The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-5/W4-2025
Philippine Geomatics Symposium (PhilGEOS) 2025 "Enhancing Human Quality of Life through Geospatial Technologies",
24–25 November 2025, Quezon City, Philippines

Urban Mobility Insights from CCTV: A Deep Learning Approach to Traffic Flow Monitoring

Mario G. Ugalino Jr.[1], Albert Francis P. Florin[1], Ma. Bea Angela I. Zamora[1], Laurelly Joyce A. Aporto[1], Leonardo Miguel Garcia[1], Pia Franchesca R. Maralit[1], Kim Elijah M. Aguilan[1], Jarence David D. Casisirano[1], Karlo Mark C. Tablang[1], Fatima Joy O. Pamittan[1], Dominic C. Fargas Jr. [1], Czar Jakiri S. Sarmiento[1]

[1] UP Training Center for Applied Geodesy and Photogrammetry, University of the Philippines Diliman, Quezon City, Philippines – {mgugalino1, apflorin1, mizamora, laaporto1, lgarcia1, prmaralit, kmaguilan, jdcasisirano, kctablang, fopamittan, dcfargas, cssarmiento}@up.edu.ph

Keywords: Traffic Flow, Computer Vision, Closed-Circuit Television (CCTV), YOLO, ByteTrack.

Abstract

This study explores the feasibility of using closed-circuit television (CCTV) data and computer vision methods, particularly YOLO (You Only Look Once) and ByteTrack, for traffic flow monitoring in Pavia, Iloilo. Two 24-hour videos from opposing lanes of a major road section were processed using YOLOv8 for vehicle detection and ByteTrack for multi-object tracking, using image rectification through homography and known vehicle measurements to estimate vehicle counts, speed, volume, and density. The models achieved high performance metric scores, with mAP50 values of 0.91 (GT Mall) and 0.89 (Robinsons). Results of the traffic flow analysis showed expected patterns: (a) vehicle speeds decreased as traffic volume and density increased, and (b) peak volumes and densities occurred during commuting hours. The interaction with and effects of between illumination and occlusion were considered during data preparation, resulting in mean absolute error rates below 3% compared with manual vehicle counts. By generating per-frame logs of traffic flow, the study shows the potential of computer vision–based monitoring systems serving as a supplement for evidence-based policymaking and urban planning for congestion management, road safety, and infrastructure planning at a local scale.

1. Introduction

1.1 Overview

Transportation and mobility are essential to sustainable and functional urban areas. They influence access to services and urban metabolism by enabling the flow of resources, energy, and waste. Effective transportation systems contribute to sustainability by reducing environmental impact, improving quality of life, and fostering economic growth through better connectivity (Fróes & Lasthein, 2020).

Among sixteen industries in Iloilo Province, located in the Western Visayas region of the Philippines, transportation and storage ranked third in terms of growth rate at 22.8 percent (PSA, 2023). The growth in this field highlights the relevance of developing tools and methods to monitor traffic and mobility patterns, particularly in Iloilo's rapidly urbanizing municipalities. Pavia, located adjacent to Iloilo City, has become one of the fastest-developing towns in the province, with its strategic location attracting residential, commercial, and industrial growth, supported by a complex network of surveillance cameras. Thus, the study proposes the utilization of closed-circuit television (CCTV) data to create a vehicle detection and tracker using YOLOv8 and ByteTrack along a select road section in Pavia.

1.2 Significance

The field of transportation and mobility is significant to the 2030 Agenda for Sustainable Development (SDG), as they are directly tied to how cities function and how people access opportunities. By emphasizing the topic of sustainable transport, the study links to Sustainable Cities and Communities (SDG 11) and Climate Action (SDG 13). It highlights the need for building sustainable transportation in the form of infrastructure, transit systems, freight and delivery networks. The study also highlights the importance of affordability, efficiency, and convenience of transportation as a whole, along with enhancing urban air quality and public health while lowering greenhouse gas emissions. Monitoring traffic and mobility patterns in municipalities such as Pavia helps policymakers develop urban planning strategies that reduce congestion, improve road safety, and enhance accessibility.

The study also contributes to Decent Work and Economic Growth (SDG 8). As Iloilo province's growth rate in transportation and mobility rises, so does the need towards improving reliable and efficient transport. Doing so supports productivity, logistics, and connectivity across industries. The focus on using CCTV data with YOLOv8 and ByteTrack also supports Industry, Innovation, and Infrastructure (SDG 9), as the study applies advanced technology to strengthen transport systems and create data-driven solutions for infrastructure development and policy-crafting.

1.3 Objectives

The study aims to develop a foundation for a traffic flow monitoring system that logs key traffic characteristics, particularly vehicle speed, volume, and density. The data can support policy-making by local government units and infrastructure planners.

Specific objectives of the paper include:

1. Assessment of using YOLO and ByteTrack for vehicle detection and tracking in Pavia;
2. Identification of temporal patterns in the obtained traffic speed, volume, and density across the study area; and
3. Validate detected traffic flow patterns with traffic flow theory to assess its potential for planning and policy-making.

2. Materials and Methods

2.1 Study Area and Data Collection

A specific section of Pavia's Old Iloilo-Capiz Road, locally known as Benigno Aquino Sr. Avenue, particularly the stretch across both GT Town Center and Robinsons Place were selected as the study sites, as seen in Figure 1. These two (2) sections are: the southbound lane towards Iloilo City and the northbound lane away towards the provincial airport. The sites were chosen due to the road being a major thoroughfare not only for private but also for public vehicles, connecting the province's international airport to multiple municipalities and cities within southern Iloilo.

Figure 1. GT Town Center Pavia (green) and Robinsons Place Pavia (pink).

To analyze traffic flow trends and cumulative statistics, the Disaster Risk Reduction and Management (DRRM) Office of Pavia provided 24 hour-long 3840x2160 (4K) MP4 videos, one for each study area, spanning from 12:00 NN of March 1, 2025, to 12:00 NN of March 2, 2025. A total of four regions of interest were set and used in the study as see in Figures 2 and 3. GT Mall contains one (1) carriageway only and Robinsons has three (3).

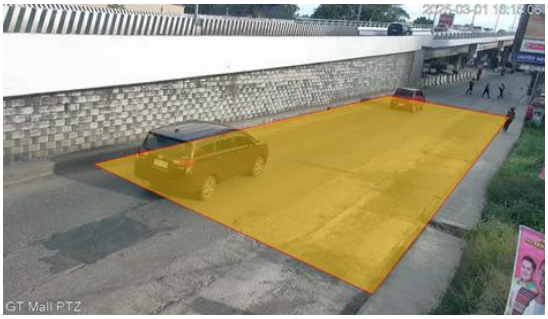

Figure 2. Region of Interest for GT Mall CCTV.

Figure 3. Regions of interest at Robinsons, showing the service road (left), the main road (middle), and the flyover (right).

2.2 Data and Preprocessing

The raw CCTV video data were first segmented into frames using FFmpeg, following an interval of one (1) frame per five (5) seconds. This resulted in 17,258 frames for GT Mall and 17,300 frames for Robinsons. The frames were then uploaded to a hosted server through the Computer Vision Annotation Tool (CVAT) platform, wherein frames were randomly sampled and annotated based on the type of vehicle. Annotations included bounding boxes to mark vehicle positions, with differentiation between the type of vehicle through color-coded bounding boxes. To ensure consistency, the annotation process followed a standardized labeling scheme where frames are evenly split into day and night periods. Vehicle annotations were also divided into six distinct classes: cars, motorcycles, tricycles, jeepneys, buses, and trucks.

Filtering and dataset splits were done by splitting annotations, where the model's training data accounted for 80 percent of the total splits and 20 percent for validation data for both GT Mall and Robinsons models. Table 1 shows the summary of training and validation splits for both GT Mall and Robinsons.

Camera	Training	Validation
GT Mall	3018	755
Robinsons	3092	744

Table 1. Annotation Splits

The annotated frames were then uploaded from CVAT into the Ultralytics YOLO Detection 1.0 format, serving as the basis for model training and validation. Figure 4 below shows the complete workflow of the study from data pre-processing, computer vision modelling, and data validation.

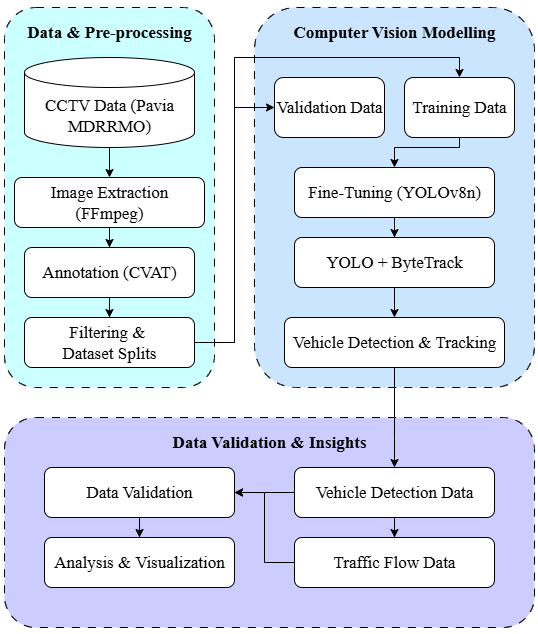

Figure 4. Workflow of the Study.

2.3 YOLOv8 with ByteTrack through Supervision

Two 24-hour videos were collected from fixed-position PTZ cameras, with footage selected to minimize camera movement. YOLOv8 nano (Ultralytics, 2025) was employed for vehicle detection, and ByteTrack (Zhang et al., 2022) via the Supervision library, was used for multi-object tracking. Prior to any measurement, a one-time perspective transformation was applied

to each video using a four-point homography derived from a reference frame (OpenCV, n.d.). This transformation rectified the view to a top-down, two-dimensional layout where the road surface appeared flat, ensuring that pixel distances remained consistent across the scene regardless of depth.

YOLOv8 nano was selected as the detection model due to its low computational overhead and fast inference speed, making it suitable for edge deployment scenarios (Ultralytics, 2025). Although it has a smaller architecture and reduced accuracy compared to larger variants, its performance was sufficient after fine tuning a separate model for every camera. This lightweight model fit a scaling strategy with two deployment options: (1) a decentralized approach, where each camera runs independently on a low-power microcomputer but comes with high maintenance demands; or (2) a centralized server-based setup, where heavier hardware handles processing for multiple video streams. The current pipeline is designed to remain open to either path.

To convert pixel displacements into real-world distances, pixel-to-meter scale factors per region of interest were derived using wheelbase measurements from known vehicle specifications. For each camera view, representative vehicle samples were manually selected over a 30-minute segment, and their apparent wheelbases in the rectified view were averaged across multiple types and lanes. This calibration allowed tracking of vehicle positions in world coordinates and estimate longitudinal displacements in meters.

2.4 Traffic Flow Theory

In traffic flow theory, one of its relationships is expressed in Equation (1) (Maerivoet and De Moor, 2008) as

$$q = k \times \overline{v_s} \tag{1}$$

where q denotes flow (vehicles per unit time)
k represents density (vehicles per unit distance); and
$\overline{v_s}$ is the average speed of the traffic stream.

This relationship provides a link between how many vehicles occupy a roadway segment and how fast they are moving, serving as a basis for the fundamental relationships for traffic flow, such as flow–density, speed–density, and speed–flow. In this study, the following variables were used as a reference for validating the direct and inverse proportionality between the data observed by the computer vision models.

Speed estimation was performed by measuring the distance each vehicle traveled in the top-down view over a fixed time interval, using the historical world-coordinate positions of tracked vehicles maintained by ByteTrack. For each vehicle, the displacement along the road axis was divided by the time elapsed and multiplied by the computed pixel-to-meter scale factors to obtain speed in kilometers per hour. To ensure reliability, only vehicles that remained in the region of interest for at least one second were included in the calculation.

Vehicle volume was defined as the total number of uniquely identified vehicles passing through the region of interest, normalized by total elapsed time (vehicles per second). **Density**, on the other hand, was computed at fixed intervals by counting the number of vehicles within the predefined zone and expressing this as vehicles per 100 meters of roadway. All traffic flow variables—including timestamps, class labels, bounding boxes, and computed speeds—were logged per frame, while per-category counts were aggregated per video to characterize temporal variations in traffic behavior across the 24-hour period.

2.5 Metrics for Model Accuracy

2.5.1 Intersection over Union (IoU): Intersection over Union (IoU) is a metric used to evaluate the overlap between predicted bounding box and its corresponding ground-truth bounding box. It is directly used in performance metrics such as confusion matrices and mean Average Precision (mAP), as it measures the accuracy of predictions and the quality of localization or how well the model places the bounding box on the detected object. An IoU value of 1 indicates a perfect overlap, while a value of 0 indicates no overlap. Figure 5 shows the measurement for Intersection over Union.

$$IoU = \frac{|B_{pred} \cap B_{gt}|}{|B_{pred} \cup B_{gt}|} = \frac{\text{predicted}}{\text{ground truth}}$$

Figure 5. Intersection Over Union

IoU is not solely a performance evaluator as it is also integrated into the loss functions of models such as YOLOv8 and YOLO11. It helps the models learn to predict bounding boxes that account for overlap, center distance and aspect ratio alignment.

2.5.2 Mean Average Precision (mAP): Mean Average Precision (mAP) is used as an overall measure of model performance across all objects and classes in a dataset. It is essentially calculated by averaging precision values across multiple IoU thresholds and all classes. The following metric is useful in multi-class object detection scenarios to provide a comprehensive evaluation of the model's performance (Ultralytics, 2025).

The mean Average Precision is defined as:

$$mAP = \frac{1}{C}\sum_{c=1}^{C} AP_c, \tag{2}$$

where C is the number of classes, and
APc is the average precision for class c.

2.5.3 Mean Absolute Error (MAE): Vehicle counts were manually validated to evaluate the accuracy of traffic flow measurements. Data collected on vehicle counts were split into five-minute segments and selected based on highest vehicle count across day and night periods. The number of vehicles that appeared within the defined region of interest were manually counted and considered as ground-truth data. These ground-truth counts were compared against the system-generated vehicle counts using Mean Absolute Error (MAE), a standard metric in vehicle counting and density estimation studies (Hu et al., 2021; Hu et al., 2022).

The absolute error was computed using Equation (3):

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|D_{n,\text{auto}} - D_{n,\text{manual}}\right| \quad (3)$$

where $D_{n,\text{auto}}$ is the automated estimate,
$D_{n,\text{manual}}$ is the manual count, and
N is the number of sampled intervals.

This methodology allows quantifying how closely the system approximates real vehicle counts and provides a basis for comparing performance across conditions such as lighting and congestion.

2.6 Error Analysis under Challenging Conditions

In addition to the metrics for model accuracy, the system's performance was assessed across three common sources of error: (1) low lighting; (2) occlusion; and (3) vehicle overlap.

Low lighting (e.g., nighttime or early morning) reduces contrast and introduces image noise lowering object detection performance and accuracy especially for smaller objects (Rodríguez-Rodríguez et al., 2024). YOLO-based detection models are known to degrade in performance under such conditions without finetuning.

Vehicle occlusion and overlap, particularly in congested scenes or multi-lane areas (e.g., Robinsons Ungka), may lead to missed detections, incorrect counting, or object ID switching during tracking. These issues can skew both volume and density metrics. While CVAT allows labeling of occluded properties such as in Figure 6, Ultralytics YOLO Detection 1.0 annotations inherently ignores occlusion as exported datasets do not support attributes and are only divided into images and *.txt files listing each object's class, center coordinates, and bounding box dimensions (CVAT, 2025).

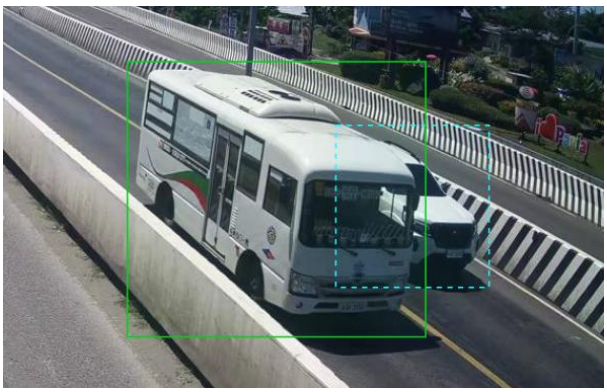

Figure 6. Annotation boxes showing occluded objects (dashed lines) and non-occluded objects (solid lines).

Absence of lane markings also contributes indirectly to calibration uncertainty, since real-world distance estimation relies on fixed visual references on the road plane.

Validation samples were stratified by day and night periods and error patterns in density output were reviewed to understand the impact of these conditions. This does not fully quantify the compounded effect of these challenges. However, it provides practical insights into the expected variation in performance under non-ideal conditions.

3. Results and Discussion

3.1 YOLOv8 Detection & Classification

Confusion matrices with vehicle types as classes were generated to assess the performance of each model. The results for GT Mall in Figure 7. and Robinsons in Figure 8. show that vehicles can be classified accurately, where the true positive ranges of each category ranges from 0.72 to 0.91. However, it is also observed that false positives occur most when background regions are misclassified as cars and motorcycles for both models. The following result can be mainly attributed to the default IoU value set at a half overlap or 0.50, where predictions that fall below this threshold are counted as false positives even if they are close to the ground truth object.

This behavior is particularly noticeable for cars and motorcycles because of their high number of annotations and detected instances across the datasets. The two categories' high frequency directly affects the model and generates more bounding box predictions, increasing the chances of localization and overlap issues compared to less common classes such as buses or tricycles. As a result, their higher annotation frequency affects the evaluated metrics under stricter IoU criteria.

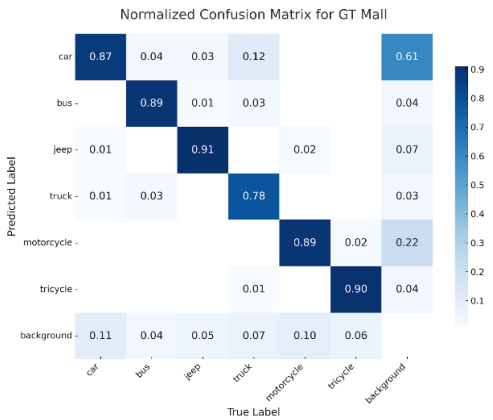

Figure 7. Normalized confusion matrix for GT Mall model.

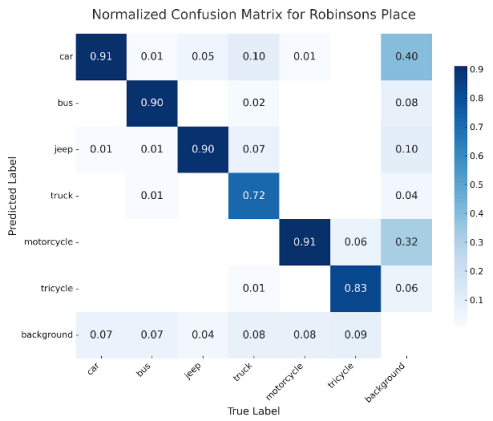

Figure 8. Normalized confusion matrix for Robinsons model.

A total of one hundred (100) learning iterations through the entire dataset, also known as an epoch, was set during the training of models for GT Mall and Robinsons. Figure 9 and Figure 10 show the training and validation performance of the model for GT Mall and Robinsons. The first row contains the trend of training metrics, with the first three graphs pertaining to localization

(box_loss and dfl_loss) and classification (cls_loss) losses and the other two graphs containing precision and recall, which measures the correctness of predictions made and measures how many of the actual objects are detected, respectively. The second row contains the trend of validation metrics, with the first three graphs similar to the first row, pertaining to localization and classification losses from the validation dataset. The last two graphs, mAP50 and mAP50-95, measure the models' mean Average Precision at specific IoU thresholds, with mAP50 at 0.50 IoU and mAP50-95 at IoU thresholds 0.5 to 0.95.

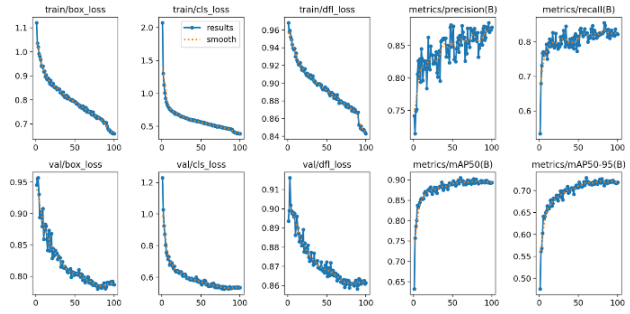

Figure 9. Training and validation performance of the GT Mall model over 100 epochs.

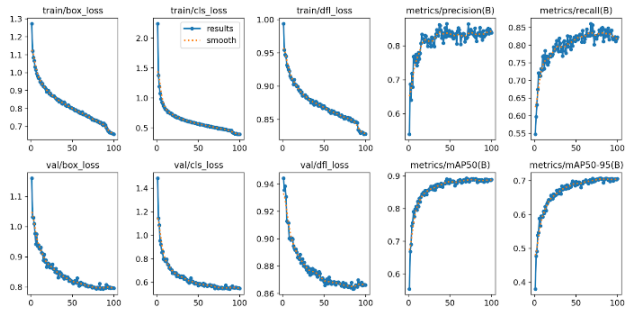

Figure 10. Training and validation performance of the Robinsons model over 100 epochs.

Tables 2 and 3 show that both the GT Mall and Robinsons model demonstrate steadily decreasing and converging training and validation losses per epoch, indicating that the model learned from the training data without overfitting.

GT Mall metrics	Initial value	Best value
Precision	0.74	0.89
Recall	0.53	0.86
mAP50	0.63	0.91
mAP50-95	0.48	0.73

Table 2. GT Mall Performance Metrics

In the context of vehicle detection for traffic monitoring and interpreting YOLOv8 results, the values of performance metrics indicate positive results as both models exhibit high peak mAP50 values of 0.91 and 0.89 with no significant drop-off in their respective mAP05-95 values. Relative to Ultralytics' benchmarks on Microsoft's Common Objects in Context (COCO), a large-scale object detection, segmentation, and captioning dataset, the finetuned models performed exceptionally well, with the benchmark evaluations for COCO's mAP50-95 under YOLOv8n equal to 0.37 compared to GT Mall and Robinsons' 0.71 and 0.73.

Robinsons metrics	Initial value	Best value
Precision	0.54	0.87
Recall	0.55	0.86
mAP50	0.55	0.89
mAP50-95	0.38	0.71

Table 3. Robinsons Performance Metrics

This indicates that the domain-specific traffic datasets used in the study are contextualized to a local scale, as opposed to COCO's highly diverse dataset, allowing the models to achieve superior performance under stricter evaluation criteria. The study's models also used only six vehicle types as object classes and consistent camera angles. These factors contributed to higher localization accuracy and in turn, mAP50-95 values that outperform general-purpose benchmarks such as COCO. Results for mAP50 show that the GT Mall and Robinsons model was performant in detecting and classifying vehicles accurately, while mAP50-95 is to bounding box placements under stricter localization parameters due to increasing IoU thresholds.

The models' MAE of object detection through observed manual vehicle counts as seen on Table 4 revealed positive results across overall, day, and night periods. The GT Mall model tallied errors under 1%, with a small difference between day and night periods, while the Robinsons model showed a higher rate of error and greater deviation, ranging from 2% to 3.6% across day and night periods. This suggests that the GT Mall model maintained consistent detection accuracy regardless of lighting conditions. On the other hand, the results of the Robinsons model show that its model is more sensitive to the CCTV's field of view and lighting conditions, with the camera directly facing the vehicles and its headlights during nighttime. Overall, both models demonstrate reliability and a strong potential for integration into automated traffic monitoring systems, particularly when camera placement and lighting conditions are optimized.

Category	Day	Night	Total
GT Mall	0.56%	0.89%	0.72%
Robinsons	2.01%	3.55%	2.78%
Overall	1.28%	2.22%	1.75%

Table 4. Mean Absolute Error for GT Mall and Robinsons

3.2 Traffic Flow Characteristics

Using the best model weights generated from training, per-frame logs of detected vehicles from GT Mall and Robinsons used to generate per-vehicle logs and aggregated traffic flow statistics. Figure 11 to Figure 14 show the plots of vehicle counts aggregated over hourly intervals for each region of interest, wherein GT Mall and Robinsons' main and service road were dominated by motorcycles and cars while the flyover was mostly utilized by cars for the entire dataset period. Usage rate for all roads except for the service road had similar hourly peaks while the service road was mostly utilized by motorcycles during various times of day, reaching up to 220 vehicles in an hour.

Overall, the plots reveal that peak vehicle usage occurred at late afternoon to early evening (4-7 PM) of March 1 and early morning (6 AM) of March 2, which is in line with common commuting hours. The high motorcycle and car counts throughout the day indicate high private vehicle usage, making up 90% of peak hour traffic. Meanwhile, jeepneys and buses show limited activity at midnight, reflecting expected public transport operations. Both trucks and tricycles maintained a low but steady presence, providing transport and freight services.

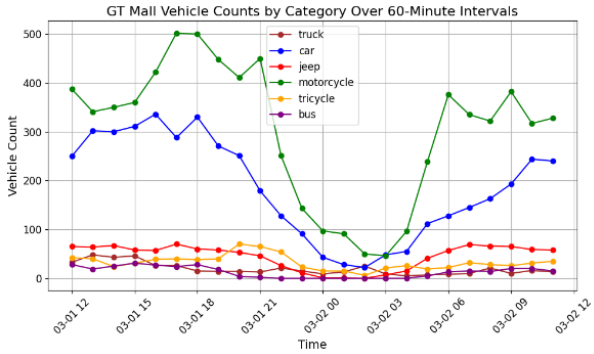

Figure 11. GT Mall Vehicle Count Over 60-minute Intervals.

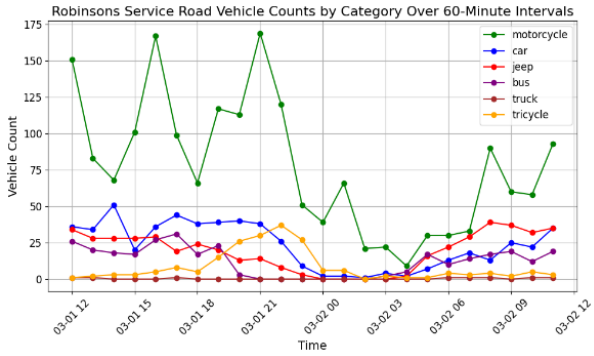

Figure 12. Robinsons Service Road Vehicle Count Over 60-minute Intervals.

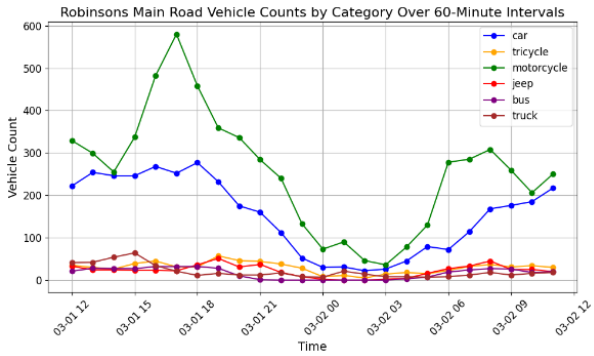

Figure 13. Robinsons Main Road Vehicle Count Over 60-minute Intervals.

Figure 15 presents the results of ByteTrack's vehicle tracking and speed estimation. Figures 16 to 19 display hourly traffic flow statistics for the regions of interest, showing an inverse relationship between mean speed to its volume and density. The measurements for volume and density for all roads have consistently peaked around mid-afternoon to early-evening (4-7 PM), with its mean speed also reaching its relative lows during the same period. Mean speeds for all roads all peaked around midnight to early morning (12 MN-6 AM), when vehicle volume and density are at its lowest, with peak-hour speeds averaging 25 kilometers per hour for GT Mall, 15 kilometers per hour for Robinsons' main and service roads, and 40 kilometers per hour for the flyover.

The traffic flow characteristics indicate that ByteTrack successfully estimated vehicle speeds and counts since the observed patterns align with concepts under traffic flow theory, wherein higher volumes and densities are indirectly proportional to speed. This consistency between expected traffic behavior and the measured outputs suggests the system can provide reliable inputs for congestion management in urban Iloilo.

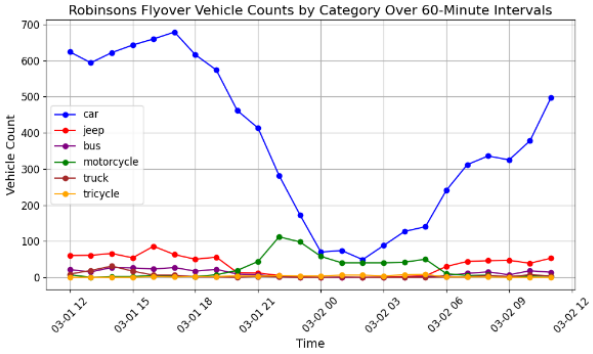

Figure 14. Robinsons Flyover Vehicle Count Over 60-minute Intervals.

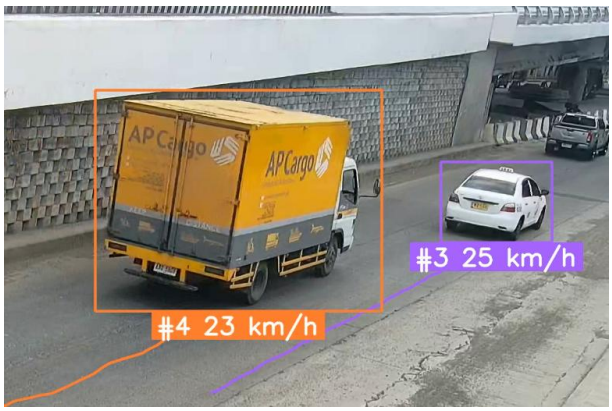

Figure 15. Vehicle ID assignment and speed estimation using ByteTrack.

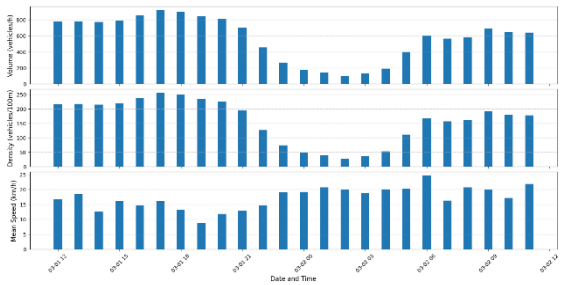

Figure 16. Hourly Traffic Flow Statistics for GT Mall. (Full quality: https://tinyurl.com/4p7dydup)

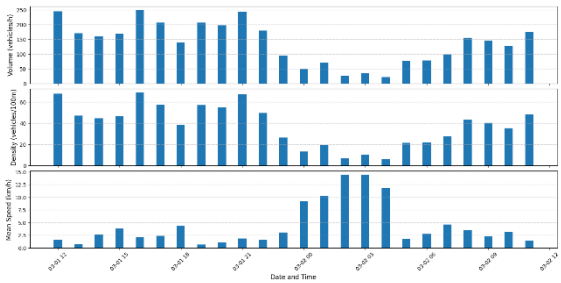

Figure 17. Hourly Traffic Flow Statistics for Robinsons Service Road. (Full quality: https://tinyurl.com/56ybm9be)

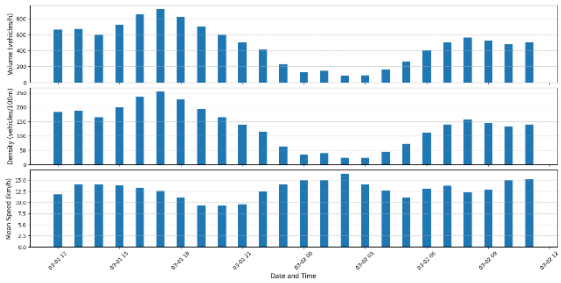

Figure 18. Hourly Traffic Flow Statistics for Robinsons Main Road. (Full quality: https://tinyurl.com/4w4mxsju)

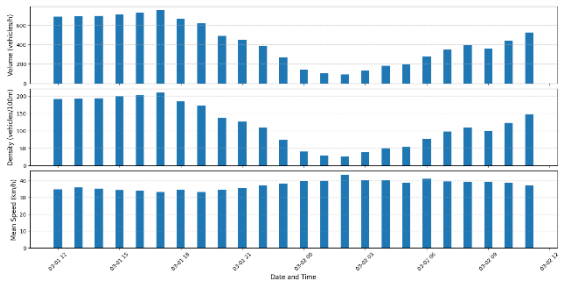

Figure 19. Hourly Traffic Flow Statistics for Robinsons Flyover. (Full quality: https://tinyurl.com/3dcath6r)

4. Conclusions and Recommendation

The study evaluated the feasibility of using YOLO-based computer vision for traffic monitoring in Pavia, Iloilo. Key results were obtained from the performance of YOLOv8 models, tracking of ByteTrack, and the derived traffic flow characteristics of vehicle counts, speed, volume, and density across the study sites. GT Mall and Robinsons models achieved high performance metrics, with mAP50 values of 0.91 and 0.89, respectively. Mean Absolute Errors relative to manual vehicle counts were low for both models, while the Robinsons site showed slightly higher error rates due to lighting and camera orientation. Overall, YOLOv8 with ByteTrack proved effective for vehicle detection and multi-object tracking in CCTV-based monitoring in the case of Pavia's Benigno Aquino Sr. Avenue.

Hourly trends in speed, volume, and density across both study sites reflected normal commuting patterns. Traffic volume and density peaked during the early morning (6–9 AM) and the late afternoon to early evening (4–7 PM). On the other hand, mean speeds reached their peak past midnight when volume and density were lowest, revealing that the patterns observed by the models are in line with established theory of traffic flow. While still in its initial stages, the system's ability to generate continuous logs of vehicle volume, density, and speed provides a foundation for local policymakers. The identified patterns such as traffic flow characteristics, daytime and nighttime vehicle activities, and peak and off-peak hours could provide insights for urban infrastructure planning, including traffic management strategies, road improvement scheduling, and commuter advisories for public transport planning.

Further improvements and recommendations for system development include adjustments in training criteria, CCTV orientation, and focus on data analysis.

In terms of training criteria, the choice of IoU threshold directly impacts the performance metrics of the model being trained. While the default IoU value of 0.5 offers a balance in terms of evaluation, further analysis on the specific needs and accuracy of vehicle detection can be studied to train better models that fit the specific needs in traffic monitoring.

Optimizing the orientation and placement of sensors for data collection such as CCTVs also serve a vital role in model performance, as observed in the limitation imposed by direct view of cameras to vehicle headlights in Robinsons. This includes validation of pixel-to-meter conversions to actual site measurements and additional documents that could support observed data such as known speeds in specific roads and highways.

Lastly, further analysis on the dynamics of transportation, particularly traffic congestion modelling, road capacity, level of service (LOS), private vehicle usage, public transit, and freight scheduling using the data collected from the study can further help policymakers gain insights on policies that could enact positive change within the transportation and mobility field on a local scale.

In conclusion, while further refinements to the methodology are possible, a traffic monitoring system via computer vison is viable to enact in a local government level.

Acknowledgments

This research was done as part of the Modern Geospatial and Collaborative Solutions for the Development of Smart Regions (SMART METRO) Project. The Project was implemented by the University of the Philippines Training Center for Applied Geodesy and Photogrammetry (TCAGP), through the support of the Department of Science and Technology (DOST) of the Republic of the Philippines and the Philippine Council for Industry, Energy, and Emerging Technology Research and Development (PCIEERD), under Project No. 1212042. The authors would also like to acknowledge the Municipality of Pavia, particularly the Municipal Disaster Risk Reduction and Management Office (MDRRMO), for providing the datasets used in this study.

References

CVAT.ai, n.d. Ultralytics YOLO: How to export and import data in Ultralytics YOLO formats. docs.cvat.ai (12 August 2025).

Fróes, I., Lasthein, M.K., 2020. Co-creating sustainable urban metabolism towards healthier cities. *Urban Transformations,* 2(1), Article 5. doi.org/10.1186/s42854-020-00009-7.

Hu, Y.X., Sun, Q., Jia, R.S., Li, Y.C., Liu, Y.B., Sun, H.-M., 2022. Le-SKT: Lightweight traffic density estimation method based on structured knowledge transfer. *Information Sciences,* 607, 947–960. doi.org/10.1016/j.ins.2022.06.047.

Hu, Z., Lam, W.H., Wong, S.C., Chow, A.H.F., Ma, W., 2023. Turning traffic surveillance cameras into intelligent sensors for traffic density estimation. *Complex & Intelligent Systems,* 9(6), 6587–6604. doi.org/10.1007/s40747-023-01117-0.

Maerivoet, S., De Moor, B., 2005. Traffic flow theory. arXiv. arxiv.org (15 August 2025).

OpenCV, n.d. Basic concepts of the homography explained with code. docs.opencv.org (8 August 2025).

Philippine Statistics Authority, 2024. Province of Iloilo's economy grows by 4.6 percent in 2023. rsso06.psa.gov.ph (20 July 2025).

Rodríguez-Rodríguez, J.A., López-Rubio, E., Ángel-Ruiz, J.A., Molina-Cabello, M.A., 2024. The impact of noise and brightness on object detection methods. *Sensors,* 24(3), 821. doi.org/10.3390/s24030821.

Ultralytics, n.d-a. Explore Ultralytics YOLOv8. docs.ultralytics.com (8 August 2025).

Ultralytics, n.d.-b: Model benchmarking with Ultralytics YOLO. docs.ultralytics.com (8 August 2025).

Ultralytics, n.d.-c: Performance metrics deep dive. docs.ultralytics.com (28 July 2025).

United Nations, Department of Economic and Social Affairs, 2015. The 17 Goals. sdgs.un.org (30 July 2025).

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., 2021. ByteTrack: Multi-object tracking by associating every detection box. arXiv. arxiv.org (8 August 2025).