IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3518~3527
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3518-3527

Performance evaluation of YOLOv11-based vehicle detection and tracking for urban intelligent transportation systems

Thang C. Vu[1], Dung T. Nguyen[1], Minh T. Nguyen[2], Long Q. Dinh[1], Mui D. Nguyen[2]
[1]Faculty of Engineering and Technology, Thai Nguyen University of Information and Communication Technology, Thai Nguyen, Vietnam
[2]Faculty of International Training, Thai Nguyen University of Technology, Thai Nguyen University, Thai Nguyen, Vietnam

Article Info

Article history:

Received Aug 13, 2025
Revised Jun 14, 2026
Accepted Jul 9, 2026

Keywords:

Convolutional neural network
Deep learning
DeepSORT technique
Vehicle detection
YOLOv11

ABSTRACT

This paper proposes and evaluates an integrated vehicle detection and tracking framework based on you only look once (YOLO)v11 for intelligent transportation systems (ITS). The combination of the deep simple online and real-time tracking (DeepSORT) algorithm helps maintain vehicle identity across consecutive frames, thereby enhancing the stability of the multi-object tracking system. Additionally, the slicing-aided hyper inference (SAHI) technique is integrated to improve the detection efficiency of small vehicles in remote sensing imagery and urban surveillance video data collected in Thai Nguyen, Vietnam. The system's performance is comprehensively evaluated through several key quantitative indicators, including mean average precision (mAP), multiple objects tracking accuracy (MOTA), and identification F1-score (IDF1), across realistic urban traffic scenarios. The results show that the framework significantly improves detection accuracy, tracking consistency, and small object recognition efficiency in real-world urban traffic scenarios. This paper provides useful insights for selecting appropriate detection and tracking configurations in ITS applications.

Corresponding Author:

Minh T. Nguyen
Faculty of International Training, Thai Nguyen University of Technology, Thai Nguyen University
Thai Nguyen, Vietnam
Email: nguyentuanminh@tnut.edu.vn

1. INTRODUCTION

Nowadays, in the field of smart transportation in urban areas, image data collected from traffic surveillance cameras is widely used for tasks such as vehicle tracking, traffic flow monitoring, as well as traffic congestion analysis and processing [1]–[3]. Solutions based on image processing and computer vision offer high flexibility and reasonable deployment costs [4], [5]. This contributes to improving environmental awareness and supporting decision-making for intelligent systems [6], [7]. Widely applicable fields include medical and healthcare [8], industry [9], agriculture [10], [11], transportation [12], [13], search and rescue [14], disaster response [15], and environmental research [16].

Machine learning has been widely used in computer vision fields. The background subtraction method is used [11] to identify moving vehicles, however the applicability of this method to many real traffic situations is limited because the data used for processing are standard images. For example, Chen *et al.* [17] proposed an improved version of you only look once (YOLO)v4 with an attention mechanism to enhance detection accuracy, although it also increased computational complexity. Among current methods, YOLO is considered one of the most efficient object detection architectures due to its high accuracy in real-time processing [18]. However, systems combining YOLOv3, YOLOv5, or faster region-based convolutional neural network (Faster R-CNN) with tracking algorithms such as simple online and real-time tracking (SORT) and DeepSORT still face limitations in complex urban traffic environments [2], [19].

This paper presents a comprehensive evaluation of the performance of a YOLOv11-based vehicle detection and tracking system for urban intelligent transportation systems (ITS) applications. The models were evaluated under the same experimental conditions using metrics such as mean average precision (mAP), multiple object tracking accuracy (MOTA), and identification F1-score (IDF1). The combination of the DeepSORT algorithm helps maintain vehicle recognition across consecutive frames, thereby enhancing the stability of the multi-object tracking system. Additionally, the slicing-aided hyper inference (SAHI) technique was integrated to improve the detection efficiency of small vehicles in remote sensing imagery and urban surveillance video data collected in Thai Nguyen, Vietnam.

The remainder of the paper is organized as follows. Section 2 presents an overview of the proposed methodology. Section 3 describes the experimental setup and detection-tracking methods based on YOLOv11-DeepSORT and YOLOv11-SAHI. Section 4 presents experimental results on still images and traffic surveillance video, along with analyses and discussions. Finally, section 5 provides conclusions and directions for future research.

2. METHOD OVERVIEW

2.1. Estimation theory

In (1) and (2) describe a set of states and measurements [20].

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{1}$$

$$z_{k+1} = Cx_{k+1} + v_{k+1} \tag{2}$$

Where w_k represents system noise and v_k is measurement noise. A is the system state matrix, B is the input control matrix, and C is the measurement matrix. The Kalman filter is constructed according to the predictor-corrector mechanism [21]. In the prediction stage, the system state is estimated based on the previous state value combined with the system's dynamic model as presented in (3). Then, the state error covariance matrix is determined according to (4).

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + u_k \tag{3}$$

$$P_{k+1|k} = AP_{k|k}A^T + Q_k \tag{4}$$

During the update phase, the Kalman gain coefficient is used to correct the prediction state based on new measurement information. The estimated state and the error covariance matrix are updated according to (5) and (6), respectively.

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\left(Z_{k+1} - C_{k+1}\hat{x}_{k+1|k}\right) \tag{5}$$

$$P_{k+1|k+1} = (I - K_{k+1}C_{k+1})P_{k+1|k}(I - K_{k+1}C_{k+1})^T + K_{k+1}R_{k+1|k}K_{k+1}^T \tag{6}$$

Where I is an identity matrix of size $n \times n$ with n being the number of states of the system; k represents the time step; $k|k$ corresponds to the value updated at the previous iteration, $k + 1|k$ represents the predicted value at time $k + 1$.

2.2. Overview of DeepSORT algorithm

The DeepSORT tracking process is illustrated in Figure 1 [22]. A sequence of video frames is input, and the endpoint is checked. The detected frames are compared with the predicted frames based on the intersection over union (IoU) index, thereby creating a cost matrix. This matrix is processed by the Hungarian algorithm to find the optimal matching pairs for three cases: trajectory mismatch, detection mismatch, and trajectory match. When there is a match, the Kalman filter is used to update and predict the state of the objects. This matching and updating process is repeated continuously until the trajectory of the object is confirmed or the video ends.

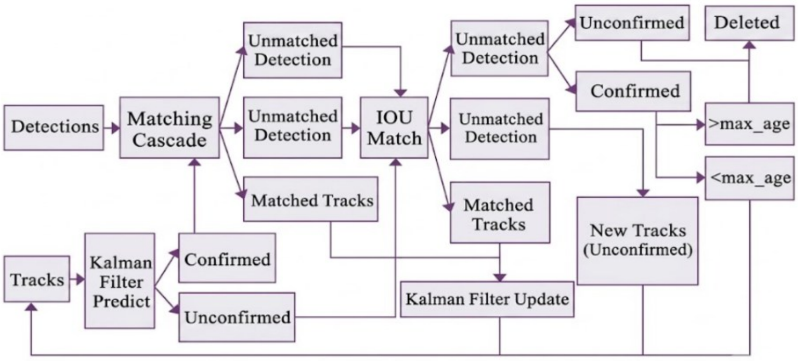


Figure 1. Vehicle tracking is based on DeepSORT algorithm

3. EXPERIMENTAL SETUP AND EVALUATION

3.1. Experimental dataset

All images were collected from real-world urban traffic in Thai Nguyen, Vietnam. Data, including both images and videos, were collected using urban traffic surveillance cameras and cameras integrated into unmanned aerial vehicles (UAVs). Image collection was conducted at various times of the day and under different traffic conditions (such as vehicle density, weather conditions, and partial occlusion). The collected data was used to evaluate vehicle detection and multi-object tracking performance in practical ITS applications such as traffic monitoring, vehicle counting, and flow analysis. A portion of the collected data was manually annotated to quantitatively assess the detection and tracking performance of the test scenarios. The remaining data was used for experimental analysis and evaluation of the proposed system. The test scenarios were implemented based on a customized YOLO model (100 epochs, 16 batches).

3.2. YOLOv11-based object detection and multi-object tracking with DeepSORT

Figure 2 illustrates the architecture of the YOLOv11 model and the integration process with the DeepSORT algorithm. As shown in Figure 2(a), YOLOv11 is built with spatial pyramid pooling – fast (SPPF), C2 with partial spatial attention (C2PSA), and C3 with kernel size 2 (C3K2) modules [23] to enhance feature extraction capabilities and object detection efficiency in complex traffic scenarios. To improve multi-object tracking efficiency, the DeepSORT algorithm integration process is illustrated in Figure 2(b). In this process, YOLOv11 performs vehicle detection on each frame independently, while DeepSORT handles linking and maintaining object identification between consecutive frames.

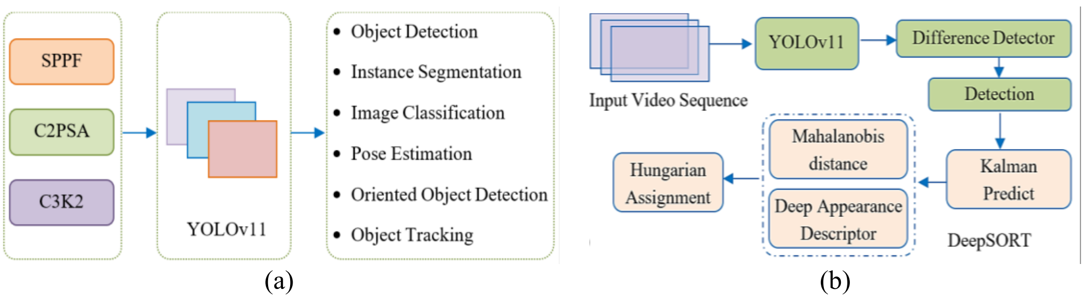
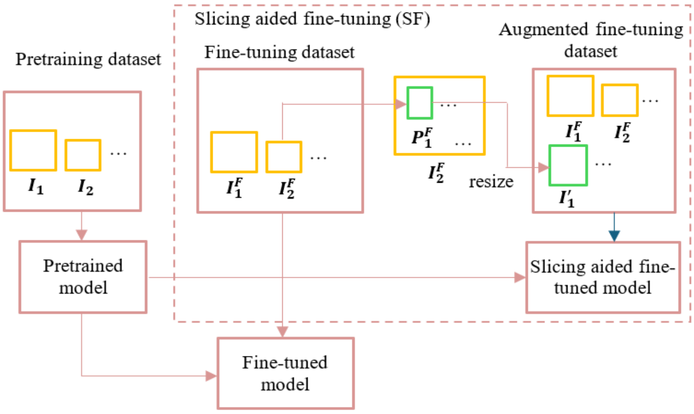


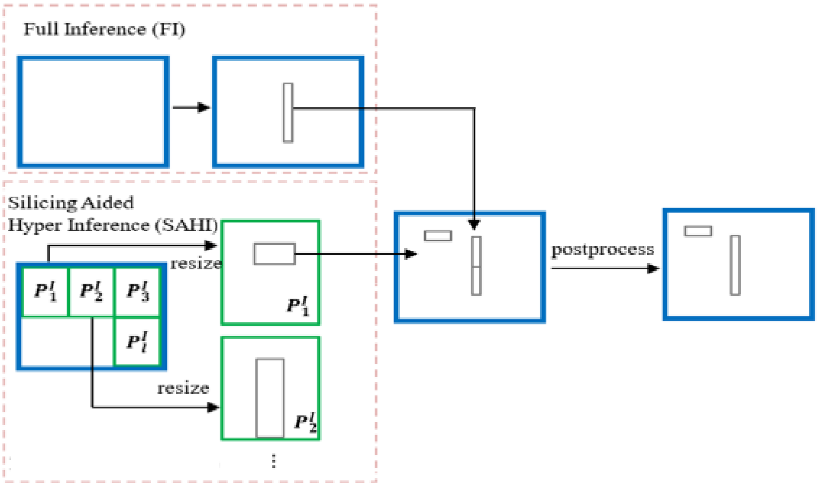
Figure 2. YOLOv11 model architecture and integration mechanism with the DeepSORT algorithm: (a) YOLOv11 model architecture and (b) YOLOv11 combined with the DeepSORT algorithm

3.3. Small object detection based on YOLOv11 and SAHI technique

Figure 3 shows an overview of the image slicing-based refinement and inference pipeline. As illustrated in Figure 3(a), during the training phase, the input image is divided into overlapping slices, thus increasing the representation rate of small objects in each image region. In the inference phase, described in Figure 3(b), the input image is further divided into smaller slices before being fed into the YOLOv11 model for object detection in each independent image region. Subsequently, the detection results from all slices are combined and processed using the non-maximum suppression (NMS) algorithm to eliminate overlapping predictions and produce a final detection result with higher accuracy.



(a)



(b)

Figure 3. YOLOv11 and SAHI-based small object detection: (a) image slicing-based refinement stage and (b) image slicing-based inference stage

Per image $I_1^F, I_2^F, \ldots, I_j^F$ cut into overlapping patches $P_1^F, P_2^F, \ldots P_k^F$ with fixed sizes M and N. During the fine-tuning phase, these slices are resized while maintaining the original aspect ratio to ensure consistency in spatial features. Then, the enhanced images along with the original image are used to train the model [24], [25]. This approach improves feature learning capabilities for small objects appearing in traffic monitoring data.

Slicing techniques are also applied in the inference phase. Specifically, the input query image I is split into multiple overlapping slices $P_1^I, P_2^I, \ldots P_l^I$ with dimensions $M \times N$. Each slice is resized and passed through the YOLOv11 model for independent object detection. Additionally, full inference (FI) can be performed on the original image to aid in the detection of larger objects. Finally, the detection results from the slices and the original image are mapped back to the original image size and merged using NMS.

4. SIMULATION RESULTS

4.1. Static image-based object detection

Figure 4 illustrates the experimental results of the proposed system framework. The detection results obtained using the YOLOv5 model are shown in Figure 4(a), where some small or partially obscured vehicles were missed despite the faster inference speed of 31.5 ms. In contrast, Figure 4(b) shows that the YOLOv11 model achieved more reliable detection with a higher confidence score, although it required a longer inference time of 83.3 ms. Furthermore, unobstructed objects achieved confidence scores up to 0.92, while smaller or obscured vehicles maintained confidence values ranging from 0.47 to 0.76. However, the cost of the YOLOv11 model is the increased computational complexity and higher resource requirements.

(a) (b)

Figure 4. Object detection results from a digital image (photo taken in central Thai Nguyen, Vietnam): (a) YOLOv5 model and (b) YOLOv11 model

4.2. Track vehicles in traffic monitoring video based on YOLOv11-DeepSORT

This section presents some experimental results of the proposed system framework for the problem of detecting and tracking vehicles from surveillance cameras. The experimental results on roads in Thai Nguyen, Vietnam are illustrated in Figure 5. After a vehicle is detected by YOLOv11, the system assigns a unique identifier (ID) to each object to support tracking multiple objects in consecutive frames. As shown in Figure 5(a), the system continuously tracks each ID in sequential frames. Furthermore, as described in Figure 5(b), the system enables the effective detection and tracking of partially or completely obscured vehicles by applying a Kalman filter. In addition, YOLOv11 shows effective object detection, especially with objects that are quite small in size or only partially visible in the frame. After the object detection process of the given model is completed, the video output with the detected objects is identified, and the parameters of the number of vehicles are displayed.

(a)

(b)

Figure 5. Vehicle tracking results using the YOLOv11-DeepSORT framework on traffic surveillance videos: (a) consistent vehicle ID assignment across consecutive frames and (b) robust tracking performance under partial occlusion

The system's video processing performance was evaluated using frames per second (FPS). The experimental hardware system was based on an Intel Core i5-6500 processor, NVIDIA GTX 1660Ti GPU, 16 GB of RAM, and 256 GB of storage. The results in Table 1 show an improvement in the detection accuracy of YOLOv11 compared to some previous generation models. Simultaneously, the integration of the DeepSORT algorithm significantly enhanced multi-object tracking efficiency, reflected in higher MOTA and IDF1. In particular, the YOLOv11-DeepSORT combined model demonstrated an effective balance between detection accuracy, recognition consistency, and tracking stability in a complex urban traffic environment. In terms of computing performance, the system achieved a processing speed of approximately 10.13 FPS. YOLOv11 has a higher model complexity, resulting in lower processing speeds compared to lighter versions of YOLO.

Table 1. Some results comparing the combination of YOLO models with the DeepSORT algorithm

Method	mAP@0.5	mAP@0.5:0.95	MOTA	IDF1	FPS
YOLOv5 + DeepSORT	0.58	0.36	0.62	0.66	19.2
YOLOv9 + DeepSORT	0.63	0.42	0.68	0.70	13.4
YOLOv11 + DeepSORT	0.66	0.45	0.70	0.73	10.1

To evaluate the system's accuracy, a manual validation process was performed on a selected subset of the test video dataset. Specifically, video segments with a total duration of approximately one minute (approximately 4,500 frames in total) were selected for experimentation in this scenario. In this process, the number of vehicles was manually counted as vehicles passed through established virtual zones, and the results were compared with the experimental results of the proposed system. Experimental results showed that the system's output achieved an average error rate of less than 5% (Table 2). However, in some experimental situations, the results were affected by short-term occlusion or the partial appearance of vehicles at the edge of the frame.

Table 2. The results count the number of vehicles using the integrated YOLOv11-DeepSORT framework

Video code	Manual count	System count	Absolute error	Error rate (%)
Video 1	70	67	3	4.29
Video 2	74	72	2	2.70
Video 3	61	58	3	4.92

4.3. Object detection based on static remote sensing images

The simulation of the YOLOv11 hybrid program with an advanced inference algorithm using image slicing in object detection from remote sensing images was performed by the authors on the Google Colab platform (allowing execution of Python commands on the cloud platform). The GPU parameters selected are as follows: GPU (T4), RAM (15 GB). The set of still images used in the study is in JPEG format (.png/.jpg/.jpeg), with a uniform size of 1276×788 pixels. This is an initial test dataset collected in a controlled manner, suitable for verifying the application potential of the system. The images are selected to simulate urban traffic situations with low to medium object density, reflecting real urban conditions, while ensuring the presence of many small objects.

In Figure 6, object detection experiments were conducted using the YOLOv11 model on aerial photographs taken in Thai Nguyen, Vietnam. Figure 6(a) presents the vehicle detection results based on the YOLOv11 model alone. In this scenario, some objects were missed due to their small size or strong occlusion. Conversely, Figure 6(b) shows that combining the YOLOv11 model with the SAHI technique significantly increased the number of detected objects, including small, distant, and partially occluded vehicles. In the test scenario, the confidence ranged from 0.57 to 0.89, indicating that the system was sufficiently sensitive and maintained its filtering capability.

Table 3 compares the detection performance of YOLOv11 with and without the SAHI strategy on aerial imagery. The results indicate that incorporating SAHI significantly improves detection accuracy for small-scale vehicles. This improvement demonstrates its effectiveness for small object detection.

4.4. Aerial video-based object detection

The experiment was conducted on a dataset extracted from UAV videos recorded in the inner city of Thai Nguyen, Vietnam. The extracted videos are in MP4 format (1280×720 pixels, 30 FPS, 15 seconds). The entire video frame will be divided into many small slices of fixed size with enough overlap to not miss objects at the boundary. Each slice is then fed into the YOLOv11 model for independent object detection. The results are then merged, and the original video frames containing the bounding boxes are reconstructed.

Experimental results show that the system is capable of effectively detecting many small objects such as people, cars, and motorcycles across the entire depth of the image. The detection boundaries are displayed along with their corresponding reliability, with larger objects generally achieving higher reliability levels, ranging from 0.83–0.92. The minimum detection threshold was set at approximately 0.3 to enhance recognition capabilities in complex traffic environments, as illustrated in Figure 7. Objects are detected in the right position and are classified quite accurately, even when their display size is very small. Experimental results show that detection reliability depends on object size and scene complexity. In aerial video scenarios, small or distant vehicles generally yield lower reliability results compared to larger vehicles (0.3-0.5). Furthermore, when using only the YOLOv11 model, some small or hidden objects may not be fully detected.

(a) (b)

Figure 6. Small object detection results in remote sensing images in Thai Nguyen, Vietnam: (a) detection results based on the YOLOv11 model and (b) detection results using YOLOv11 combined with the SAHI technique using the image slicing method

Table 3. Comparative results of YOLOv11 and YOLOv11 combined with SAHI for small object detection

Configuration	mAP@0.5	mAP@0.5:0.95	AP_S
YOLOv11	0.66	0.45	0.29
YOLOv11 + SAHI	0.70	0.48	0.40

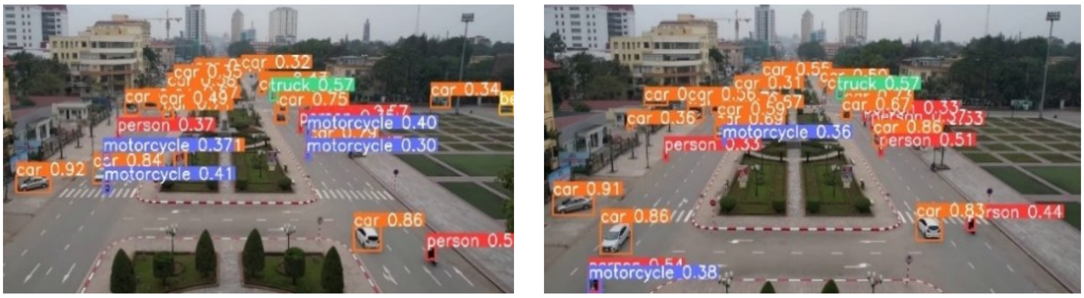


Figure 7. Sample object detection results of YOLOv11 with SAHI on UAV traffic data in Thai Nguyen, Vietnam

4.5. Discussion

Experimental results show that the YOLOv11 model, combined with the DeepSORT algorithm, reliably detects and tracks vehicles in urban traffic video. The proposed solution framework maintains object recognition across consecutive frames. The use of Kalman filter contributes to maintaining trajectory continuity when vehicles are partially obscured. However, experimental results show that video quality and frame resolution significantly impact system performance.

In test scenarios using the YOLOv11 model with SAHI technology, the reliability score of detection results was relatively high for vehicles clearly visible in still images. However, reliability results were lower, especially for smaller or distant vehicles in test scenarios with collected video sequences. The main factors affecting the experimental results can be mentioned as: slight fluctuations of the UAV camera cause image quality to be blurred (motion blur); UAV changes in altitude during operation; changes in lighting conditions; and background motion of the video affect the results. By dividing the image into overlapping slices, the SAHI strategy helps to improve the relative scale of objects within each slice. While this method does not eliminate the influence of the aforementioned factors, it somewhat minimizes the detection errors related to scale commonly encountered in UAV traffic monitoring.

5. CONCLUSION

This study proposes and evaluates an integrated vehicle detection and tracking framework based on YOLOv11 for ITS applications. The incorporation of the DeepSORT algorithm allows for the maintenance of vehicle identification across consecutive frames, thereby enhancing the stability and consistency of the multi-object tracking system. Simultaneously, SAHI techniques are integrated to improve the detection of small vehicles in remote sensing imagery and traffic surveillance video data collected in Thai Nguyen, Vietnam. Experimental results show that the system achieves high efficiency in both vehicle detection and tracking tasks through evaluation metrics such as mAP, MOTA, and IDF1. However, system performance may still be affected in some cases such as object obscuration, motion blur, or changes in viewing angle. Future work will focus on extending the assessment to larger-scale annotated datasets and incorporating statistical validation on longer video sequences. Additionally, integrating multi-sensor data sources such as radar or light detection and ranging (LiDAR), as well as exploring time-aware learning strategies, is a promising direction for further improving sustainability in complex urban traffic conditions.

ACKNOWLEDGMENTS

The authors would like to thank Thai Nguyen University of Technology (TNUT), Vietnam.

FUNDING INFORMATION

Authors state no external funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Thang C. Vu	✓		✓	✓			✓			✓	✓		✓	✓
Dung T. Nguyen	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Minh T. Nguyen		✓				✓		✓	✓	✓	✓	✓		
Long Q. Dinh					✓		✓			✓		✓		✓
Mui D. Nguyen			✓		✓			✓		✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [MTN], upon request.

REFERENCES

[1] Y. Zhang, Z. Guo, J. Wu, Y. Tian, H. Tang, and X. Guo, "Real-time vehicle detection based on improved YOLO v5," *Sustainability*, vol. 14, no. 19, Sep. 2022, doi: 10.3390/su141912274.

[2] Q. Lin, S. Zhang, and S. Xu, "Construction of traffic moving object detection system based on improved YOLOv5 algorithm," in *2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME)*, Jun. 2023, pp. 268–272, doi: 10.1109/ICDIIME59043.2023.00057.

[3] G. Liu *et al.*, "Smart traffic monitoring system using computer vision and edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12027–12038, Aug. 2022, doi: 10.1109/TITS.2021.3109481.
[4] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.
[5] H. Ghahremannezhad, H. Shi, and C. Liu, "Object detection in traffic videos: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 6780–6799, Jul. 2023, doi: 10.1109/TITS.2023.3258683.
[6] M. Trinh, D. Nguyen, L. Dinh, M. Nguyen, D. Setiadi, and M. Nguyen, "Unmanned aerial vehicles (UAV) networking algorithms: communication, control, and AI-based approaches," *Algorithms*, vol. 18, no. 5, Apr. 2025, doi: 10.3390/a18050244.
[7] M. D. Nguyen and M. T. Nguyen, "Artificial intelligence for human detection, identification and tracking: methods and applications," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 1, pp. 79–94, Apr. 2025, doi: 10.62411/faith.3048-3719-87.
[8] A. Soni and A. Rai, "YOLO for medical object detection (2018–2024)," in *2024 IEEE 3rd International Conference on Electrical Power and Energy Systems (ICEPES)*, Jun. 2024, pp. 1–7, doi: 10.1109/ICEPES60647.2024.10653506.
[9] S. Kang, Z. Hu, L. Liu, K. Zhang, and Z. Cao, "Object detection YOLO algorithms and their industrial applications: overview and comparative analysis," *Electronics*, vol. 14, no. 6, Mar. 2025, doi: 10.3390/electronics14061104.
[10] A. S. Mekled, S. Abdennadher, and O. M. Shehata, "Performance evaluation of YOLO models in varying conditions: a study on object detection and tracking," in *2024 International Conference on Computer and Applications (ICCA)*, Dec. 2024, pp. 1–6, doi: 10.1109/ICCA62237.2024.10927807.
[11] M. T. Nguyen, L. H. Truong, and T. T. H. Le, "Video surveillance processing algorithms utilizing artificial intelligence (AI) for unmanned autonomous vehicles (UAVs)," *MethodsX*, vol. 8, 2021, doi: 10.1016/j.mex.2021.101472.
[12] J. Tang, C. Ye, X. Zhou, and L. Xu, "YOLO-fusion and internet of things: advancing object detection in smart transportation," *Alexandria Engineering Journal*, vol. 107, pp. 1–12, Nov. 2024, doi: 10.1016/j.aej.2024.09.012.
[13] M. T. Nguyen, L. H. Truong, T. T. Tran, and C.-F. Chien, "Artificial intelligence based data processing algorithm for video surveillance to empower Industry 3.5," *Computers & Industrial Engineering*, vol. 148, Oct. 2020, doi: 10.1016/j.cie.2020.106671.
[14] S. Caputo, G. Castellano, F. Greco, C. Mencar, N. Petti, and G. Vessio, "Human detection in drone images using YOLO for search-and-rescue operations," in *AIxIA 2021 – Advances in Artificial Intelligence (AIxIA 2021)*, Cham, Switzerland: Springer, 2022, pp. 326–337, doi: 10.1007/978-3-031-08421-8_22.
[15] Y. Zhang, C. Song, and D. Zhang, "Deep learning-based object detection improvement for tomato disease," *IEEE Access*, vol. 8, pp. 56607–56614, 2020, doi: 10.1109/ACCESS.2020.2982456.
[16] M. Kristo, M. I. -Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE Access*, vol. 8, pp. 125459–125476, 2020, doi: 10.1109/ACCESS.2020.3007481.
[17] C. Chen, C. Wang, B. Liu, C. He, L. Cong, and S. Wan, "Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 13023–13034, Nov. 2023, doi: 10.1109/TITS.2022.3232153.
[18] M. Hussain, "YOLOv1 to v8: unveiling each variant–a comprehensive review of YOLO," *IEEE Access*, vol. 12, pp. 42816–42833, 2024, doi: 10.1109/ACCESS.2024.3378568.
[19] V. G. Dhanya *et al.*, "Deep learning based computer vision approaches for smart agricultural applications," *Artificial Intelligence in Agriculture*, vol. 6, pp. 211–229, 2022, doi: 10.1016/j.aiia.2022.09.007.
[20] H. W. Sorenson, "Least-squares estimation: from Gauss to Kalman," *IEEE Spectrum*, vol. 7, no. 7, pp. 63–68, Jul. 1970, doi: 10.1109/MSPEC.1970.5213471.
[21] M. Avzayesh, M. A. -Hafez, M. AlShabi, and S. A. Gadsden, "The smooth variable structure filter: a comprehensive review," *Digital Signal Processing*, vol. 110, Mar. 2021, doi: 10.1016/j.dsp.2020.102912.
[22] A. Pujara and M. Bhamare, "DeepSORT: real time & multi-object detection and tracking with YOLO and TensorFlow," in *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Nov. 2022, pp. 456–460, doi: 10.1109/ICAISS55157.2022.10011018.
[23] D. S. B. Naik, Y. G. Bala, G. G. Krishna, T. B. Surendhra, and Y. Saketh, "Vehicle counting for traffic analysis: a comparative study using YOLOv5, YOLOv8, and YOLO11," in *Proceedings of the International Conference on AI and Robotics*, 2026, pp. 157–168, doi: 10.1007/978-3-032-05545-3_13.
[24] M. Muzammul, A. Algarni, Y. Y. Ghadi, and M. Assam, "Enhancing UAV aerial image analysis: integrating advanced SAHI techniques with real-time detection models on the VisDrone dataset," *IEEE Access*, vol. 12, pp. 21621–21633, 2024, doi: 10.1109/ACCESS.2024.3363413.
[25] R. Kejriwal, R. H. J., A. Arora, and Mohana, "Vehicle detection and counting using deep learning based YOLO and deep SORT algorithm for urban traffic management system," in *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Feb. 2022, pp. 1–6, doi: 10.1109/ICEEICT53079.2022.9768653.

BIOGRAPHIES OF AUTHORS

Dr. Thang C. Vu received the B.S. and M.Sc. degrees in Electronics and Communications Technology from Hanoi University of Science and Technology in 2006, 2009; and Ph.D. in Telecommunications Engineering from Vietnam Research Institute of Electronics, Informatics, and Automation in 2015. He is currently a lecturer at Thai Nguyen University of Information and Communication Technology. He has published over 30 journal articles. His research interests include wireless communications networks, internet of things, and embedded systems. He can be contacted at email: vcthang@ictu.edu.vn.

Dung T. Nguyen is a lecturer at the Faculty of Engineering and Technology, Information and Communication Technology University (ICTU), Vietnam. He graduated with a bachelor's degree in Electronics and Telecommunications Engineering from ICTU in 2011 and completed his master's degree in Biomedical Engineering from Hanoi University of Science and Technology (HUST) in 2014. He is currently pursuing a Ph.D. in Electronics Engineering at the University of Technology and Education (TNUT). He is a member of the Advanced Wireless Communication Networks (AWCN) laboratory directed by Dr. Minh T. Nguyen. His main research interests include intelligent control and wireless sensors. He can be contacted at email: ntdungcndt@ictu.edu.vn.

Assoc. Prof. Dr. Minh T. Nguyen is currently the director of the Human Resource Training and Development Center at Thai Nguyen University, Vietnam, and also the director of the Advanced Wireless Communication Networks (AWCN) lab. He has interest and expertise in a variety of research topics in the communications, networking, and signal processing areas, especially compressive sensing and wireless/mobile sensor networks. He serves as a technical reviewer for several prestigious journals and international conferences. He also serves as an editor for the Wireless Communication and Mobile Computing journal and an editor-in-chief for ICSES Transactions on Computer Networks and Communications. He can be contacted at email: nguyentuanminh@tnut.edu.vn.

Long Q. Dinh is currently working as a lecturer at the Department of Engineering and Technology, Thai Nguyen University of Information and Communication Technology (ICTU) - Thai Nguyen University. He graduated with a bachelor's degree in Automation in 2011 from Hanoi University of Science and Technology and completed his master's degree in Control and Automation in 2015 from the same university. He is currently pursuing his Ph.D. in Control and Automation Engineering at Thai Nguyen University of Technology - Vietnamese. He is a member of the Advanced Wireless Communication Networks (AWCN) laboratory directed by Dr. Minh T. Nguyen. His main research interests include intelligent control. He can be contacted at email: dqlong@ictu.edu.vn.

Mui D. Nguyen is currently working as a lecturer at the Department of Electrical Engineering, Faculty of International Training, Thai Nguyen University of Technology - Thai Nguyen University. He graduated from university with a major in electronics and telecommunications in 2015 and completed his master's degree in electronic engineering in 2018. Currently, he is pursuing a Ph.D. in Control and Automation Engineering at the same university where he has been attached for many years. He is a member of the Advanced Wireless Communication Networks (AWCN) laboratory. He can be contacted at email: ducmui@tnut.edu.vn.