Data Processing and Augmentation Techniques for Vehicle Counting and Detection in the Philippines

Ethyl Joy I. Badiang, Antonio C. Briza, Eduardo D. Piedad Jr, Elmer C. Peramo
Automated Labelling Machine,
DOST-Advanced Science and Technology Institute,Quezon City, Philippines
ethyljoy.badiang,antonio.briza,eduardojr.piedad,elmer@asti.dost.gov.ph

***Abstract*—Manual vehicle counting in the Philippines is challenged by rapid urbanization, while artificial intelligence (AI) offers a solution, hindered by the lack of representative vehicle datasets. This study addresses the dataset gap in vehicle detection by applying data engineering techniques to improve quality and mitigate class imbalance. Three methods were tested: grayscaling, blocking overrepresented classes, and grouping classes with similar instance counts. Grayscaling notably improved night-time detection, boosting precision and recall by 247% and 155%, respectively. Grouping cars and motorcycles yielded significant gains in both day and night settings, with day-time precision and recall increasing by 351% and 279%, and night-time by 157% and 164%. These results highlight the vital role of data preprocessing and balanced class distribution in optimizing AI performance for traffic management.**

***Index Terms*—Artificial intelligence, Data augmentation**

I. Introduction

Urbanization often drives increased automobile ownership in cities. As urban areas expand, the demand for personal vehicles increases, influenced by economic growth and the development of infrastructure that supports private transportation [1]. Establishing an efficient traffic monitoring system is crucial to managing congestion proactively, promoting strategic urban development [2].

Traditional traffic monitoring methods, such as manual vehicle counting by traffic personnel, are labor-intensive and prone to inaccuracies. Traffic officers may work extended hours under varying weather conditions, posing risks to their health and compromising data accuracy. Another challenge would be during weekends and holidays; during these periods, agencies rely heavily on CCTV footage for analysis. However, manually reviewing recordings is time-consuming and often delayed, preventing real-time traffic management.

Several studies on the application of artificial intelligence models for vehicle classification have been published in recent years. Wang et al. [3] compared state-of-the-art deep learning algorithms for vehicle detection at the time that were trained using KITTI dataset; however, the KITTI dataset [4] does not contain classe s characteristic of the Philippine setting. Moreover, robust AI model architectures have since been developed [5].

The variety of vehicles on Philippine roads—especially underrepresented types like public utility jeepneys (PUJs) and tricycles—adds complexity to traffic monitoring. Some studies have included these modes [6]–[8], but often use camera angles from sidewalks that capture more pedestrians than vehicles. Maclang et al. [9] developed an active learning-based system to classify vehicles and measure traffic flow, covering buses, cars, PUJs, tricycles, trucks, and vans. However, their dataset primarily features side-view images and generalizes across varying conditions (e.g., day/night, clear/cloudy).

This study proposes a framework for vehicle detection and counting tailored to the Philippine context, utilizing locally sourced data and incorporating vehicle classifications common to the region, namely, PUJs. The research leverages advanced data augmentation techniques and explores inno-

979-8-3315-6946-4/25/$31.00 ©2025 IEEE

vative data preprocessing and engineering methods to enhance model performance. Additionally, it presents a proof-of-concept and offers actionable recommendations for local government units on the implementation of AI-driven solutions for traffic management and optimization.

II. EXPERIMENTAL SET-UP

This section outlines the methodology employed in this study, detailing the framework used for data processing and augmentation techniques and AI model training.

A. Data Acquisition

(a) Day setting.

(b) Night setting.

Fig. 1: CCTV frames from the data gathered.

Quantitative data was obtained from CCTV footage provided by the Metropolitan Manila Development Authority (MMDA) Traffic Engineering Center. The dataset consists of video recordings from four major roads over a two-day period, spanning 6:00 AM to 8:00 PM daily. An additional recording was captured at Roces St., Diliman, Quezon City, Metro Manila. The footage includes both day and night conditions, offering a comprehensive view of typical and atypical traffic scenarios. Sample frames are presented in Figures 2a (day) and 2b (night).

Vehicle classes follow MMDA's classification, with class labels and instance counts detailed in Table I. Annotation was performed using the Computer Vision Annotation Tool (CVAT) for both the collected footage. An open-source dataset comprising nine vehicle classes [10] was added to the dataset. All frames were manually labeled to ensure consistency and quality.

B. Dataset Labeling

The image labeling process was carried out over five months using semi-automatic annotation in CVAT. In addition, open-source labels

Label	Vehicle Description
0	Car
1	Motorcycle
2	Public Utility Jeepney (PUJ)
3	Utility Vehicle (UV) express
4	Taxi
5	Standard Bus
6	Pick-up
7	Truck(2-Axles)
8	Truck Traile
9	Tanker
10	Ambulance
11	Fire Truck
12	Small Van
13	Bicycle
14	Motorized Tricycle
15	E-bike
16	Pedi Cab
17	Police Car
18	Other Vehicles

TABLE I: Vehicle labels

from [10] were adapted to match the annotation scheme of the manually collected dataset. Due to limited manpower, class validation was not performed. The resulting dataset exhibits a marked class imbalance, with Public Utility Jeepneys (PUJs)—a common vehicle type in the Philippines—significantly underrepresented compared to other vehicle classes [11]. The final dataset comprises 8,778 images captured during the day and 2,248 at night. Class imbalance was assessed using the Chi-square Goodness of Fit test, with p-values presented in Table II, based on a significance level of $\alpha = 0.05$.

Technique	Day	Night
Benchmark	<0.001	<0.001
Grayscale	<0.001	<0.001
Blocking	<0.001	<0.001
Group 1	0.003	0.353
Group 2	<0.001	<0.001
Group 3	<0.001	<0.001
Group 4	<0.001	<0.001
Group 5	<0.001	0.003

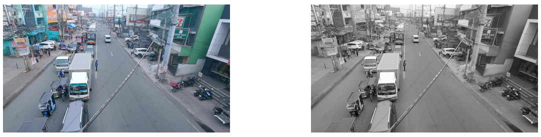
TABLE II: Dataset p-values for Chi-square Goodness of Fit Test at $\alpha = 0.05$

C. Data processing

The labeled data are statistically tested using Chi-Square Goodness of Fit test to determine if the distribution of frequencies match the expected frequencies based on a specified distribution [12]. Its

p-value gives insight on the significant difference of the instance counts of classes in the dataset. Dataset imbalance is mitigated by employing three data augmentation/processing techniques: Grayscaling, Blocking and Class-grouping. These are described in the succeeding subsections.

1) Grayscaling Technique: Color information is a significant feature in image-based object detection, particularly in tasks where distinguishing between object classes relies on differences in hue, saturation, or intensity [13]. However, in object detection tasks where variations in shape, texture, or structural features are more defining than color, incorporating color information may not provide substantial benefits. In such cases, models can be trained using grayscale images, reducing computational complexity while maintaining detection accuracy [14]. Converting images to grayscale not only simplifies the feature space but can also enhance robustness by minimizing the impact of lighting conditions and color variations that may introduce noise or inconsistencies in the training data. To improve performance metrics, this study investigates the impact of removing color information by training models on grayscale images [15]. This approach also aims to reduce headlight glare in night-time settings [15].



(a) RGB image. (b) Grayscale image.

Fig. 2: Side-by-side comparison of RGB and grayscale images

2) Blocking Technique: To address the issue of limited class representation, [16] introduced a ”cut, paste, and learn” method using the BigBird dataset to increase the number of instances. However, this approach simply pastes objects onto scene images, sometimes covering existing underrepresented instances. To improve this, [17] developed a technique that copies, transforms, and places instances in random locations within an image. While this method helps balance class distribution, adding similar instances just to balance out the class with least count may lead to overfitting, reducing the model’s ability to generalize effectively [17] .

In our study, we implement a strategic blocking technique that selectively removes non-jeepney instances to increase the number of example images containing jeepneys, while preserving the number of instances from other vehicle classes. An illustration of this approach is provided in Figure 3. The corresponding image label files were also updated to reflect the changes introduced by the augmentation process.

Fig. 3: Blocking technique for non-PUJ vehicle classes

3) Class Grouping Technique: Another data processing technique we explored is grouping the classes according to their instance counts, putting classes with instance counts close to each other into the same group. A total of 5 groups were identified for both the day and night settings. These groupings are shown in Table III.

Group	Vehicle Classes
1	Car and motorcycles
2	Pick-up, Bus, and PUJ
3	PUJ, and UV
4	Pick-up, Bus, and taxi
5	UV, truck trailer, bicycle, E-bike, and police car

TABLE III: Vehicle class groupings

D. Model development

YOLO works by dividing an image into a grid and predicting bounding boxes and class probabilities for objects directly from each grid cell in a single forward pass through a neural network. This allows for fast, real-time object detection by treating the entire detection process as a single regression problem. To cater deployment of the models, inference speed must be considered and YOLO11 achieved the best inference speed among YOLO models [18]. The researchers employed YOLO11n, a state-of-the-art AI model pre-trained on the COCO dataset [19]. Built on the PyTorch framework, YOLO11 represents a significant advancement in real-time object detection compared to its predecessors. Its key innovations include the Spatial Pyramid Pooling-Fast (SPPF) block, the C3k2 block, and C2PSA, all of which enhance feature extraction and object detection capabilities [5].

To adapt the model to the specific task, the researchers fine-tuned YOLO11n using a custom-labeled dataset. The dataset was split into 70-20-10 for train-val-test. The model was trained for 50 epochs with a batch size of 16, a learning rate of 0.01, and an intersection-over-union (IoU) threshold of 0.70. Training was conducted on an NVIDIA V100 GPU. In this study, precision and recall were used to evaluate the detection performance of the models. Additionally, mean Average Precision (mAP) at various thresholds—including mAP@50 and mAP@50–95—was employed to assess the models' overall classification performance.

III. Results and Discussion

In this section we discuss the results of the experiments and give insights on future research that can be done on vehicle detection with an imbalanced dataset.

A. Data Processing and Model Performance

A summary of the metrics and improvements of the data augmentation/processing techniques that were employed by the AI models are given in Table IV for day setting and Table V for night setting. These two tables present a comparative analysis of the performance metrics.

1) Grayscale: As shown in Tables IV and V, grayscale conversion can lead to notable gains in both speed and accuracy, primarily due to reduced computational complexity. Grayscale images use a single channel instead of three (RGB), lowering memory usage, bandwidth, and processing time [14]. When object recognition relies on shape and texture rather than color, grayscale inputs may enhance model generalization by reducing sensitivity to color variations introduced by lighting, camera sensors, or environmental conditions.

In this study, grayscale conversion improved precision and recall under daylight conditions (Table IV). However, its effectiveness diminished in low-light settings (Table V), suggesting that color information may be essential for object distinction and contrast at night. These findings highlight the context-dependent utility of grayscale processing, emphasizing its benefits in high-light environments while cautioning against its use in scenarios where color aids visual discrimination.

2) Blocking technique: Applying blocking technique led to a doubling of PUJ instances—a vehicle class commonly seen in the Philippines which, according to Table II still resulted to an imbalanced dataset. It was further used during the fine-tuning of the YOLO11n model. The results, shown in Tables IV and V , reflect performance in both day and night settings, respectively. The observed decline in performance may stem from overfitting, as the same instances were merely duplicated, a limitation also noted in [17], [20]; moreover, such can also be attributed to the persistent dataset imbalance used in training.

3) Class Grouping Technique: This technique evaluates the impact of grouping classes based on similar instance counts on AI model performance. Implementation involved retaining images while removing annotations for classes outside each group, encouraging the model to treat excluded classes as background noise. Class groupings are summarized in Table III.

In the day setting (Table IV), Group 1 yielded a p-value of 0.003, while Groups 2–5 all reported p-values below 0.001, indicating increasing imbalance. Correspondingly, Group 1 achieved the highest precision (67.531) and recall (61.757) (Table IV), suggesting that balanced class distribution contributed to improved model performance. Models trained on the more imbalanced groups showed degraded precision and recall, likely due to bias toward majority classes and reduced detection accuracy for minority classes [17].

For the night setting (Table V), Group 1 and Group 5 exhibited no significant deviation from expected distributions, indicating balance, whereas Groups 2–4 had p-values below 0.001. The model trained on Group 1 again outperformed others, with precision of 54.5 and recall of 47.1, demonstrating better generalization and reduced overfit-

Technique	Precision		Recall		mAP50		mAP50-95	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Benchmark	14.986	—	16.317	—	10.498	—	5.3189	—
Grayscale	52.00	247%	41.6	155%	45.10	329.61%	34.30	544.87%
Blocking	13.942	-7.00%	16.461	1.23%	10.149	-3.32%	5.196	-2.31%
Grouping 1	67.531	351%	61.757	279%	56.10	434.39%	91.70	1624.04%
Grouping 2	54.4	263%	43.7	168%	40.5	285.79%	26.7	401.98%
Grouping 3	31.0	107%	33.3	104%	23.8	126.71%	10.7	101.17%
Grouping 4	19.3	28.8%	36.1	121%	23.8	126.71%	19.0	257.22%
Grouping 5	11.4	-24.0%	9.07	-44.3%	5.85	-44.28%	3.16	-40.59%

TABLE IV: Performance of Data Preprocessing and Augmentation for the Day Setting.
Note: % Change refers to improvement over the benchmark.

Technique	Precision		Recall		mAP50		mAP50-95	
	Value	% Change	Value	% Change	Value	% Change	Value	% Change
Benchmark	52.0	—	17.8	—	11.2	—	5.70-	—-
Grayscale	21.7	2.36%	15.3	-14.0%	10.70	-4.46%	5.34	-6.32%
Blocking	13.942	-34.4%	16.461	-7.30%	10.149	-9.38%	5.196	-8.42%
Grouping 1	54.5	157%	47.1	164%	43.8	-4.46%	17.5	207.02%
Grouping 2	2.57	-87.9%	16.8	-5.62%	3.09	-72.41%	1.40	-75.44%
Grouping 3	21.2	0%	17.8	0%	11.2	0%	5.70	0%
Grouping 4	13.9	-34%	20.8	16.9%	7.59	-32.23%	3.91	-31.40%
Grouping 5	11.4	-46.2%	9.07	-49.0%	4.60	-58.93%	2.39	-58.07%

TABLE V: Performance of Data Preprocessing and Augmentation for the Night Setting.
Note: % Change refers to improvement over the benchmark.

ting. In contrast, models trained on imbalanced groups showed poorer performance, reaffirming the negative impact of skewed class distributions.

The findings underscore the importance of class distribution in AI model performance. Grouping classes with similar instance counts enhances model generalization by promoting a more balanced training dataset. The superior performance of models trained on Group 1 in both day and night settings indicates that maintaining relatively equal class representation reduces bias and improves recall.This highlights the effectiveness of data-centric optimization, which serves as a powerful alternative to model-centric approaches that modify the network architecture with components like attention modules to boost performance [21].

IV. Conclusion and Recommendations

This study proposed a framework for vehicle counting and detection in the Philippine context by applying data processing and augmentation techniques. The findings showed that these methods significantly impacted data quality, which, in turn, affected AI model performance. Notably, the grayscaling technique improved night-time detection, increasing precision and recall by 247% and 155%, respectively. Similarly, grouping cars and motorcycles significantly enhanced detection in both day and night settings, with precision and recall improvements of 351% and 279% during the day, and 157% and 164% at night.

However, not all techniques yielded positive outcomes. For instance, grayscaling did not consistently improve performance across all settings, and grouping techniques applied to other vehicle subsets were less effective. These results highlight the critical role of class distribution and dataset quality in shaping model learning and effectiveness.

V. Future Work

The data collection campaign will be extended to six months, covering at least ten arterial and secondary roads across Metro Manila and two provincial cities. Each site will be sampled over 24-hour periods under dry, rainy, and holiday conditions, yielding ≥ 30 hours of footage per location. Dataset splits will exclude entire roads and days to avoid temporal leakage. At least 2,000 annotated instances per minority class (PUJ, tricycle, bicycle, e-bike) will be ensured, with diffusion-based augmentation used as needed. Protocols, camera specifications, annotation guidelines, and a 10% public subset will be released.

Acknowledgment

This study was funded and monitored by the Department of Science and Technology (DOST) – Philippine Council for Industry, Energy, and Emerging Technology Research and Development (PCIEERD) under Project No. 1213385. We extend our sincere gratitude to ASTI-VIROS and the Metropolitan Manila Development Authority (MMDA) for providing the datasets used in this research.

We would also like to give special thanks to Aunhel Adoptante, Gary Chris Lacdang, Samuel Harrison Cerrudo, and Kent Roger Truita for their valuable insights and contributions to this study.

Our appreciation also goes to the student interns who assisted in data labeling: Danielle Verna J. Caasi, Eden V. Dorato, Jared Ralexander Manansala, Kyle Dominic Albano, Julian Nite Yu, Ryan Joseph J. Tingco, Bien Liam Cromwell S. Caño, and Nevin Kenneth B. dela Paz.

References

[1] F. Ostermeijer, H. R. A. Koster, J. van Ommeren, and V. M. Nielsen, "Automobiles and urban density," *Journal of Economic Geography*, vol. 22, no. 5, pp. 1073–1095, 02 2022. [Online]. Available: https://doi.org/10.1093/jeg/lbab047

[2] H. S. Lidasan, I. C. Espada, and M. R. C. D. Leon, "A needs assessment of transport planning and traffic management of local cities: The case of the philippines," *Proceedings of the Eastern Asia Society for Transportation Studies*, vol. 2009, pp. 3–3, 2009.

[3] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Q. Liu, "A comparative study of state-of-the-art deep learning algorithms for vehicle detection," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 2, pp. 82–95, 2019.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[5] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.

[6] EAN, "mayondon roa dataset," https://universe.roboflow.com/ean/mayondon-roa , nov 2023, visited on 2025-01-23. [Online]. Available: https://universe.roboflow.com/ean/mayondon-roa

[7] M. 7K, "Vehicle detection dataset," https://universe.roboflow.com/mayondon-7k/vehicle-detection-gakkc , nov 2023, visited on 2025-01-23. [Online]. Available: https://universe.roboflow.com/mayondon-7k/vehicle-detection-gakkc

[8] N. Crasto, "Class imbalance in object detection: An experimental diagnosis and study of mitigation strategies," *arXiv preprint arXiv:2403.07113*, 2024.

[9] A. R. Maclang, M. L. Orante, R. D. Salvador, D. J. Del Carmen, and R. D. Cajote, "Video dataset labeling using active learning with applications in vehicle classification and traffic flow rate measurement," in *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, 2023, pp. 936–941.

[10] LogicLaboratories, "Final na i2 dataset," https://universe.roboflow.com/logiclaboratories-stxe9/final-na-i2-hilo0 , nov 2024, visited on 2025-01-23. [Online]. Available: https://universe.roboflow.com/logiclaboratories-stxe9/final-na-i2-hilo0

[11] T. R. Hoens and N. V. Chawla, "Imbalanced datasets: from sampling to classifiers," *Imbalanced learning: Foundations, algorithms, and applications*, pp. 43–59, 2013.

[12] P. Newbold, W. L. Carlson, and B. M. Thorne, *Statistics for business and economics*. Pearson, 2013.

[13] L. Bo, X. Ren, and D. Fox, "Unsupervised feature learning for rgb-d based object recognition," in *Experimental Robotics: The 13th International Symposium on Experimental Robotics*. Springer, 2013, pp. 387–402.

[14] Y. Ge, Q. Zhang, Y. Sun, Y. Shen, and X. Wang, "Grayscale medical image segmentation method based on 2d&3d object detection with deep learning," *BMC Medical Imaging*, vol. 22, no. 1, p. 33, 2022.

[15] M. Powelleit, S. Winkler, and M. Vollrath, "Cooperation through communication–using headlight technologies to improve traffic climate," *Proceedings of the Human Factors and Ergonomics Society Europe*, pp. 149–160, 2018.

[16] D. Dwibedi, I. Misra, and M. Hebert, "Cut, paste and learn: Surprisingly easy synthesis for instance detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1301–1310.

[17] K. Lee, S. Lee, and H. Y. Kim, "Bounding-box object augmentation with random transformations for automated defect detection in residential building façades," *Automation in Construction*, vol. 135, p. 104138, 2022.

[18] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, and M. Karkee, "Comprehensive performance evaluation of yolo11, yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments," *arXiv preprint arXiv:2407.12040*, 2024.

[19] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[20] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "Bigbird: A large-scale 3d database of object instances," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 509–516.

[21] J. Zhao, S. Hao, C. Dai, H. Zhang, L. Zhao, Z. Ji, and I. Ganchev, "Improved vision-based vehicle detection and classification by optimized yolov4," *IEEE Access*, vol. 10, pp. 8590–8603, 2022.