E-ISSN 2586-9027
Homepage : *https://tci-thaijo.org/index.php/SciTechAsia* Science & Technology Asia
Vol.30 No.2 April - June 2025 Page: [193-203]
Original research article

Real-Time Traffic Surveillance: Aerial Vehicle Detection, Tracking, and Counting with YOLOv7

Vishakha Singh, Phisan Kaewprapha*

Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathum Thani 12120, Thailand

Received 16 February 2024; Received in revised form 21 April 2025
Accepted 30 April 2025; Available online 18 June 2025

ABSTRACT

In this research, we present an advanced aerial surveillance system powered by the YOLOv7 object detection model, designed for automatic and on-demand collection of traffic data. The system uses unmanned aerial vehicles (UAVs) to capture real-time video, making it especially valuable in areas without fixed surveillance cameras, such as rural roads and busy highways. It accurately detects, classifies, and tracks eight types of vehicles, and includes vehicle counting with directional analysis (left, right, or straight). This comprehensive approach enables the extraction of detailed traffic statistics, including flow rates, movement patterns, and vehicle density. Our classification model achieved an overall accuracy of 98.6%, with some vehicle types reaching up to 99.6%, demonstrating the system's strong performance and practical utility for traffic monitoring.

Keywords: Aerial surveillance system; Detection; Intelligent traffic management; Vehicle classification; YOLOv7

1. Introduction

In the backdrop of global technological advancements, the evolution of vehicle technology has been swift. Enhanced vehicle technology translates to increased speed, often leading to reckless driving practices. Notably, speeding is a prominent factor contributing to approximately 60% of road accidents [1]. However, these road-related mishaps and the challenges of traffic congestion can be mitigated through the implementation of an intelligent traffic management system.

Traditionally, traffic management has relied heavily on static infrastructure such as fixed sensors (e.g., loop detectors,

***Corresponding author:** kphisan@engr.tu.ac.th **doi:** 10.14456/scitechasia.2025.35

inductive loops) and stationary traffic cameras. While these conventional methods have provided essential data for traffic monitoring and control, they inherently possess limitations. Fixed sensors offer localized data and lack comprehensive spatial coverage, struggling to provide an overarching view of traffic flow across wider areas. Similarly, stationary cameras, despite their visual information, are restricted by their fixed viewpoints and coverage areas, often missing crucial events or data in unmonitored zones or during dynamic traffic changes. These limitations in capturing a holistic and adaptable view of traffic dynamics underscore the need for more flexible and comprehensive data acquisition systems, which forms the central theme of our research.

Intelligent traffic management not only has the potential to enhance road safety but also to significantly reduce traffic-related accidents and alleviate the burdens of congestion. As we look to the future, the role of surveillance, particularly through closed-circuit cameras (CCTV), will remain pivotal [2]. These unblinking sentinels have seamlessly integrated into contemporary traffic management systems, generating indispensable data streams for vehicle detection, tracking, and beyond [3]. The data they provide not only empowers real-time traffic analysis but also bolsters a nation's intelligence infrastructure, enhancing security and situational awareness on multiple fronts.

The analysis of the data yields a wealth of statistical information, including vehicle counts, directional flow patterns, and congestion levels, not only for the entire traffic ecosystem but also for specific vehicle types. This innovation, as explored in our study, stands as a beacon of hope in advancing traffic management capabilities.

Our research, centered on a vehicle detection and tracking system utilizing the YOLO algorithm [4], aims to contribute to this evolving landscape. It seeks to not only understand the intricacies of traffic dynamics but to proactively address the challenges posed by modern transportation. In the following sections of this paper, we delve deeper into our methodology, results, and insights, ultimately advocating for the broader adoption of intelligent traffic management systems to ensure safer, more efficient, and more secure roadways in the world's ever-advancing technological landscape.

2. Literature Review

The domain of object detection is currently undergoing a series of diverse experiments aimed at advancing its capabilities. These endeavors involve the exploration and integration of novel technologies across various applications. An illustrative example of this progress can be found in the work of Jang and Turk, who introduced the ”Car-Rec” recognition system, harnessing a blend of detection and extraction algorithms to discern both moving and stationary vehicles [5]. In a similar vein, the research presented in [6] employed YOLOv3 for detection and tracking, particularly from the vantage point of surveillance cameras. Furthermore, the study detailed in [7] stands out for its use of high-order statistics in the vehicle classification process post-identification. This research initiative involved the collection of class-related information through images, subsequently employed to predict the presence of a vehicle. These contributions collectively exemplify the dynamic landscape of object detection and its multifaceted applications

The paper by Ammar et al. [8], addresses car detection in aerial images using

Convolutional Neural Networks (CNNs) and compares three state-of-the-art algorithms: Faster R-CNN, YOLOv3, and YOLOv4. The study analyzes two datasets with different conditions and hyperparameters to evaluate algorithm performance. YOLOv4 and YOLOv3 generally outperform Faster R-CNN, except when there are significant differences in object sizes between training and testing datasets.

In a like manner, the paper by Berwo M. A. et al. [9], conducts a comprehensive survey of using Deep Learning (DL) techniques in vehicle detection and classification, examining successes, limitations, benchmark datasets, loss and activation functions, recent experiments, and technical advancements. It also discusses challenges and suggests future research directions, offering insights for the development of neural networks and related learning frameworks.

In the paper by Alamgir R. M. et al. [10], they focused on the significance of vehicle detection in various automation and intelligent systems, especially in Bangladesh. They highlighted the growing influence of Deep Learning models in this field, with a specific emphasis on real-time applications. Their research compares different YOLO-based architectures, including YOLOV3, YOLOV5s, and YOLOV5x, for efficient vehicle detection in Bangladeshi traffic images. The evaluation, based on a diverse dataset of 21 vehicle types, concludes that the YOLOV5x variant outperforms YOLOv3 and YOLOv5s with a 7% and 4% higher mean Average Precision (mAP), and 12% and 8.5% higher accuracy, respectively.

Similarly, [11] addresses the significant issue of wrong-way driving, a leading cause of road accidents and traffic congestion worldwide. The paper introduces an automated wrong-way vehicle detection system for on-road surveillance camera footage. The system comprises three stages: vehicle detection using the YOLO algorithm, vehicle tracking with the centroid tracking algorithm, and identification of wrong-way driving vehicles. YOLO is employed for precise object detection, while centroid tracking efficiently tracks moving objects. Their experiments conducted on various traffic videos demonstrated the system's ability to detect and identify wrong-way vehicles under diverse lighting and weather conditions.

Likewise, [12] talks about the need for accurate traffic sign detection in the context of intelligent driving technology. Existing methods based on color or shape recognition are limited in terms of recognition categories and accuracy. To tackle these limitations, the paper proposes an enhanced YOLOv5 method. It incorporates the SIoU loss function to optimize the training model and introduces the CSP1_3CBAM model, combining the CSP1_3 model with the convolutional block attention model (CBAM) to improve feature extraction. Additionally, the activation function ACONC enhances YOLOv5's generalization ability. Experimental results on the TT100k dataset demonstrate significant improvements in precision, recall, mAP, and frames per second (FPS). The algorithm's generalization ability is further validated on the GTSDB traffic sign dataset.

The above research inspired us to select YOLO model for our work. The paper by Redmon et al. [13] provides a good foundation for the YOLO model. Their model is an innovative object detection approach that frames the problem as regression, predicting bounding boxes and class probabilities directly in one neural network evaluation.

YOLO is exceptionally fast, offering real-time processing. It outperforms other detectors in generalization across different domains, making it a powerful detection system.

The original model has undergone numerous iterations and improvements, evolving from YOLO to versions 2, 3, 4, and onward, ultimately reaching version 7.

In the paper by Wang et al. [14], YOLOv7 was introduced. It sets new standards for both speed and accuracy in object detection, performing exceptionally well in the 5 FPS to 160 FPS range. This model attains the highest accuracy among real-time detectors and surpasses the performance of other notable models, including both transformer-based and convolutional-based detectors. Notably, YOLOv7 excels in both speed and accuracy compared to several state-of-the-art object detectors, even when trained exclusively on the MS COCO dataset from scratch.

In our previous study [15], we trained YOLOv7 on a dataset comprising eight vehicle types and achieved an overall detection accuracy of 99.5%, with individual class accuracies reaching up to 93%. However, the dataset used in that work was relatively limited in diversity and size. In this paper, we significantly enhance our previous work by expanding the dataset to include a larger and more diverse set of images, thereby improving the model's generalization capability. Furthermore, we extend the scope of our research by incorporating object tracking and counting, aiming to develop a more comprehensive and practical real-time vehicle analysis system.

3. Dataset

Our experiment involves a dataset comprising eight distinct vehicle types commonly used in Thailand, which encompass cars, buses, motorcycles, vans, trailer trucks, semitrailers, 10-wheeler trucks, and pickup trucks. Fig. 1 exhibits all the vehicle types below. We partitioned our dataset that consisted of approximately 5000 annotated images into training, validation, and testing subsets, with an 80:10:10 ratio, respectively. The data acquisition was primarily conducted during daytime hours, capturing scenes under sunny and slightly cloudy weather conditions. The majority of the images (approximately 85%) were taken with clear or mostly clear skies, while the remaining 15% include instances with scattered clouds that subtly alter the lighting. While the dataset effectively captures variations in traffic density, ranging from free-flowing to moderately congested scenarios during the day, it does not currently include images captured during nighttime, heavy

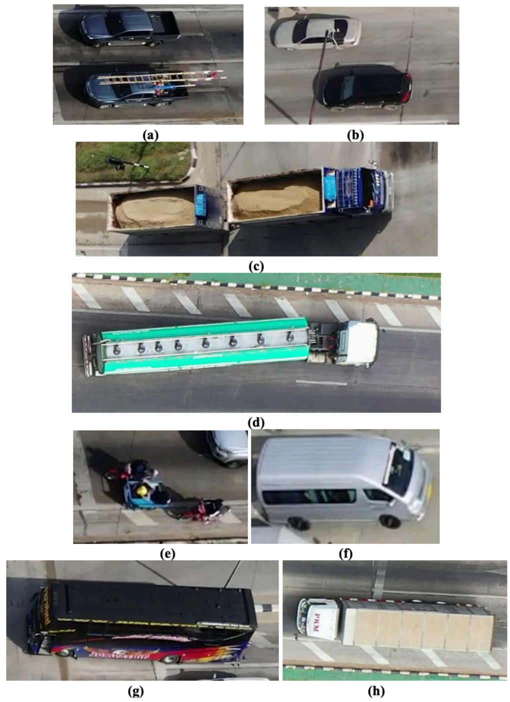


Fig. 1. (a) Pickup Truck (b) Car (c) Trailer (d) Semitrailer (e) Motorcycle (f) Van (g) Bus (h) 10-wheeler truck.

Fig. 2. Training dataset.

rainfall, or other adverse weather conditions. The aerial viewpoint remained consistent throughout the data collection process.

To improve the precision and robustness of our vehicle detection model, we implemented a suite of image augmentation techniques during the training phase. These augmentations included horizontal flipping, saturation adjustments (applied within a range of ±20%), random rotation (up to ±10○), and exposure modifications (adjusted by a factor between 0.8 and 1.2). To evaluate the effectiveness of these augmentation strategies, we conducted ablation experiments. We trained the YOLOv7 model both with and without the application of these augmentations, keeping all other hyperparameters constant.

The results of these experiments, presented in Table 2, demonstrate the positive impact of our augmentation pipeline on the model's performance. Specifically, we observed an improvement in the mean Average Precision (mAP) at an IoU threshold of 0.5 (mAP@0.5) from 96.1% without augmentation to 98.6% with augmentation. This 2.5% increase in mAP@0.5 indicates that the introduced variations in the training data helped the model generalize better and improve its ability to accurately detect vehicles under slightly different conditions of orientation, color, and brightness. Furthermore, we noted a consistent improvement across individual vehicle classes, suggesting that the augmentations contributed to a more robust feature learning process. These findings underscore the importance of employing appropriate data augmentation techniques to enhance the overall detection accuracy of our traffic monitoring system Fig. 2 shows the training dataset examples.

Each image file was accompanied by a corresponding text file containing its annotations. The annotation process was simplified through the use of ”labelImg” software, a widely recognized graphical annotation tool in the field. This software affords the flexibility of defining classes either prior to or during the annotation process, and it compiles all the class definitions into a corresponding text file. Two primary issues we encountered were the time-consuming nature of annotating thousands of individual vehicles and the potential for annotation errors and inter-annotator variability.

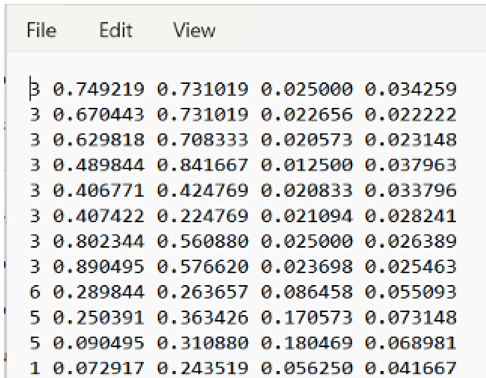
The task of manually drawing accurate bounding boxes around the numerous vehicles present within our 5000-image dataset proved to be exceptionally time-consuming. The sheer volume of annotation required significant time investment, and the manual process introduced potential for errors and inter-annotator inconsistencies. To address the time constraint, a team of two annotators worked concurrently. To mitigate annotation errors, we implemented a two-stage process: initial annotation followed by a review of 15% of images by a second annotator, with discrepancies resolved through consensus. Inter-annotator variability was minimized through the development of detailed annotation guidelines and the consensus-based review process. These measures ensured the high reliability of our ground truth data, crucial for the robust training of our vehicle detection model.

An illustration of the software interface is presented in Fig. 3, while Fig. 4 showcases the text file housing the annotations.

4. Methodology

We opted for the YOLOv7 algorithm in our study, primarily for its notable combination of speed and accuracy. YOLOv7 represents the latest iteration in the YOLO model series, renowned for its real-time object detection capabilities. Alexey Bochkovskiy, Chien-Yao Wang,

Fig. 3. LabelImg labelling annotations.



```
File    Edit    View

3 0.749219 0.731019 0.025000 0.034259
3 0.670443 0.731019 0.022656 0.022222
3 0.629818 0.708333 0.020573 0.023148
3 0.489844 0.841667 0.012500 0.037963
3 0.406771 0.424769 0.020833 0.033796
3 0.407422 0.224769 0.021094 0.028241
3 0.802344 0.560880 0.025000 0.026389
3 0.890495 0.576620 0.023698 0.025463
6 0.289844 0.263657 0.086458 0.055093
5 0.250391 0.363426 0.170573 0.073148
5 0.090495 0.310880 0.180469 0.068981
1 0.072917 0.243519 0.056250 0.041667
```

Fig. 4. Annotated file.

and Hong-Yuan Mark Liao are the creators of the model as outlined in their paper [14]. YOLO notably operates as a one-stage object detection model.

The YOLO machine learning model's architecture is designed for real-time object detection. It is a single-stage object detection system known for its speed and accuracy. YOLO consists of three primary components:

Backbone: This component acts as a feature extractor, extracting valuable information or feature maps from input data.

Neck: The neck component is responsible for merging representations from Convolutional Neural Network (ConvNet) layers before they are passed on to the head.

Head: YOLO comprises multiple heads, with the primary head responsible for generating the final output. Auxiliary heads assist during the training process.

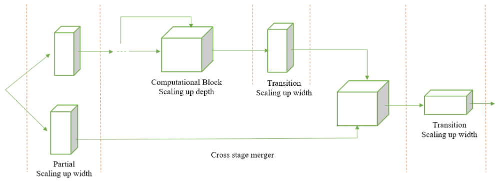


Fig. 5. Compound scaling up depth and width for concatenation-based model.

YOLO is recognized for its efficiency in real-time object detection tasks and has seen various iterations.

4.1 YOLOv7

It is an evolution of the YOLO model, and it comes with several improvements such as speed and efficiency, accuracy, and stability. YOLOv7 introduces two significant changes, primarily in its architecture and the integration of trainable bag-of-freebies.

YOLOv7 has restructured its architecture by incorporating the Extended Efficient Layer Aggregation Network (E-ELAN), enhancing its ability to acquire a broader range of features for improved learning. Furthermore, YOLOv7 scales its architecture by integrating elements from models like YOLOv4, Scaled YOLOv4, and YOLO-R from which it is derived. This adaptation enables the model to cater to varying requirements for different inference speeds. Fig. 5 shows the compound scaling for the concatenation-based model.

The concept of ”bag-of-freebies” is central to YOLOv7's approach, focusing on improving model accuracy without increasing training costs while simultaneously enhancing inference speed and detection accuracy. The re-parameterization planning, which follows training, is a technique that extends the training time but leads to improved inference results. Model-level reparametrization can be achieved through two primary methods: training multiple models with different training data but identical settings and averaging their weights for the final model, or by calculating the average of model weights at various training epochs.

This version is an advancement of the YOLO model series, focusing on improved speed, accuracy, and overall performance in real-time object detection which is why we have selected it for our work.

5. Results

This section will elucidate the parameters utilized, the performance metric applied, and the results obtained. The hyperparameters for YOLOv7 employed in our study are outlined in Table 1 below.

The selection of hyperparameters for the YOLOv7 model, outlined in Table 1, involved an iterative process guided by monitoring the model's performance on the validation set. We started with the default hyperparameters suggested in the YOLOv7 repository and then systematically adjusted key parameters such as the initial learning rate (lr0), momentum, and weight decay based on the observed training and validation loss, as well as the mAP. For instance, we experimented with different learning rates and learning rate schedules to find one that facilitated stable convergence and high performance. The warmup epochs and momentum were tuned to avoid instability during the initial training phase. The final values presented in Table 1 represent the configuration that yielded the best mAP on our validation set after several rounds of adjustments. While a comprehensive grid search was not feasible, this iterative tuning allowed us to find a set of hyperparameters well-suited to our specific dataset and task.

Training was executed by using 80% of the dataset in batches of 32, spanning

Table 1. Hyperparameters of the model.

Parameters	Description	Values
lr0	Initial learning rate	0.01
lrf	Final OneCycleLR learning rate	0.2
momentum	SGD momentum/Adam beta1	0.937
weight_decay	Optimizer weight decay 5e-4	0.005
warmup_epochs	Warmup epochs (fractions ok)	3.0
warmup_momentum	Warmup initial momentum	0.8
warmup_bias_lr	Warmup initial bias lr	0.1
box	Box loss gain	0.05
cls	Cls loss gain	0.5
cls_pw	Cls BCELoss positive_weight	1.0
obj	Obj loss gain	1.0
obj_pw	Obj BCELoss positive_weight	1.0
iou_t	Iou training threshold	0.20
anchor_t	Anchors per output layer (0 to ignore)	4.0
fl_gamma	Focal loss gamma	0.0
hsv_h	Image HSV-Hue augmentation (fraction)	0.015
hsv_s	Image HSV-Saturation augmentation (fraction)	0.7
hsv_v	Image HSV-Value augmentation (fraction)	0.4
degrees	Image rotation (+/- deg)	0.0
translate	Image translation (+/- fraction)	0.1
scale	Image scale (+/- gain)	0.5
shear	Image shear (+/- deg)	0.0
perspective	Image perspective (+/- fraction)	0.0
flipud	Image flip up-down (probability)	0.0
fliplr	Image flip left-right (probability)	0.5
mosaic	Image mosaic (probability)	1.0
mixup	Image mixup (probability)	0.0

a total of 200 epochs. Table 2 shows the comparison between the results we obtained with the augmentation versus without the augmentation. As seen below, our experiment achieved an overall mean average precision (mAP) of 98.6% across all classes. A detailed breakdown of individual class accuracies with mAP metrics is detailed in Table 2 below. The table above provides compelling evidence of YOLOv7's exceptional detection capabilities. The accompanying figure below illustrates the accuracy trends through plotting.

Table 2. Results of all classes.

Class	mAP (with augmentation)	mAP (Without augmentation)
All	98.6%	96.1%
10-wheeler truck	99.6%	97.8%
Bus	99.5%	98.9%
Car	99.2%	99%
Motorcycle	97.4%	95.2%
Pickup Truck	99.1%	97.6%
Semitrailer	98.5%	97.4%
Trailer	99%	98.5%
Van	96.3%	96%

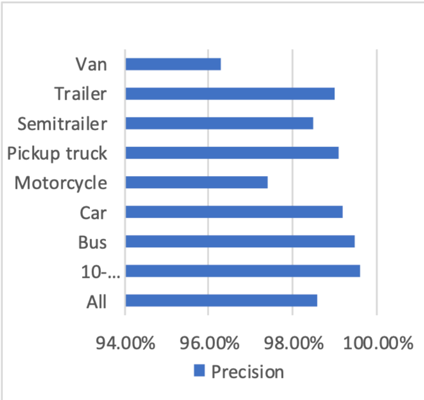


Fig. 6. Results of our dataset on YOLOv7.

Table 3. Results of YOLO models.

Metrics	YOLOv5	YOLOv6	YOLOv7
mAP	98.2	83.5	98.6
F1 Score	95	81.1	95.8
Precision	95.39	74.50	96.2
Recall	95.8	68.9	97.4

The achieved results are indeed remarkable, with an impressive overall accuracy of 98.6%, and some individual class accuracies reaching as high as 99.6%.

Table 3 presents a comprehensive comparison of three different YOLO models using multiple evaluation metrics to assess their performance. In addition to the widely used mean Average Precision (mAP), we also report precision, recall, and the F1-score to provide a well-rounded evaluation of each model's effectiveness in detecting vehicles. These metrics help highlight not only the accuracy of the models but also their balance between correctly detecting vehicles and minimizing false positives and false negatives.
Our system not only detects and tracks but also counts the number of vehicles passing through, taking their direction into account. Figs. 7-8 depict detection and tracking, while Fig. 9 illustrates detection, tracking, and vehicle counting. Figs. 7-8 not only

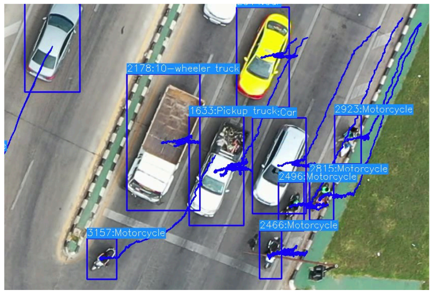


Fig. 7. Detection and tracking.

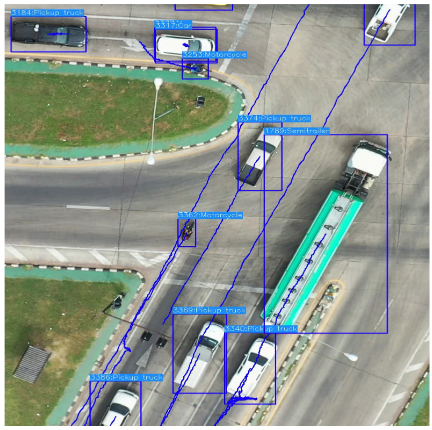


Fig. 8. Detection and tracking.

Fig. 9. Detection, tracking and counting of vehicles.

Fig. 10. Counting of vehicles.

display effective detection but also exhibit the trajectory of tracked vehicles, showcasing the precision and robustness of our real-time object detection system. It was also observed that vehicle classes with a higher number of training images achieved greater accuracy during the testing phase, underscoring the significance of data volume in model performance.

To enhance the functionality of our system, we integrated a robust tracking and counting mechanism. For tracking, we employed the ByteTrack algorithm, known for its reliability in maintaining object identity even in crowded scenes. For counting, we implemented a loop-based method, where virtual entry and exit zones are defined, and vehicle IDs are counted as they pass through these predefined loops. This method, as shown in Fig. 9, includes a visual loop at the top of the frame, and its corresponding count is reflected in the table at the bottom. Fig. 10 further demonstrates how vehicles are counted based on their direction, contributing to a comprehensive and accurate traffic monitoring solution.

6. Conclusion

Our experiment focused on employing YOLOv7 for the detection and tracking of eight distinct vehicle types commonly utilized in Thailand. We extracted frames from aerial view footage and proceeded to annotate them with the assistance of the LabelImg software. Our dataset encompassed around 5000 images, accompanied by four diverse augmentations. The dataset was partitioned in an 80:10:10 ratio for training, validation, and testing purposes. The hyperparameters employed in our study are comprehensively detailed in Table 1. As a result of our experimentation, we achieved an impressive overall accuracy of 98.6%, with individual detection accuracy reaching remarkable heights of up to 99.6%.

During the development, we encountered challenges such as managing the computational resources required for real-time processing of aerial video streams. Our solution involved optimizing the YOLOv7 model and utilizing cloud-based GPU acceleration to ensure efficient performance. While our initial experiments were conducted under favorable conditions, future work will need to address the integration of more robust drone platforms and power solutions for extended operational periods and varied environmental conditions.

Our work has achieved the capabilities of real-time detection, tracking, and vehicle counting, rendering it valuable for various applications. In the future, we aim to further enhance and refine our counting function to broaden its utility and effectiveness along with loops and tallying for every road.

Acknowledgements

We would like to thank the Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat School of Engineering (Thammasat University) for their valuable support and resources.

References

[1] World Health Organization. Road traffic injuries. Geneva: World Health Organiza-

tion; [cited 2022 Jun 20]. Available from: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] Caputo AC. Digital Video Surveillance and Camera. 2nd ed. Oxford: Butterworth-Heinemann; 2014.

[3] Li K, Deng R, Cheng Y, Hu R, Shen K. Research on vehicle detection and recognition based on infrared image and feature extraction. Mob Inf Syst. 2022;2022:1–10.

[4] Nelson A. Computer Vision: You Only Look Once (YOLO) Custom Object Detection with Colab GPU. 1st ed. Birmingham: Packt Publishing; 2020.

[5] Jang DM, Turk M. Car-Rec: A real time car recognition system. In: IEEE Workshop on Applications of Computer Vision (WACV); 2011 Jan 5–7; Kona, HI, USA. IEEE; 2011. p. 599–605.

[6] Song H, Liang H, Li H, Dai Z, Yun X. Vision-based vehicle detection and counting system using deep learning in highway scenes. Eur Transp Res Rev. Springer; [cited 2023 Feb 19]. Available from: https://link.springer.com/article/10.1186/s12544-020-00431-6.

[7] Rajagopalan AN, Burlina P, Chellappa R. Higher order statistical learning for vehicle detection in images. In: Proceedings of the Seventh IEEE International Conference on Computer Vision; 1999 Sep 20–27; Kerkyra, Greece. IEEE; 1999. p. 1204–9. vol. 2.

[8] Ammar A, Koubbaa A, Ahmed M, Saad A, Benjdira B. Vehicle detection from aerial images using deep learning: A comparative study. Electronics. 2021;10(7):820.

[9] Berwo MA, Khan A, Fang Y, Fahim H, Javaid S, Mahmood J, et al. Deep learning techniques for vehicle detection and classification from images/videos: A survey. Sensors (Basel). 2023;23(10):4832.

[10] Alamgir RM, Rahman M, Khan F, Fattah SA, Ahmed MR, Ali MH, et al. Performance analysis of YOLO-based architectures for vehicle detection from traffic images in Bangladesh. In: 2022 5th International Conference on Computer and Information Technology (ICCIT); 2022 Dec 17–19; Cox's Bazar, Bangladesh. IEEE; 2022. p. 982–7.

[11] Rahman Z, Ami AM, Ullah MA. A real-time wrong-way vehicle detection based on YOLO and centroid tracking. In: 2022 IEEE Region 10 Symposium (TENSYMP); 2022 Jul 1–3; Dhaka, Bangladesh. IEEE; 2022. p. 916–20.

[12] Zhang R, Zheng K, Shi P, Mei Y, Li H, Qiy T. Traffic sign detection based on the improved YOLOv5. Appl Sci. 2023;13(17):9748.

[13] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. IEEE; 2016. p. 779–88.

[14] Wang C, Bochkovskiy A, Liao L. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [Preprint]. arXiv:2207.02696 [cs.CV]. 2022. Available from: https://doi.org/10.48550/arXiv.2207.02696.

[15] Singh V, Kaewprapha P, Boonmee C. Ad-hoc Aerial-view Vehicle Detection and Tracking for Real-Time Traffic Monitoring Using YOLOv7. 2023 International Electrical Engineering Congress (iEECON). 2023 Mar;8:327–31.

[16] Cyganek B. Object Detection and Recognition in Digital Images. 1st ed. Chichester: John Wiley & Sons; 2013.