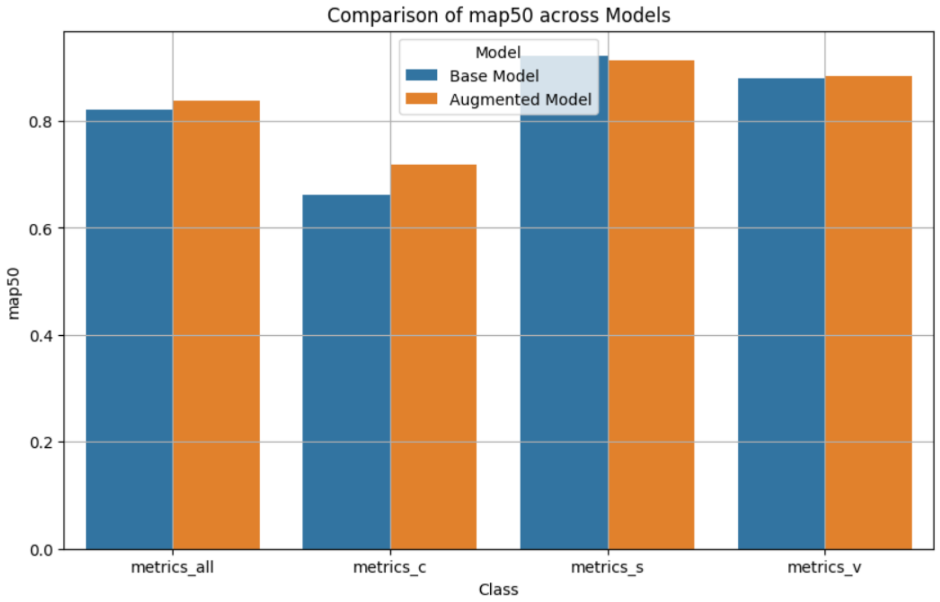


Fig 23. Comparison of map50

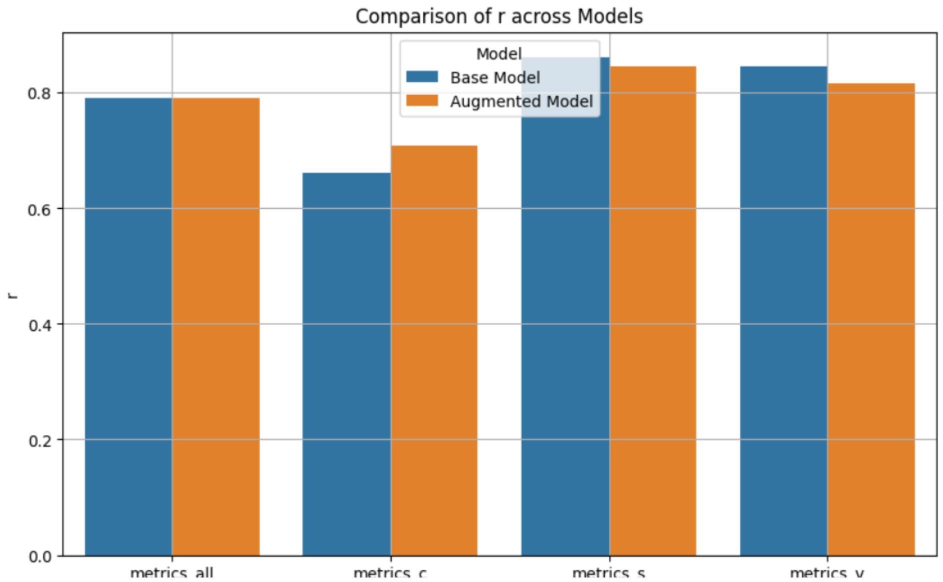


Fig 22. Comparison of r

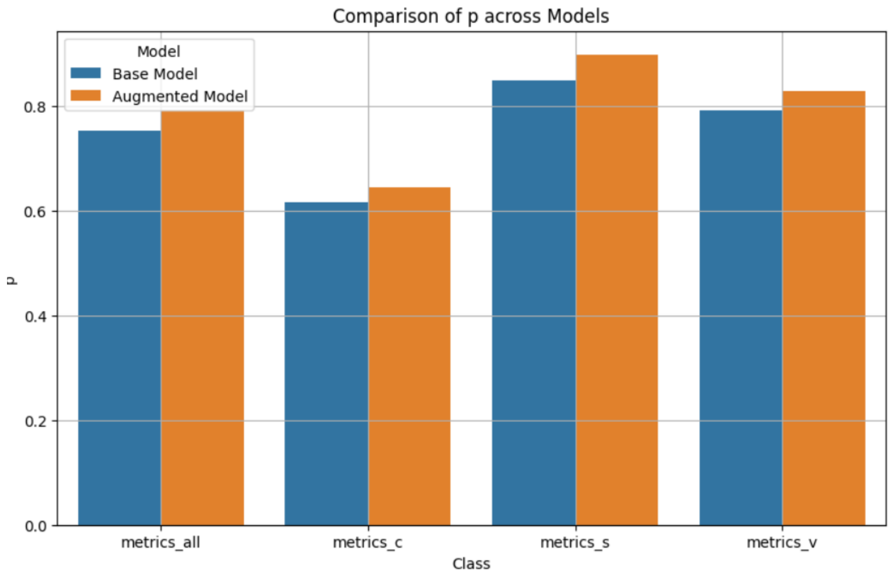


Fig 24. Comparison of p

Evaluation of Base and Augmented Model on Unlabeled Data

Key Evaluation Areas:

- Class Distributions
- Confidence Scores (average, spread, and variability)
- Agreement Ratios
- Bounding Box Dimensions (width and height)

Models used .predict() function to analyze the test sets.

Metrics derived include:

- Agreement ratios for model predictions.
- Average confidence scores and their variability.
- Class distribution percentages.
- Average bounding box dimensions.

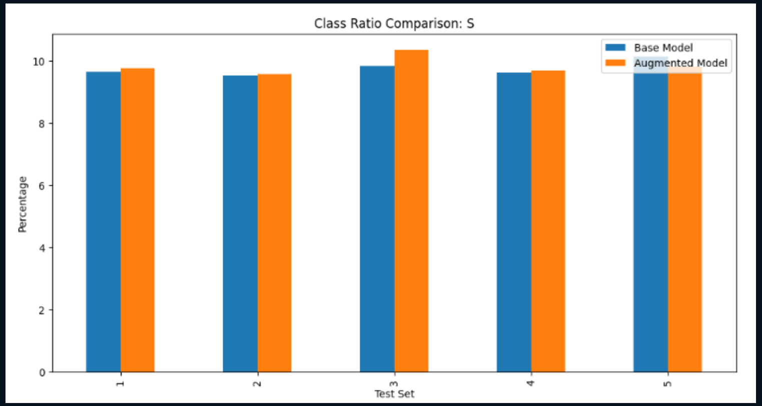


Fig 25. Class Ratio Comparision: S

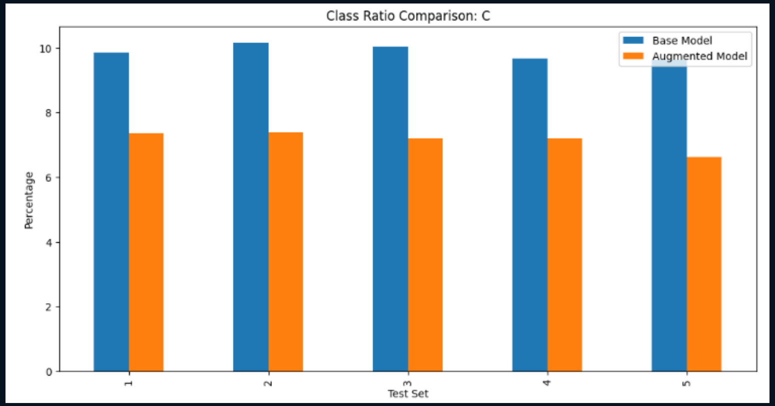


Fig 26. Class Ratio Comparision: C

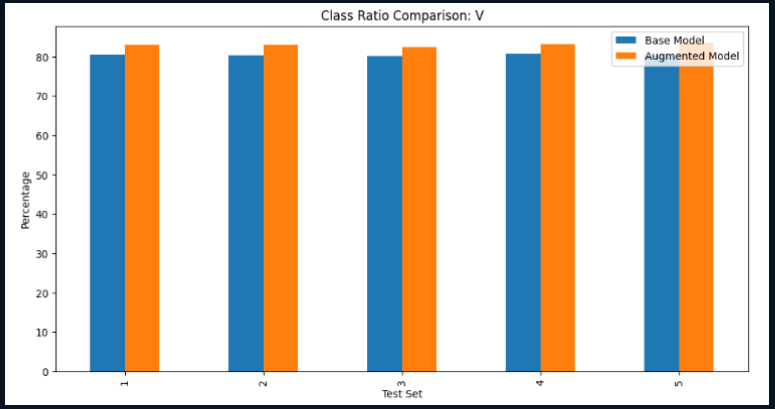


Fig 27. Class Ratio Comparision: V

Class Distribution Comparison

- Base Model:
  - Passenger Vehicles (V): 80.30%
  - Cargo Vehicles (C): 10.11%
  - Buses (S): 9.59%
- Augmented Model:
  - Passenger Vehicles (V): 83.01% (increase)
  - Cargo Vehicles (C): 7.16% (decrease)
  - Buses (S): 9.83% (slight increase)

Confidence Score Comparison

- Overall Confidence Scores:
  - Base Model: 0.8099
  - Augmented Model: 0.7833 (slightly lower overall confidence)
- Confidence Variability Across Test Sets:
  - Test Set 5: Base = 0.8079, Augmented = 0.7814 .
  - Test Set 4: Base = 0.8109, Augmented = 0.7838 .
  - Test Set 3: Base = 0.8106, Augmented = 0.7822 .
  - Test Set 2: Base = 0.8115, Augmented = 0.7863 .
  - Test Set 1: Base = 0.8084, Augmented = 0.7827 .

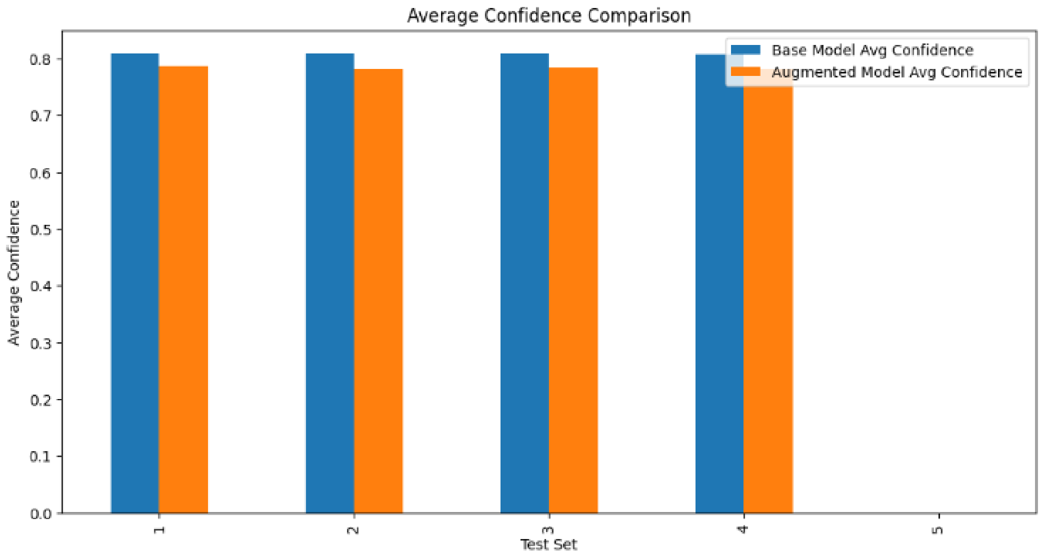


Fig 28. Average Confidence Comparision

Agreement Ratio Comparison

- Agreement ratios measure prediction consistency between models.
- Findings:
  - Agreement ratio ranges from 3.45% to 4.43% across test sets.
  - Indicates divergent behavior in predictions for some samples.

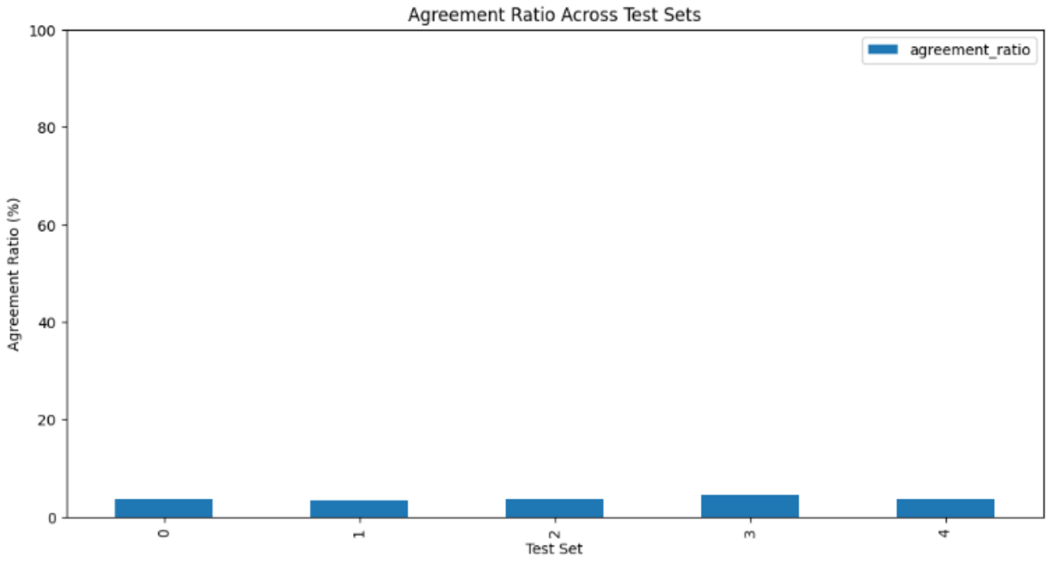
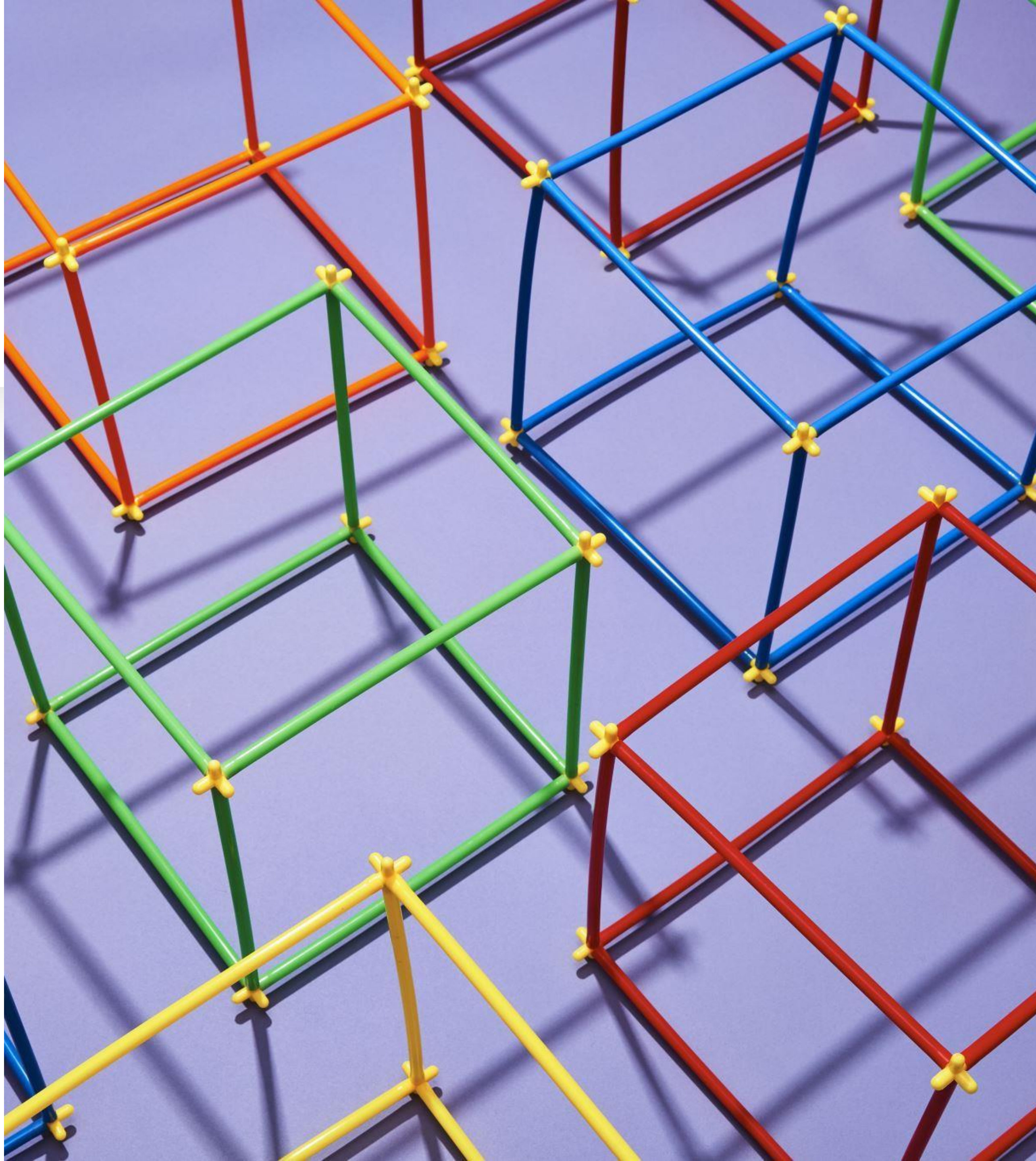


Fig 29. Agreement Ratio

Average Bounding Box Comparison

- Base Model:
  - Average width: ~120 pixels
  - Average height: ~60 pixels
- Augmented Model:
  - Average width: Slightly reduced compared to the base model.
  - Average height: Consistently smaller than base model.



Unlabeled Data Evaluation Summary

- Class Distributions:
  - Augmented model predicts fewer Cargo Vehicles but more Passenger Vehicles.
  - Similar performance on Buses across models.
- Confidence Scores:
  - Base model maintains higher confidence overall.
  - Augmented model shows consistent confidence spread across test sets.
- Agreement Ratios:
  - Low agreement ratios indicate significant differences in predictions.
- Bounding Box Dimensions:
  - Augmented model predicts smaller boxes, potentially improving specificity.

Conclusions and Recommendations

Questions?

Reference list

- Kornblith, S., Shlens, J. and Le, Q.V. (2019). Do Better ImageNet Models Transfer Better? *arXiv:1805.08974 [cs, stat]*. [online] Available at: https://arxiv.org/abs/1805.08974.
- Ma, C. and Xue, F. (2024). A Review of Vehicle Detection Methods Based on Computer Vision. *Journal of Intelligent and Connected Vehicles*, 7(1), pp.1–18. doi:https://doi.org/10.26599/jicv.2023.9210019.
- Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M. and Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126, p.103514. doi:https://doi.org/10.1016/j.dsp.2022.103514.

Presentation Rubric

Should cover the main areas documented in the project's technical report, including:

- - research conducted on the topic area.
- - team coordination, delegation, and project management. (Gonzalo)
- - technical implementation, including challenges encountered and how they were addressed.
- - conclusions, recommendations and reflection

Final presentation (30%, rubric components are equally weighted)				
Presentation delivery (Team)	Very hard to follow/unsuitable for a lay audience.	Hard to follow in places, lacking logical flow	Largely easy to follow, understandable, good quality presentation materials	Highly coherent, understandable, all key terms explained, well-rehearsed
Research depth and knowledge (Team)	Research wholly lacking, poor quality/ quantity or largely disconnected from the project focus.	Reasonable knowledge of the topic. Some research, though patchy in places or not based on good/current sources.	A clear grasp of the topic area, though perhaps unevenly distributed in the team, supported by generally good depth and breadth of research	Excellent team knowledge of the problem area shared throughout the team. Contemporary and quality evidence reviewed and actively/effectively harnessed.
QA session (Individual)	Unable to satisfactorily answer questions relating to the topic/project implementation	Some unsatisfactory answers but clear evidence of understanding of project implementation and outcomes	Good understanding of the topic and research implementation, but overall knowledge in the domain is limited	Excellent understanding of the topic area; clear and nuanced awareness of the scope and limitations of the project outcomes