Vision-Based Vehicle Classification for Smart City

Ahsiah Ismail[1*], Amelia Ritahani Ismail[2], Nur Azri Shaharuddin[3], Muhammad Afiq Ara[4], Asmarani Ahmad Puzi[5], Suryanti Awang[6], Roziana Ramli[7]
[1,2,3,4,5]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Malaysia
[6]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia
[7]Department of Computer and Information Sciences, Northumbria University, United Kingdom
[1]ahsiah@iium.edu.my, [2]amelia@iium.edu.my, [3]n.azri@live.iium.edu.my, [4]afiq.ara@live.iium.my,
[5]asmarani@iium.edu.my,[6]suryanti@umpsa.edu.my,[7]roziana.ramli@northumbria.ac.uk
*Corresponding Author

Article Info

Article history:

Submission June 30, 2024
Revised September 23, 2024
Accepted June 19, 2025
Published July 11, 2025

Keywords:

Vehicle Classification
Image Recognition
Vision-Based
Smart City

ABSTRACT

Vehicle detection systems are essential for improving **traffic management**, enhancing safety, supporting law enforcement, facilitating toll collection, and contributing to **smart city initiatives** through real-time monitoring and data analysis. With the rapid growth of smart city technologies, the need for efficient, scalable, and high-accuracy vehicle detection models has become increasingly critical. This study aims to propose an advanced vehicle detection system using **Convolutional Neural Networks (CNNs)** in combination with the YOLOv5 model, which is known for its high-speed performance and superior accuracy in image recognition tasks. The proposed model is evaluated using a custom-trained YOLOv5s model, tested on a dataset comprising 1460 images of vehicles. These images are divided into five classes which are cars, motorcycles, trucks, ambulances, and buses. Performance evaluation metrics such as precision, recall, and mean Average Precision (mAP50-95) are used to assess the model's effectiveness. The results indicate that the YOLOv5-based model achieved impressive detection accuracy, with precision, recall, and mAP values exceeding 87%. The proposed system demonstrates its robustness in detecting and classifying various vehicle types across different conditions, including small, partially visible, and distant vehicles. The findings suggest that this model holds significant potential for real-world applications in urban traffic management and smart city infrastructure.

DOI: https://doi.org/10.34306/att.v7i2.446

1. INTRODUCTION

The advancement of technologies such as the Internet of Things (IoT) and Artificial Intelligence (AI), coupled with the vast availability of data, plays a critical role in the development of smart cities by enabling predictive analytics and data-driven decision-making particularly in urban transportation networks [1]. These technologies help enhance transportation infrastructure and improve road safety by employing image recognition and computer vision techniques to detect vehicles for effective traffic management, congestion control, and future planning [2]. Vision-based vehicle detection leverages sensors like cameras to automatically identify and monitor vehicles and their movements on the road [2].

This technology enables the automatic regulation of traffic flow in urban environments and facilitates incident detection such as accidents or unusual traffic patterns [3]. However, this depends on the availability of an accurate and reliable vehicle classification system capable of precisely identifying various types of vehicles under diverse conditions [4]. Therefore, this paper investigates the implementation of image recognition methods for vehicle classification using deep learning [5]. In particular, the ability to classify vehicles that are small, far away, or partially visible in low-light conditions remains a significant challenge that can directly affect system performance [6].

Despite recent advances, many models still struggle to maintain high classification accuracy in challenging real-world scenarios such as occlusion, dense traffic, and small object detection [7]. Previous studies have focused mainly on clear and centered vehicle images, lacking sufficient attention to faraway or overlapping vehicles [8]. This gap highlights the need for a more robust model capable of generalizing to various vehicle sizes and visibility levels [9]. Addressing this gap not only contributes to more resilient vehicle classification models but also strengthens intelligent transportation systems by improving traffic regulation, law enforcement, and road safety outcomes [10].

In this context, enhancing vehicle detection and classification contributes to broader societal goals, particularly the United Nations Sustainable Development Goals (SDGs) [11]. This research aligns with SDG 11 (Sustainable Cities and Communities) by promoting safer and smarter urban transportation and SDG 9 (Industry, Innovation, and Infrastructure) through the development of advanced AI-based systems [12]. By enabling real-time traffic monitoring and responsive management strategies, this research supports the transition toward more sustainable, efficient, and livable cities [13].

2. LITERATURE REVIEW

In vehicle classification, Selective Search Algorithm for small vehicles is investigated by [14]. They tested on Pascal VOC 2007 dataset to evaluate its performance in vehicle detection [15]. Their research highlighted the adaptability of Selective Search for small vehicles and introduced deep learning method [16]. In their research, using Selective Search Algorithm able to achived high classification [17]. However, they found that, it is challenges to balance between the accuracy and time complexity, especially for small instances [18].

Similarly, [19] explored advanced vehicle detection focusing on deep learning frameworks of R-CNN and YOLO [19]. They investigate on the enhancements in YOLO versions to improve small object detection and addressed computational complexities, highlighting advancements of Faster R-CNN and YOLOv4 for better efficiency [20]. They found that original Faster R-CNN less effecitive in detecting small objects and persistent challenges in detecting small vehicles, indicating areas for further innovation in vision-based vehicle detection systems [21].

[22] investigate the vehicle detection on the Real-Time Vehicle Classification and Tracking Using a Convolution Neural Networks (CNN), Transfer Learning-Improved Deep Learning Network which uses a combination of deep learning, YOLO-based object detection, transfer learning, and a multi-vehicle tracking method [22]. In their research, it is observed that the YOLOv5l, YOLOv5m, YOLOv5n and YOLOv5s model only able to detect the vehicle on a single straight sections of road an unable to detect the vehicle appear on bi-directional roads [23].

From the review conducted, the most commonly used methods in vehicle classification are CNN using YOLO model, particularly YOLOv5 [24]. The CNN is chosen due to its ability to effectively extract features from images, making it highly suitable for complex object detection tasks like vehicle recognition [22]. While YOLO's able to detects and classifies multiple vehicles at the same time, making it ideal for dynamic traffic surveillance scenarios [25]. Therefore, in this research, we investigate the CNN method with YOLOv5 model to improves the vehicle classification especially multiple, small and far away vehicle [26]. YOLOv5 was chosen over other versions due to its ability in detecting small objects [27].

3. METHODS

Generally, the proposed work can be divided into two phases, proposed model and web development [28]. Each of the phases will be discussed in the following subsections [29].

3.1. Proposed Model

The proposed model for vision-based vehicle classification using CNN with YOLOv5 model [30]. YOLOv5 was chosen due to its accuracy performance in detecting small objects [19]. The YOLOv5 architecture used as shown in Figure 1 [31].

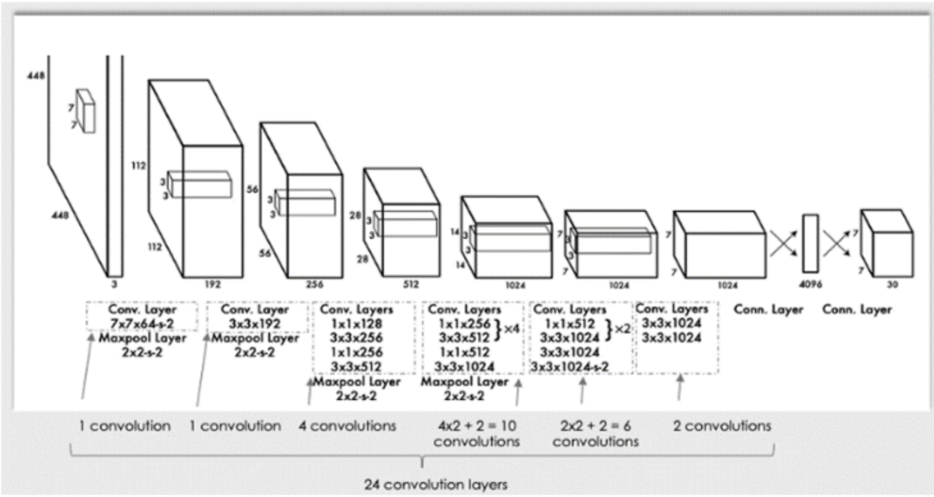

Figure 1. YOLOv5 Architecture

It consists of a series of convolutional and pooling layers that extract spatial features from the input image, followed by fully connected layers that perform classification based on the detected patterns [32]. The architecture is optimized for real-time object detection and is particularly effective in identifying small and overlapping vehicles in complex environments [33].

YOLOv5 architecture used for vehicle classification consists of 24 convolutional layers, four max-pooling layers, and two fully connected layers [34]. As illustrated in Figure 1, it focuses on localized feature analysis, analyzing vehicle parts of the image rather than the entire image to reduce computational complexity for real-time detection [35]. The convolutional network begins with a 1×1 convolution, followed by a 3×3 convolution that generates a cuboidal output [36]. A 1×1 convolution is used to reduce channel numbers, followed by a 3×3 convolution to extract feature maps [37]. To extract feature maps, the YOLO algorithm utilizes convolutions repeatedly [38].

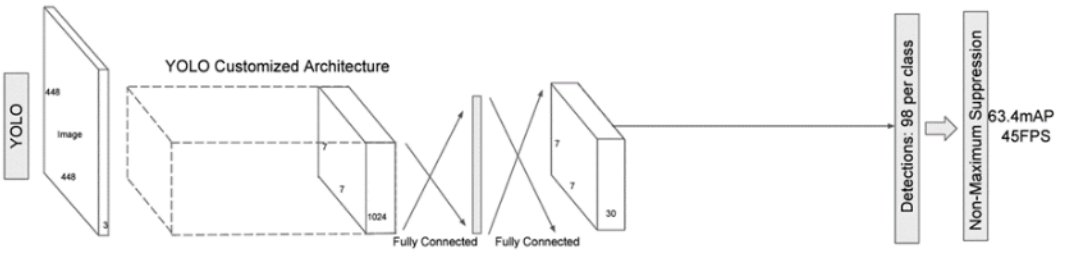

Figure 2. YOLO Algorithm Structure

The YOLO Algorithm structure is shown in Figure 2 [39]. Convolution extracts important features from vehicle input image by parameters Kernels, Strides, and paddings, which structure the operations and help in feature extraction [40]. The feature extraction process in our model incorporates multiple convolutional layers, CSP (Cross Stage Partial) modules, an SPPF (Spatial Pyramid Pooling Fast) layer, and a combination of upsampling and concatenation operations [41]. CSP modules is used to enhance gradient flow and improve learning efficiency [42]. The SPPF layer is included to generate fixed-size feature maps regardless of input vehicle image size, enabling the model to capture spatial information effectively [43].

A combination of upsampling and concatenation operations is used to merge features from different scales and levels of abstraction [44]. This multi-scale feature extraction ensures that the model can accurately detect various sizes vehicles within the image [45]. The refined features extracted through this process are then used by the detection layers to precisely identify and locate the vehicle, ensuring high accuracy and robustness in detection phase [46]. A ReLU and linear activation functions are used in our model architecture [47]. ReLU activation function is used throughout the network except in the final layer where a linear activation function is used in the final layer. To enhance performance and prevent overfitting, batch normalization techniques and dropout are used in our proposed model. Batch normalization helps to standardize inputs to each layer, making training more stable and efficient, while dropout randomly disables some neurons during training to prevent reliance on certain features and improve generalization [48].

3.2. Web Development

In the web development phase, the proposed model in the previous phase are integrated in the webpage. The webpage architecture for the proposed vision-based vehicle classification utilized the HTML, CSS, and JavaScript components. HTML structures the user interface, buttons, labels, headers, and images elements. CSS used to design the color schemes, button styles, font choices, and layout configurations. Meanwhile, JavaScript used to for image handling, server communication, response processing, and dynamic interface enhancements. The vision-based vehicle classification used Flask server to integrate with the proposed model. The Flask serve as the server which will call the proposed model functionalities in previous phase to load and produce inference results based on the input image from the user. Users can interact with the webpage by uploading images and initiate the ”scan image” function. The uploaded images will be sent to the Flask server, which employs sub processes and invoke the proposed model for predictions.

Figure 3. Vision-Based Vehicle Classification

Figure 3 shows the user interface of the vision-based vehicle classification system developed in the web implementation phase. The interface allows users to upload images and perform real-time vehicle detection using the integrated YOLOv5 model. On the left side of the interface, several traffic-related images are displayed to represent various conditions under which vehicles may be detected. The central section features the uploaded image preview, while the right section presents the classification results, including the types of detected vehicles (e.g., car, truck) and the computational time required for processing.

This visual interface enables users to interact with the system intuitively. As illustrated in figure above, once an image is uploaded and the ”Scan Image” button is clicked, the system processes the image through the back-end Flask server and returns the detection results in real time [49]. The design of the interface ensures clarity, accessibility, and responsiveness, which are essential for practical deployment in smart city applications [50].

3.3. Experiments

To demonstrate the reliability of the proposed model, a series of comprehensive experiments is conducted between two versions of YOLOv5 namely; YOLOv5s and YOLOv5x. Their performance are evaluated and compared. YOLOv5s is chosen due to its smallest and fastest model, which is suitable for real-time applications with limited computational resources. Although YOLOv5x is the largest and slowest model, the model is also consired in this research due to its promising accuracy performance in many object detection and classification task. Both of these models are trained using Google Colab with T4 GPU for faster processing. A vehicle image dataset is created to evaluate the performance of the two models.

Generally, the dataset consists of 1460 vehicle images classified into five categories of vehicle which are Car, Bus, Truck, Motorcycle and Ambulance. The vehicle images collected with different poses and angles including multiple, small and distant or far away vehicle. The properties and sample of vehicle dataset are shown in Table 1 and Figure 4.

Table 1. Details of the Characteristics in Vehicle Dataset

Category	**Total**
Car	370
Bus	329
Truck	312
Motorcycle	275
Ambulance	174
Total	**1460**

Table 1 presents the characteristics of the vehicle dataset, which consists of 1460 images categorized into five vehicle types: Car, Bus, Truck, Motorcycle, and Ambulance. The dataset includes a varied collection of images, captured from different poses and angles, which may feature vehicles from a close-up perspective, as well as distant or far-away views. The distribution of images across these categories is as follows: 370 images for Cars, 329 images for Buses, 312 images for Trucks, 275 images for Motorcycles, and 174 images for Ambulances. This diverse set of images provides a comprehensive sample for vehicle classification, ensuring that the dataset covers a wide range of vehicle types and visual perspectives, which is essential for building a robust model for vehicle recognition and analysis. The details outlined in Table 1 highlight the well-balanced nature of the dataset, with each category contributing a substantial number of images to the overall collection.

Figure 4. Examples of Images in the Vehicle Dataset

Figure 4 illustrates examples of images from the vehicle dataset used in the research. The images show various vehicles in different environments and perspectives, representing different vehicle types such as buses, cars, and trucks. These examples demonstrate the diversity of the dataset, which includes images captured from various angles and distances. The vehicles depicted in the figure serve as visual samples that reflect the range of conditions under which the dataset was collected, supporting the development of a robust vehicle classification model. As shown in Figure 4, the dataset consists of images taken in various settings, helping to ensure that the model can recognize vehicles under diverse circumstances. These images play a crucial role in training and validating the proposed model's effectiveness in distinguishing between vehicle categories.

In this research, the training split technique is used to split the vehicle dataset into training, validation and testing set. The dataset were divided into 70% training, 20% validation and 10% testing set. The validation set will helps in identifying the most efficient model and preventing the overfitting [51]. The aforementioned technique is considered in this research since most of the work related to this domain from the literature use this percentage to split the training, validation and testing set in their work [52]. The performance of the proposed model is measured in terms of Precision, Recall and Mean Average Precision (mAP) metric, define as follows:

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

where:

- AP_k = the AP of class k, n = the number of classes
- TP = True positive
- FP = False positive, FN = False negative

The True Positive (TP) is the correctly classifification of a vehicle image into its actual class [53]. False Positive (FP) is the incorrectly classifies a vehicle into a wrong classes. False Negative (FN) is the model unable to recognize a vehicle in its correct class and assigns it to another classes.

To obtain the most optimal settings for the porposed model, the experiment conducted with different batch sizes and epoch values. The parameter setting of the Confidence Threshold value are also evaluated to determine the optimal Confidence Threshold value. Three Confidence Threshold value are tested in this experiment. The experiment started with a low Confidence Threshold value of 0.01 and incremented to 0.05, and 0.1 to identify the most suitable Confidence Threshold value for the proposed model. The first experiment are tested with the parameter setting batch sizes of 16 and 32, with the epoch counts of 30 and 300. The maximum number of 32 for batch size and 300 for epochs in the experiment are chosen due to the maximum limit constraints of Google Colab for T4 GPU usage. Their performances are evaluated and compared.

4. RESULTS AND DISCUSSION

To evaluate the performance of the proposed model against different Confidence Threshold, the test with three different Confidence Threshold value (i.e., 0.01,0.5 and 0.1.) are conducted. The experiments started with a low Confidence Threshold value of 0.01 and incremented accordingly to identify the most suitable Confidence Threshold value that balances precision and recall for the proposed model. The comparative results of their precision and recall are presented in Table 2. From the results, it can be seen that the performance of both Confidence Threshold value of 0.05 and 0.1 demonstrate high precision and recall for almost all the vehicle classes with both Confidence Threshold value able to obtain 1.000 precision and 1.000 recall. Only a slightly lower in precision can be observed in Car class with both obtain 0.917 precision and 0.917 recall. While the lowest precision and recall can be observed in Confidence Threshold value of 0.01 for almost all classes with both precision and recall of 0.867 for Car class. Compared to the other two Confidence Threshold value (0.05 and 0.1) tested, a lower in precision can also be observed using the Confidence Threshold value of 0.1 in all other classes with only 0.857 for motorcycle, 0.800 for Bus and 0.875 for ambulance class. The high precision and recall using the Confidence Threshold value of 0.1 can only be observed in Truck class with 1.000 precision and 1.000 recall. Based on the result obtain, the Truck class is shown able to be detected not only by the Confidence Threshold value of 0.01, but also for all Confidence Threshold value tested. All the Confidence Threshold value tested able to obtain 1.000 for both performance measure of precision and recall for the Truck class.

Table 2. Comparative Results of the Proposed Model for Different Confidence Threshold Value

Class		**Confidence Threshold**		
		0.01	**0.05 (Proposed)**	**0.1**
Car	Precision	0.867	0.917	0.917
	Recall	0.867	1.000	1.000

Motorcycle	Precision	0.857	1.000	1.000
	Recall	1.000	1.000	1.000
Bus	Precision	0.800	1.000	1.000
	Recall	1.000	1.000	1.000
Ambulance	Precision	0.875	1.000	1.000
	Recall	1.000	1.000	1.000
Truck	Precision	1.000	1.000	1.000
	Recall	1.000	1.000	1.000

Table 2 presents the comparative results of the proposed model for different confidence threshold values, specifically for vehicle classification tasks. The table shows the precision and recall values for five vehicle categories Car, Motorcycle, Bus, Ambulance, and Truck at three different confidence threshold levels is 0.01, 0.05 (proposed), and 0.1.

As indicated in Table 2, the confidence threshold value of 0.05, which is the proposed value in this research, generally provides the best performance across all vehicle classes. The precision and recall values at this threshold are consistently high, reaching 1.000 for most vehicle categories, except for the Bus class, where the precision is 0.800 at the 0.05 threshold. The confidence threshold of 0.05 produced clearer bounding box detection, effectively handling images with multiple vehicles and detecting partially visible ones, especially in complex real-world environments like crowded traffic conditions.

In contrast, the 0.01 and 0.1 thresholds show lower or less reliable performance in vehicle detection, particularly in detecting vehicles that are partially obscured or located inside the image. The precision and recall values at the 0.1 threshold do not exceed 0.917, indicating less efficient detection when compared to the 0.05 threshold. Therefore, Table 2 highlights that the confidence threshold of 0.05 is the optimal choice for ensuring robust and accurate vehicle classification in the proposed model.

Overall, from the result, it can be seen that both the Confidence Threshold value of 0.05 and 0.1 demonstrated high precision and recall rates. However, the detection in Confidence Threshold value of 0.1 not able to detect vehicles which are partly inside in the image. In contrast, the Confidence Threshold value of 0.05 produced clearer bounding boxes detection and able to detect multiple inferences on a single vehicle images compared to other Confidence Threshold value. The Confidence Threshold value detection on vehicle images are shown in Figure 4. In addition, the Confidence Threshold value of 0.05 also effective in detecting partially visible vehicles in the image. This will enhanced vehicle detection especially for real-world applications where vehicles may be partially obscured or occluded especially in crowded traffic conditions or urban environments. From the result obtain, the Confidence Threshold value of 0.05 also shows the highest performance measure for all performance measure of prescion and recall across all the tested vehicle class with the value exceeding 0.917. Thus, the Confidence Threshold value of 0.05 are used in the proposed model as it able to produce clear and accurate detections, making it the optimal Confidence Threshold value for robust and reliable vehicle classification.

(a)

(b)

(c)

Figure 5. Confidence Threshold Value Detection on Vehicle Images (a) Confidence Threshold of 0.01 (b) Confidence Threshold of 0.05 and (c) Confidence Threshold Value of 0.1

To obtain the most optimal model, the proposed Confidence Threshold value of 0.05 are utilized in YOLOv5 and are tested in vehicle dataset. The YOLOv5 model is chosen due to its high-speed performance, making it ideal for vehicle classification. The selection of YOLOv5 models in this research are YOLOv5s and YOLOv5x model. YOLOv5s is the smallest and fastest model,which suitable for real-time applications with limited computational resources. Although YOLOv5x is the largest and slowest model, YOLOv5x model is also considered in this research due to its promising accuracy performance in many classification task. In searching for the best vehicle classification model, the effect of hyper-parameter tuning of batch size and epochs are evaluated to find the right balance between training efficiency, model accuracy, and generalization. The parameter settings used for each model and their results are summarized in Table 3.

Table 3. Parameters Settings and Comparative Results for Each Models

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5 (Proposed)**
Batch Size	16	16	32	32	32
Epochs	30	300	30	300	30
Precision	0.898	0.942	0.874	0.921	0.961
Recall	0.892	0.855	0.917	0.875	0.892
mAP50	0.938	0.930	0.952	0.940	0.947
mAP50-95	0.796	0.821	0.820	0.826	0.871

From the experiment conducted on the evaluation of batch sizes and epoch counts as presented in Table 3, Model 3 achieved the highest performance of recall and mAP50 (across all classes) with 0.917 recall and 0.952 mAP50. Model 3 based on the YOLOv5s models, which was trained with a batch size of 32 and 30 epochs able to detect all the five vehicle classes, particularly for Ambulance and Truck classes. From the result, it shows that the selection batch size of 32 able to balance between training speed and model stability, ensuring efficient use of computational resources while maintaining its performance. In addition, training for 30 epochs prevents overfitting, which can occur with prolonged training periods, ensuring the model generalizes well to unseen data. Although the Model 4 that based on YOLOv5s model was trained with a batch size of 32 and 300 epochs demonstrates a slightly higher mAP50-95 (across all classes) which only minimal diffrence of 0.006 compared to Model 3, it required more training time as the number of epochs increase in Model 4. Increases the number of epoch may risk overfitting, which may degrade performance on the new data [11]. Model 3 able to obtained high recall value with 0.917 and mAP50 of 0.952 and outperform the other models. The results indicate that the selection batch size of 32 with 30 epochs on YOLOv5s models able to optimize the performance where it balance the training data with strong generalization capabilities. A batch size of 32 ensure the stablity for training, while 30 epochs prevent overfitting, helping the model to generalize well to a new data.

Overall, Model 3 outperfomed the others especially for mAP50 and recall in the experiment conducted. On the other hand, Model 5 outperformed other for mAP50-95 and precision with 0.871 mAP50-95 and 0.961 precision. Despite having the same batch size (32) and epoch count (30) as other models, Model 5 which based on YOLOv5x, able to achive highest score of mAP50-95 and precision. This superior performance is due to YOLOv5x is a more advanced model than YOLOv5s, which allows more detailed feature extraction that leads to higher precision and mAP50-95 scores.

To further evaluate the performance of the models, the results for each vehicle classes were further explored and presented on Table 4.

Table 4. Comparative Results for Each Models on Vehicle Classes

Class		Model 1	Model 2	Model 3	Model 4	Model 5 (Proposed)
Ambulance	mAP50	0.952	0.965	0.972	0.962	0.993
	mAP50-95	0.842	0.896	0.911	0.885	0.952
Bus	mAP50	0.992	1.000	0.994	0.994	0.994
	mAP50-95	0.912	0.921	0.913	0.913	0.965
Car	mAP50	0.963	0.946	0.964	0.958	0.960
	mAP50-95	0.854	0.880	0.859	0.891	0.914
Motorcycle	mAP50	0.860	0.842	0.878	0.861	0.875
	mAP50-95	0.605	0.633	0.621	0.639	0.695
Truck	mAP50	0.924	0.904	0.951	0.924	0.916
	mAP50-95	0.765	0.773	0.798	0.800	0.828

Overall, Model 5 recorded the highest mAP50-95 for all the vehicle classes. Model 5 shows consistency in obtaining highest performance across all vehicle classes. The highest mAP50-95 of 0.965 can be observed on Bus detection. While the lowest mAP50-95 for all the model can be seen on Motorcycle. For the Motorcycle class, Model 5 able to obtain the highest mAP50-95 compared to other models with 0.695 mAP50-95. The results indicates Model 5 which is based on the YOLOv5x models with batch a size of 32 and epoch count 30 are more robust compared to other models. Thus in this research, Model 5 is proposed for vision-based vehicle classification.

Despite the effectiveness of the proposed model in detecting the vehicle images, missclassification still can be observed especially on white-color vehicles. The white-color vehicles tend to be missclassified as an ambulance. The example of the missclassifcation of ambulance are shown in Figure 6.

Figure 6. Misclassification of White-Color Vehicles

Figure 6 illustrates the misclassification of white-color vehicles, specifically ambulances, in the vehicle detection model. The images show examples where white-colored vehicles, including ambulances, are incorrectly classified by the model. The bounding boxes in the images highlight the misclassified vehicles, with the confidence scores displayed next to each detection. In the first image, an ambulance is misclassified with a confidence score of 0.39, and in the second image, it is misclassified with a higher confidence score of 0.65. Similarly, in the third image, another ambulance is misclassified with a confidence score of 0.62.

As shown in Figure 6, these misclassifications of white-color vehicles, particularly ambulances, are a significant challenge for the model. Despite the model's strong performance in detecting other vehicles, white-colored vehicles tend to be misclassified, possibly due to their color and design, which makes them difficult to distinguish accurately from other types of vehicles in certain conditions. This issue highlights a limitation in the model's ability to correctly classify all vehicle types, especially in real-world scenarios where environmental conditions and vehicle colors can impact detection accuracy.

5. CONCLUSION

The proposed Vision-Based Vehicle Classification model for smart city applications demonstrates significant potential in enhancing traffic management and congestion control. The experimental results reveal that the YOLOv5x-based Model 5 effectively classifies vehicles under various conditions, including the detection of multiple, small, and distant vehicles, making it highly suitable for real-world dynamic traffic environments. The model outperforms the YOLOv5s model, achieving the highest mAP50-95 value of 0.871 across all vehicle classes. Despite its high accuracy, the model encounters challenges in classifying white-colored vehicles, which are sometimes misclassified as ambulances.

Future improvements will focus on addressing this misclassification issue, specifically with white-colored vehicles. This can be achieved by incorporating more advanced detection models and expanding the dataset to include a greater diversity of vehicle types, especially ambulances and other white-colored vehicles. Additionally, overcoming GPU computation limitations, which affect training epochs and batch sizes, is essential for optimizing model performance. Exploring newer YOLO versions, such as YOLOv9 and YOLOv10, may offer improved accuracy and processing efficiency, enhancing the model overall performance.

The findings of this research have significant implications for the development of intelligent transportation systems in smart cities. Accurate vehicle classification, even in complex and crowded environments, supports the advancement of traffic regulation, law enforcement, and safety measures. The integration of this system into urban infrastructures can lead to more efficient road traffic management, enhancing public safety and contributing to the creation of sustainable, smart cities. Furthermore, the proposed model lays the groundwork for future developments in automated vehicle classification systems, driving innovation in smart city technologies.

6. DECLARATIONS

6.1. About Authors

Ahsiah Ismail (AI) https://orcid.org/0000-0003-3057-3734

Amelia Ritahani Ismail (AR) https://orcid.org/0000-0002-7126-4287

Nur Azri Shaharuddin (NA) -

Muhammad Afiq Ara (MA) -

Asmarani Ahmad Puzi (AA) https://orcid.org/0000-0001-8796-7289

Suryanti Awang (SA) https://orcid.org/0000-0002-5468-1150

Roziana Ramli (RR) https://orcid.org/0000-0002-3763-3149

6.2. Author Contributions

Conceptualization: AI; Methodology: AR; Software: NA; Validation: MA and AA; Formal Analysis: SA and RR; Investigation: AI; Resources: AR; Data Curation: NA; Writing Original Draft Preparation: MA and AA; Writing Review and Editing: SA and RR; Visualization: AI; All authors, AR, NA, MA, AA, and SA, have read and agreed to the published version of the manuscript.

6.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.4. Funding

This research was funded by UMP-IIUM Sustainable Research Collaboration 2022 grant (IUMP-SRCG22-015-0015).

6.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] B. Kidmose, "A review of smart vehicles in smart cities: Dangers, impacts, and the threat landscape," *Vehicular Communications*, p. 100871, 2024.

[2] D. Oladimeji, K. Gupta, N. A. Kose, K. Gundogan, L. Ge, and F. Liang, "Smart transportation: an overview of technologies and applications," *Sensors*, vol. 23, no. 8, p. 3880, 2023.

[3] B. E. Maryono, V. R. Handoko, A. Maduwinarti, and C. Gusmao, "Analysis of npm based public policy to enhance sustainability performance: A grounded theory," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 1, pp. 97–108, 2025.

[4] D. K. Dewangan and S. P. Sahu, "Towards the design of vision-based intelligent vehicle system: methodologies and challenges," *Evolutionary Intelligence*, vol. 16, no. 3, pp. 759–800, 2023.

[5] A. Sabha and A. Selwal, "Towards machine vision-based video analysis in smart cities: a survey, framework, applications and open issues," *Multimedia Tools and Applications*, vol. 83, no. 22, pp. 62 107–62 158, 2024.

[6] Q. Aini, D. Manongga, U. Rahardja, I. Sembiring, and Y.-M. Li, "Understanding behavioral intention to use of air quality monitoring solutions with emphasis on technology readiness," *International Journal of Human–Computer Interaction*, pp. 1–21, 2024.

[7] A. A. Iswandi, H. R. Ngemba, S. Hendra, M. N. D. Ulhaq, A. Iswandi, H. Ngemba, S. Hendra, S. Syahrullah, and M. Noval Daffa Ulhaq, "Implementasi algoritma rc4 pada sistem informasi pelaporan dukcapil provinsi sulawesi tengah," 2024.

[8] J. D. Trivedi, S. D. Mandalapu, and D. H. Dave, "Vision-based real-time vehicle detection and vehicle speed measurement using morphology and binary logical operation," *Journal of Industrial Information Integration*, vol. 27, p. 100280, 2022.

[9] A. Chetouane, S. Mabrouk, I. Jemili, and M. Mosbah, "Vision-based vehicle detection for road traffic congestion classification," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 7, p. e5983, 2022.

[10] R. Sivaraman, M.-H. Lin, M. I. C. Vargas, S. I. S. Al-Hawary, U. Rahardja, F. A. H. Al-Khafaji, E. V. Golubtsova, and L. Li, "Multi-objective hybrid system development: To increase the performance of diesel/photovoltaic/wind/battery system." *Mathematical Modelling of Engineering Problems*, vol. 11, no. 3, 2024.

[11] A. Rehman, T. Saba, M. Z. Khan, R. Damaševičius, and S. A. Bahaj, "Internet-of-things-based suspicious activity recognition using multimodalities of computer vision for smart city security," *Security and communication Networks*, vol. 2022, no. 1, p. 8383461, 2022.

[12] T. G. Mondal and M. R. Jahanshahi, "Applications of computer vision-based structural health monitoring and condition assessment in future smart cities," *The rise of smart cities*, pp. 193–221, 2022.

[13] S. I. Al-Hawary, J. R. N. Alvarez, A. Ali, A. K. Tripathi, U. Rahardja, I. H. Al-Kharsan, R. M. Romero-Parra, H. A. Marhoon, V. John, and W. Hussian, "Multiobjective optimization of a hybrid electricity generation system based on waste energy of internal combustion engine and solar system for sustainable environment," *Chemosphere*, vol. 336, p. 139269, 2023.

[14] L. Masello, G. Castignani, B. Sheehan, F. Murphy, and K. McDonnell, "On the road safety benefits of advanced driver assistance systems in different driving contexts," *Transportation research interdisciplinary perspectives*, vol. 15, p. 100670, 2022.

[15] I. G. N. A. K. Dwi, L. Bethany, O. Smith *et al.*, "Empowering tourism communication for sustainable village development," *Startupreneur Business Digital (SABDA Journal)*, vol. 3, no. 2, pp. 123–130, 2024.

[16] N. Shankaranarayan and S. S. Kamath, "Deep vision based vehicle retrieval for automated smart traffic surveillance systems," in *2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. IEEE, 2022, pp. 1–8.

[17] U. Rahardja, A. Sari, A. H. Alsalamy, S. Askar, A. H. R. Alawadi, and B. Abdullaeva, "Tribological properties assessment of metallic glasses through a genetic algorithm-optimized machine learning model," *Metals and Materials International*, vol. 30, no. 3, pp. 745–755, 2024.

[18] M. I. Basheer Ahmed, R. Zaghdoud, M. S. Ahmed, R. Sendi, S. Alsharif, J. Alabdulkarim, B. A. Albin Saad, R. Alsabt, A. Rahman, and G. Krishnasamy, "A real-time computer vision based approach to detection and classification of traffic incidents," *Big data and cognitive computing*, vol. 7, no. 1, p. 22, 2023.

[19] S. Pandey, B. Meenakshi, A. Sneha, N. Charan, and G. Monesh, "Research on enhancing accuracy in vehicle detection," in *2023 International Conference on Power, Instrumentation, Control and Computing (PICC)*. IEEE, 2023, pp. 1–6.

[20] M. Zaidan, N. T. Lapatta, and L. P. Pasha, "Optimalisasi pemasaran digital adaptif untuk mendorong keberlanjutan e-commerce di era transformasi digital," *ADI Pengabdian Kepada Masyarakat*, vol. 5, no. 1, pp. 7–17, 2024.

[21] D. Fernández Llorca, A. Hernández Martínez, and I. García Daza, "Vision-based vehicle speed estimation: A survey," *IET Intelligent Transport Systems*, vol. 15, no. 8, pp. 987–1005, 2021.

[22] P. Premaratne, I. J. Kadhim, R. Blacklidge, and M. Lee, "Comprehensive review on vehicle detection, classification and counting on highways," *Neurocomputing*, vol. 556, p. 126627, 2023.

[23] M. I. Pavel, S. Y. Tan, and A. Abdullah, "Vision-based autonomous vehicle systems based on deep learning: A systematic literature review," *Applied Sciences*, vol. 12, no. 14, p. 6831, 2022.

[24] M. N. Chan and T. Tint, "A review on advanced detection methods in vehicle traffic scenes," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2021, pp. 642–649.

[25] B. Neupane, T. Horanont, and J. Aryal, "Real-time vehicle classification and tracking using a transfer learning-improved deep learning network," *Sensors*, vol. 22, no. 10, p. 3813, 2022.

[26] R. G. Munthe, M. Abbas, R. Fernandez, and N. Ulita, "The impact of educational information systems on learning accessibility in higher education," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 94–103, 2024.

[27] V. A. Adewopo and N. Elsayed, "Smart city transportation: Deep learning ensemble approach for traffic accident detection," *IEEE Access*, 2024.

[28] A. Rashed, Y. Abdulazeem, T. A. Farrag, A. Bamaqa, M. Almaliki, M. Badawy, and M. A. Elhosseini, "Toward inclusive smart cities: Sound-based vehicle diagnostics, emergency signal recognition, and beyond," *Machines*, vol. 13, no. 4, p. 258, 2025.

[29] J. van der Merwe, S. M. Wahid, G. P. Cesna, D. A. Prabowo *et al.*, "Improving natural resource management through ai: Quantitative analysis using smartpls," *International Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 135–142, 2024.

[30] K. Darwhekar, A. Patil, S. Ghodke, R. Bawkar, and S. Rudrawar, "Computer vision based intelligent traffic management system," in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*. IEEE, 2022, pp. 1051–1056.

[31] D. Mavrokapnidis, N. Mohammadi, and J. Taylor, "Community dynamics in smart city digital twins: A computer vision-based approach for monitoring and forecasting collective urban hazard exposure," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2021.

[32] I. C. Kawi and W. Pontjoharyo, "Integrating buddhist principles into management control systems: A cattari ariya saccani approach," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 5, no. 2sp, pp. 146–161, 2023.

[33] T. Kustu and A. Taskin, "Deep learning and stereo vision based detection of post-earthquake fire geolocation for smart cities within the scope of disaster management: İstanbul case," *International journal of disaster risk reduction*, vol. 96, p. 103906, 2023.

[34] I. Gudmundsson and G. Falco, "Porting computer vision models to the edge for smart city applications: Enabling autonomous vision-based power line inspection at the smart grid edge for unmanned aerial vehicles (uavs)." in *HICSS*, 2022, pp. 1–10.

[35] A. H. Aditiya, H. Hamdan, S. N. W. Putra, S. Visiana, and R. Thakkar, "Transforming education with genai: Case study on chatgpt, midjourney, and policy changes," *Sundara Advanced Research on Artificial Intelligence*, vol. 1, no. 1, pp. 20–27, 2025.

[36] A. Aboah, "A vision-based system for traffic anomaly detection using deep learning and decision trees," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4207–4212.

[37] J. Mahale, D. Rout, B. Roy, and A. Khang, "Vehicle and passenger identification in public transportation to fortify smart city indices," in *Smart Cities*. CRC Press, 2023, pp. 231–255.

[38] T. Mariyanti, Y. Z. Basri, and J. Jazuli, "The basic factors driving the intention to pay zakat," *Aptisi Transactions on Management*, vol. 6, no. 1, pp. 30–41, 2022.

[39] R. Kashaf, E. P. Alegre, T. Prova, and S. Aggarwal, "Automated waste management using a customized vision-based transformer model," in *2024 IEEE World AI IoT Congress (AIIoT)*. IEEE, 2024, pp. 300–309.

[40] D. P. Lazirkha, J. Hom, and V. Melinda, "Quality analysis of digital business services in improving customer satisfaction," *Startupreneur Business Digital (SABDA Journal)*, vol. 1, no. 2, pp. 156–166, 2022.

[41] A. Boukerche and X. Ma, "Vision-based autonomous vehicle recognition: A new challenge for deep learning-based systems," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–37, 2021.

[42] D. Singh, A. Kumar, and R. Singh, "Multiple object tracking of autonomous vehicles for sustainable and smart cities," in *Object Tracking Technology: Trends, Challenges and Applications*. Springer, 2023, pp. 201–219.

[43] R. Aprianto, R. Haris, A. Williams, H. Agustian, and N. Aptwell, "Social influence on ai-driven air quality monitoring adoption: Smartpls analysis," *Sundara Advanced Research on Artificial Intelligence*, vol. 1, no. 1, pp. 28–36, 2025.

[44] J. Azimjonov and A. Özmen, "Vision-based vehicle tracking on highway traffic using bounding-box features to extract statistical information," *Computers & Electrical Engineering*, vol. 97, p. 107560, 2022.

[45] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering," *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 289–302, 2023.

[46] A. Ramadhan, B. Bhima, and T. Nurtino, "Integrated energy system systems and game theory a review," *International Transactions on Artificial Intelligence*, vol. 1, no. 1, pp. 84–101, 2022.

[47] I. Maria *et al.*, "Unlocking success: Human resource management for startupreneur," *Startupreneur Business Digital (SABDA Journal)*, vol. 3, no. 1, pp. 89–97, 2024.

[48] A. Vijayakumar and S. Vairavasundaram, "Yolo-based object detection models: A review and its applications," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83 535–83 574, 2024.

[49] Government of Malaysia, "Malaysia smart city framework (mscf)," Online; portal Malaysia, n.d. [Online]. Available: https://www.malaysia.gov.my/portal/content/30947

[50] Ministry of Housing and Local Government Malaysia (KPKT), "Smart city framework for malaysia: Executive summary," 2021. [Online]. Available: https://www.kpkt.gov.my/kpkt/resources/user_1/GALERI/PDF_PENERBITAN/FRAMEWORK/FRAMEWORK_SMART_CITY_EXECUTIVE_SUMMARY.pdf

[51] G. Fatima and S. Saeed, "A novel weighted ensemble method to overcome the impact of under-fitting and over-fitting on the classification accuracy of the imbalanced data sets," *Pakistan Journal of Statistics and Operation Research*, pp. 483–496, 2021.

[52] E. Casas, L. Ramos, E. Bendek, and F. Rivas-Echeverría, "Assessing the effectiveness of yolo architectures for smoke and wildfire detection," *IEEe Access*, vol. 11, pp. 96 554–96 583, 2023.

[53] S. Setiawan, K. A. Tiara, M. Rustine, D. Hilman, K. Joy, and I. A. Jumbri, "Experiential value and novelty cultivating brand love and behavioral intentions in technopreneurship," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 1, pp. 144–157, 2025.