JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Engineering
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn
ISSN: 2615-9740

An Integrated Approach for Multi-Object Detection and Tracking in Traffic Monitoring Using YOLOv9c and ByteTrack

Nguyen Thi Thanh Thuy, **Ho Van Luc**, **Nguyen Thi Thai An**, **Phung The Bao**, **Nguyen Thi Dinh***

[1]*Ho Chi Minh City University of Industry and Trade, Vietnam*

*Corresponding author. *Email: dinhnt@huit.edu.vn*

ARTICLE INFO

Received:	12/01/2026
Revised:	02/02/2026
Accepted:	11/02/2026
Published:	28/02/2026

KEYWORDS

Object Detection;
Object Tracking;
YOLOv9c;
ByteTrack;
Traffic Monitoring.

ABSTRACT

This paper proposes an integrated method for object detection and tracking in congested traffic environments, based on a combination of the YOLOv9c object detection model and the ByteTrack multi-object tracking algorithm. In this proposed method, the YOLOv9c model is trained and fine-tuned to enhance the performance of vehicle detection in complex conditions. Simultaneously, ByteTrack algorithm links objects across extracted video frames by leveraging both high- and low-confidence bounding boxes. This approach reduces the identity loss and increases the stability of object tracking in traffic, especially in conditions with high object density and severe occlusion. To implement this method, the object detection model was trained and refined on the BDD100K dataset, combined with the Vietnam Traffic Dataset, with a focus on common vehicle classes, including bicycles, motorbikes, cars, buses, and trucks. Experimental results showed that the model achieved a Precision of 89.8% and a Recall of 72.7% in the daytime traffic congestion scenario, and a recall rate of 90.1% in nighttime conditions. For the multi-object tracking problem, the system achieved an IDF1 of 84.3%, demonstrating its ability to maintain stable object identification even in the presence of obstructions, and achieved an MOTA of 69.9% under favorable observation conditions. These results confirm that the proposed method is highly effective in detecting and tracking traffic objects and has potential applications in intelligent traffic monitoring systems and real-time video analysis.

Doi: https://doi.org/10.54644/jte.2026.2077

Copyright © JTE. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purpose, provided the original work is properly cited.

1. Introduction

Multi-object detection and tracking in crowded environments is a crucial problem in computer vision, fundamental to practical applications such as traffic monitoring, urban security, and intelligent transportation systems. In complex conditions, challenges include object obscuration, overlap, shape similarity, and the need for real-time processing [1]. In digitalized cities with smart devices, online monitoring systems are increasingly vital for traffic management, security, and supporting decision-making for authorities. A core challenge is multi-object tracking (MOT) in video data: object detection identifies the location and type of objects in each frame, while MOT maintains object IDs over time [2]. These tasks are closely related and often combined in modern video systems. In heavy urban traffic, MOT is particularly difficult: objects frequently obscure each other, causing incomplete or inaccurate detection; vehicles often share similar shapes and colors, hindering reliable identification; and real-time processing requirements create significant computational pressure.

Many deep learning models have been proposed for this problem, but their effectiveness in real-world conditions, especially urban areas with high traffic, remains limited. Therefore, developing an efficient, stable, and deployable object detection and tracking application is urgent [3]. This report focuses on building such an application for crowded environments that uses the YOLOv9c detection model and the ByteTrack tracking algorithm to improve accuracy and stability in complex scenarios.

The contributions of the paper are: (1) proposing an integrated method for object detection and tracking based on a combination of the YOLOv9c model and ByteTrack, to effectively handle the multi-object tracking problem in crowded traffic scenes; (2) conducting a comprehensive experimental evaluation on standard BDD100K datasets [4] combined with real traffic data from the Vietnam Traffic Dataset [5], thereby demonstrating the ability to improve object identification maintenance and tracking stability compared to traditional methods; (3) building a demonstration application to show the applicability of the proposed method in real-time traffic monitoring scenarios.

2. Related works

In this paper, related works can be divided into two main groups: (1) object detection methods, where the goal is to accurately determine the location and classification of objects appearing in each video frame; (2) multi-object tracking methods, aimed at maintaining the identification of objects over time and effectively handling challenges such as obscurity, overlap, and changes in movement in crowded environments. The study performed a comparative analysis of multiple YOLO variants for vehicle detection in traffic imagery [6]. This study focused on evaluating the performance of YOLOv3, YOLOv5s, and YOLOv5x to determine the most suitable model for a real-world traffic environment, where many complex conditions exist, such as congestion, changing lighting, and diverse vehicle types. The authors used standard evaluation metrics in object detection problems, such as Precision, Recall, and mAP. Experimental results showed that YOLOv5x achieved the highest accuracy among the models surveyed, while YOLOv5s showed a good balance between accuracy and processing speed. In addition, another study proposed an improved YOLOv7-based object detection method to enhance traffic monitoring performance under complex environmental conditions [7]. The method was built by integrating an attention mechanism and optimizing a feature pyramid network, thereby building a model that better exploits contextual information and multi-scale features of objects. The model was evaluated on standard traffic data sets such as KITTI and Cityscapes, including various environmental conditions such as weather changes, lighting, and viewing angles. The results showed that the improved YOLOv7 model achieved higher accuracy than the original YOLOv7 and previous generation YOLO models, especially in detecting small and obscured vehicles. In the field of object detection and traffic prediction for autonomous vehicles, another experiment on [8] review the application of artificial intelligence in autonomous vehicle systems, with a focus on object detection, semantic segmentation, and traffic flow prediction. The authors synthesize a CNN architecture to improve the recognition of vehicles and pedestrians in complex traffic environments. They also discuss models for processing spatial and temporal data, integration of HD maps, big data, and high-performance computing, and suggest future directions for intelligent autonomous vehicles.

Following the object detection phase is real-time object tracking. ByteTrack is a modern multi-object tracking method based on an improved tracking-by-detection strategy, proposed to overcome the limitations in trackers like DeepSORT when faced with detection boxes that have low scores but are actually useful objects for tracking. The study entitled *“ByteTrack: Multi-Object Tracking by Associating Every Detection Box”* experimentally evaluated the ByteTrack method on the MOT17, MOT20, and BDD100K datasets for traffic tracking tasks [9]. The main techniques include two-stage association, Kalman filter, and using the entire detection box to optimize tracking. Experimental results show that ByteTrack outperforms traditional methods in terms of metrics such as MOTA and IDF1, especially in complex video segments with many objects and high levels of occlusion. A recent study aimed at enhancing multi-object tracking performance proposed a framework for livestock tracking by integrating optimization strategies from both DeepSORT and StrongSORT, while refining key components of the tracking-by-detection pipeline [10]. Specifically, the authors used YOLOv9-t as the object detector, then tested and compared five different loss functions for bounding box regression to improve detection accuracy, and adjusted the state vector of the Kalman filter to more accurately model the movement of the livestock object. Subsequently, the distance metric in the re-identification (Re-ID) algorithm was fine-tuned to better distinguish between individuals with similar appearances. Experimental results on real-world datasets show that the proposed method achieved HOTA (78.64%), MOTA (90.29%), and IDF1 (91.41%) while reducing the number of identification changes (IDSW) compared to the basic DeepSORT configuration. These improvements demonstrate that optimizing components within the DeepSORT and StrongSORT frameworks increases track robustness and identification accuracy in complex scenes. In addition, the study entitled *“Observation-Centric SORT:*

Rethinking SORT for Robust Multi-Object Tracking" proposed a novel approach to multi-object tracking by redesigning the conventional SORT architecture from an observation-centric perspective, rather than relying heavily on linear motion models [11]. Unlike SORT and DeepSORT, which use the Kalman filter to predict trajectories, OC-SORT directly exploits information from the nearest observations to estimate implicit velocity, making the system more flexible when the object moves non-linearly or is temporarily obscured. The results Experiment showed that OC-SORT achieved significant improvements in MOTA, IDF1, and HOTA indices compared to SORT and DeepSORT, especially in crowdedand and obscured scenarios. This demonstrates the effectiveness and sustainability of OC-SORT in real-world MOT applications.

3. Proposed methods

The proposed method addresses multi-object detection and tracking in congested traffic by integrating the YOLOv9c detector with ByteTrack in a track-by-detection framework.

3.1. Proposed overall system architecture

Based on the theoretical basis of the YOLOv9c model, ByteTrack and related works show that the problem of detecting and tracking multiple objects in traffic is often approached from a tracking-by-detection perspective, where the detection model directly determines the tracking efficiency. Modern YOLO architectures show a clear advantage in speed and real-time processing capability, while tracking algorithms like ByteTrack help maintain stable object identification in crowdedand obscured scenes [12]. However, many studies still separate the two problems or are not optimized for complex urban traffic conditions in practice. Based on the above analysis, this study designs an integrated object detection and tracking model based on YOLOv9c and ByteTrack, as illustrated in Figure 1. The model is organized into two main phases: the model training phase for detection and the object tracking phase using video, aiming to balance accuracy, tracking stability, and real-time deployment capabilities in a congested traffic environment.

The overall architecture of the proposed model for the object recognition and tracking problem in a crowded environment (Figure 1) is as follows: (1) Normalizing experimental data; (2) Training the object recognition model by using YOLOv9c and fine-tuning; the result is a trained model for the dataset, including BDD100K and Vietnam Traffic Dataset; (3) Using OpenCV to convert the video into frame images; (4) Using the trained model in phase 1 to detect and track objects on the video input; (5) Extracting the object detection and tracking results for the input data.

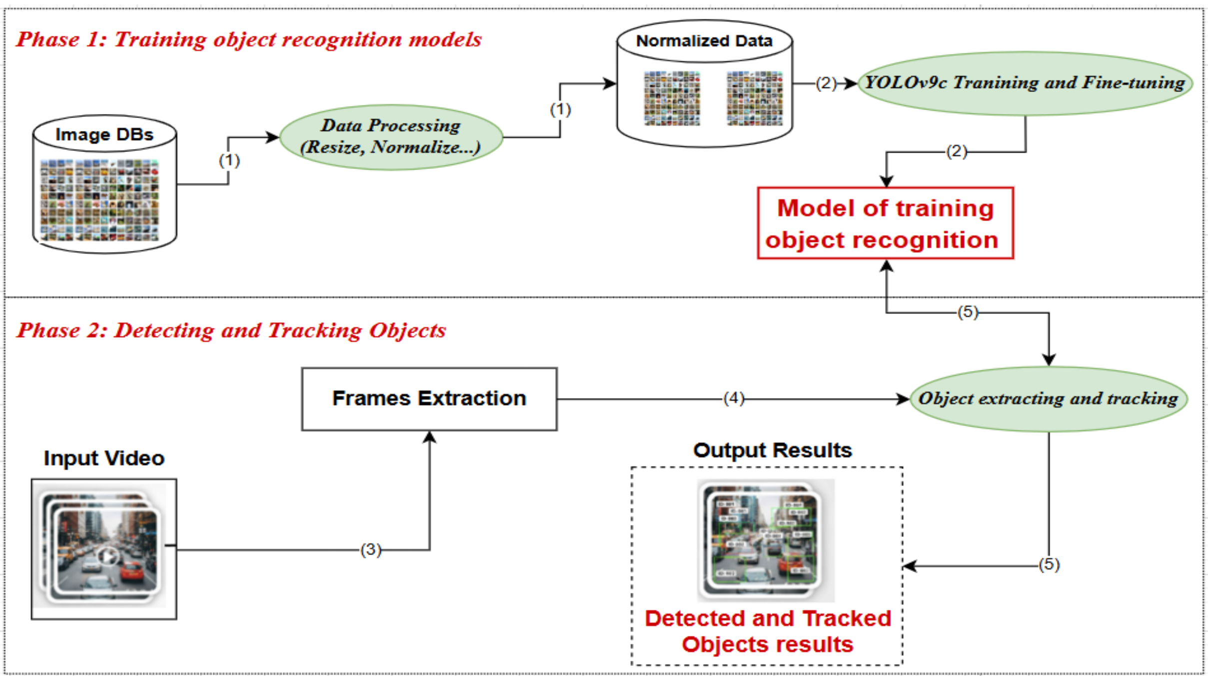

Figure 1. *Proposed model for real-time object detection and tracking application.*

3.2. The YOLOv9c object detection model

In this paper, the YOLOv9c model was used for object detection due to its effective balance between processing speed and accuracy [13]. At the same time, the YOLOv9c model was also fine-tuned on the BDD100K dataset combined with real traffic data in Vietnam to increase its adaptability to diverse environmental conditions and rotation angles in Vietnamese traffic conditions. The dataset used for training and testing includes 7,201 images (BDD100K and Vietnam Traffic Data), focusing on five common vehicle classes: bicycles, motorbikes, cars, buses, and trucks. The images are normalized to a resolution of 640×640 and divided into three sets: training, testing, and evaluation in a 7:2:1 ratio. The training process helps the model improve its ability to accurately detect vehicles in different lighting conditions and high traffic density. The output of the YOLOv9c model at each frame includes bounding boxes, class labels, and corresponding confidence scores for each detected object [13]. The results of training the YOLOv9c model are used for object detection and tracking in Section 3.2.

3.3. The ByteTrack object tracking algorithm

Following the object detection step in section 3.1, the bounding boxes are fed into ByteTrack algorithm to perform multi-object tracking. ByteTrack is a track-by-detection tracking algorithm designed to overcome the limitations of traditional methods that discard low-confidence detections [14]. In this paper, ByteTrack uses a two-stage linking strategy, in which high-confidence bounding boxes are prioritized for primary linking, while low-confidence bounding boxes are retained to support the tracking process. This approach allows for the exploitation of remaining visual information of the object when partially obscured or at a distance, thereby helping to maintain the object's trajectory continuously [15]. During tracking, each object is assigned a unique identifier upon its first appearance. Objects are linked across frames based on predicted location and physical characteristics, and their state is continuously updated to identify their appearance, disappearance, and movement over time. This algorithm significantly reduces ID switching and increases the consistency of the tracking data.

Following the implementation of ByteTrack, the integration process of YOLOv9c and ByteTrack is performed frame by frame. In each frame, the YOLOv9c model performs object detection, and then ByteTrack links and tracks objects across consecutive frames. The system's output consists of uniquely identified bounding boxes, displayed directly on the input video. This allows users to visually observe the object detection and tracking process, while also providing a foundation for traffic analysis and intelligent monitoring in real-world environments.

4. Evaluation of experimental results

4.1. Data and experimental environment

The experimental data in this study consist of two main sources to ensure diversity and closeness to real-world conditions. The first source is the publicly available BDD100K dataset [4]. The second data source is real traffic data in Vietnam [5], collected from traffic surveillance cameras, to reflect the unique characteristics of the domestic traffic environment, such as high vehicle density, diversity of vehicle types, and frequent obstructions. In total, the experimental dataset includes 7,201 images, focusing on five common vehicle classes, including motorbikes, cars, trucks, buses, and other vehicles (Table 1).

Table 1. *Description of experimental data.*

Data sets	No. Images	No. Classes	Training	Testing	Validation
BDD100K [4]	6,805	*5*	5,000	1,000	805
Vietnam Traffic Data [5]	396	5	277	80	39

The experimental setup was implemented on the PyTorch platform, where the YOLOv9c object detection model was trained using supporting libraries such as OpenCV, NumPy, and Pandas for video processing and data preprocessing. The ByteTrack multi-object tracking algorithm was integrated to maintain object identities across video frames. Experiments were conducted on a system equipped with an NVIDIA T4 GPU via Google Colab, with 12–16 GB of RAM, which is sufficient for processing high-resolution traffic videos.

4.2. Experimental results

This paper presents a demo application to illustrate the practical implementation of the proposed method. The application allows users to upload videos, select tracking algorithms, and observe the output, thereby demonstrating the feasibility of the proposed solution in real-world scenarios. Experiments were conducted to evaluate the effectiveness of a multi-object detection and tracking system based on a combination of the YOLOv9c model and ByteTrack in a congested traffic environment. The YOLOv9c detection model was trained and refined on a dataset combining the BDD100K standard dataset and real-world traffic data from Vietnam, with five common vehicle classes.

The results of the YOLOv9c training parameters for object detection are presented in Table 2. The training was conducted for 30 epochs, which is sufficient for convergence on the experimental dataset.

Table 2. *Training parameters for the YOLOv9c model.*

Model	Batch size	Number of epochs	Initial learning rate (lr0)	Weight decay	Momentum
YOLOv9c	16	30	0.005	0.0005	0.937

The results of object tracking parameters using ByteTrack are presented in Table 3.

Table 3. *Experimental parameters with the ByteTrack.*

Model	High confidence threshold	Low confidence threshold	IoU matching threshold	Track buffer (frames)	Object detector
ByteTrack	0.5	0.2	0.8	50	YOLOv9c

Experimental results on the combined datasets including BDD100K and Vietnam Traffic Dataset are presented in Tables 4, 5.

Table 4. *The results show that the YOLOv9c trained model.*

Class of Object	Precision	Recall	F1-Score	mAP@0.5
bicycle	0.825	0.817	0.821	0.829
motorcycle	0.854	0.762	0.805	0.835
car	0.871	0.678	0.762	0.752
truck	0.828	0.676	0.744	0.798
bus	0.68	0.651	0.665	0.686
AVG	**0.812**	**0.717**	**0.759**	**0.780**

Table 5. *The results of object tracking using ByteTrack.*

Video	No. Frames	Precision	Recall	IDF1(%)	MOTP (%)	Feature of Video
MVI_39031	1470	87.2	81.9	**84.3**	18.5	During the day, 2 lanes
MVI_40775	975	75.5	**90.1**	77.3	15.3	Night
MVI_40852	1150	**89.8**	**72.7**	74.8	16.2	During the day, six roads are busy
MVI_40871	1720	83.5	85.8	80.1	13.2	During the day, many lanes and high traffic density

JTE
HCM-UTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Engineering
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

To justify the selection of YOLOv9, we compared it with YOLOv7 and YOLOv8 under identical conditions (Table 6). The experimental results are illustrated in Figures 2, 3 and 4.

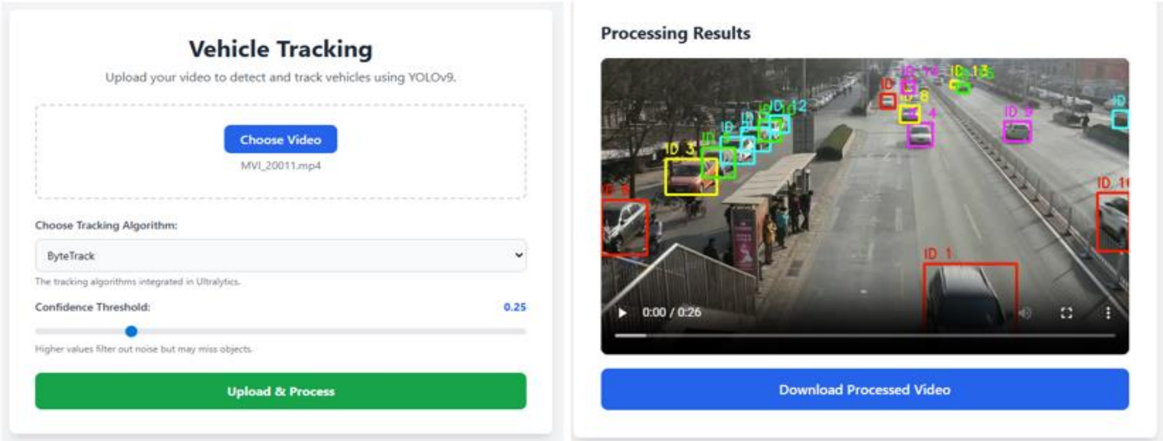

Figure 2. *Experimental application interface.*

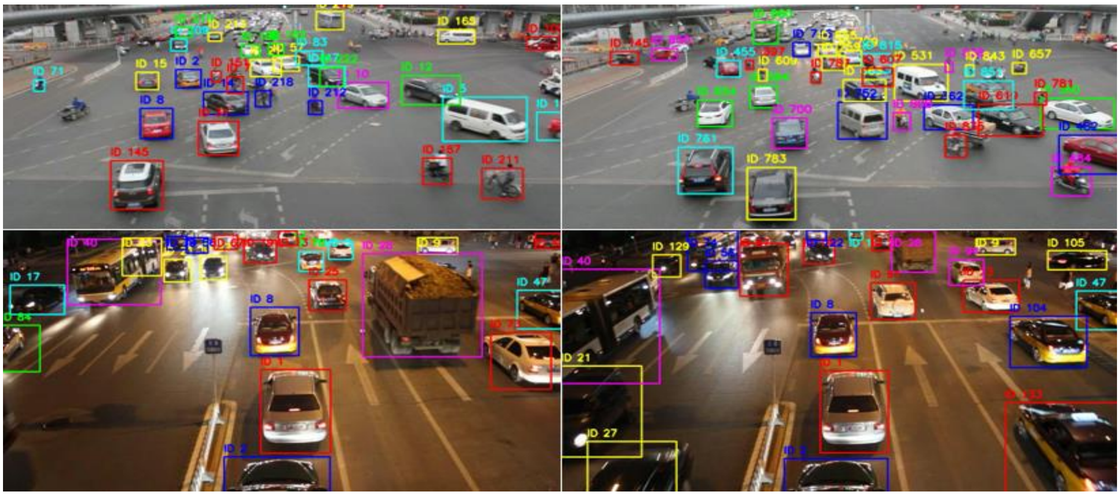

Figure 3. *Results of tracking subjects in real-life traffic videos at Vietnam.*

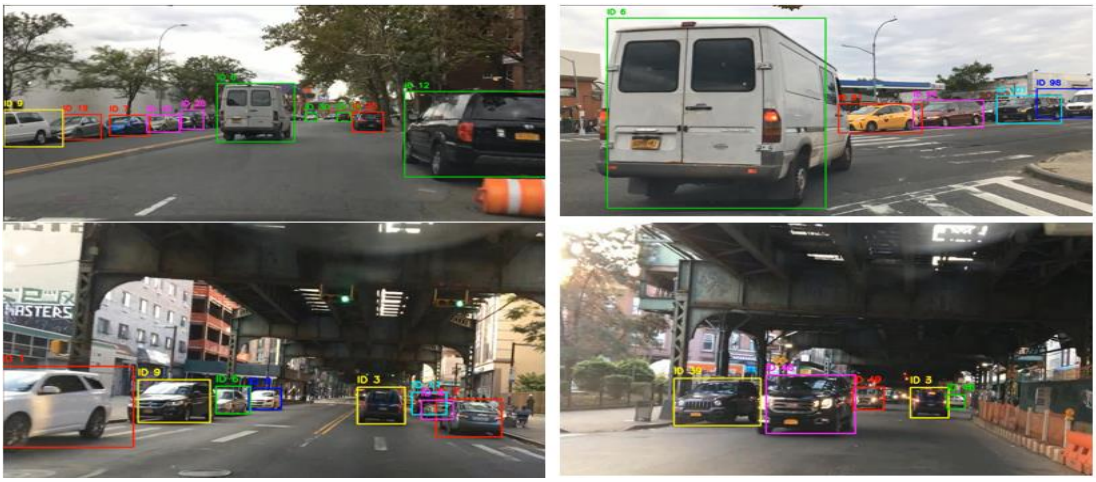

Figure 4. *Results of tracking subjects in Dataset (BDD100K).*

Experimental results show that the YOLOv9c model achieved 89.8% Precision and 72.7% Recall in heavy daytime traffic conditions, and the highest Recall rate of 90.1% in nighttime scenarios, demonstrating good adaptability to changes in lighting conditions and observation environments. These

JTE HCM-UTE
ISSN: 2615-9740
JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Engineering
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

results confirm the effectiveness of model refinement on real-world traffic data. For the multi-object tracking problem, ByteTrack demonstrated the ability to maintain stable object identification in situations with severe obstruction and overlap. The system achieved an IDF1 of up to 84.3%, reflecting a high level of consistency in assigning and maintaining object identification throughout the video sequence. In addition, the MOTA index reached 69.9% under favorable observation conditions, indicating the overall effectiveness of the tracking process.

Table 6. *Comparison of YOLOv7, YOLOv8, and YOLOv9 detection performance*.

Model	Precision	Recall	F1-Score	mAP@0.5
YOLOv7	0.785	0.668	0.722	0.721
YOLOv8	0.803	0.694	0.745	0.756
YOLOv9	**0.812**	**0.717**	**0.759**	**0.780**

Figure 5 shows the changes in the loss functions across both the training and validation sets. The convergence of the loss curves over 30 epochs indicates that the model did not overfit and learned important features from real-world traffic data.

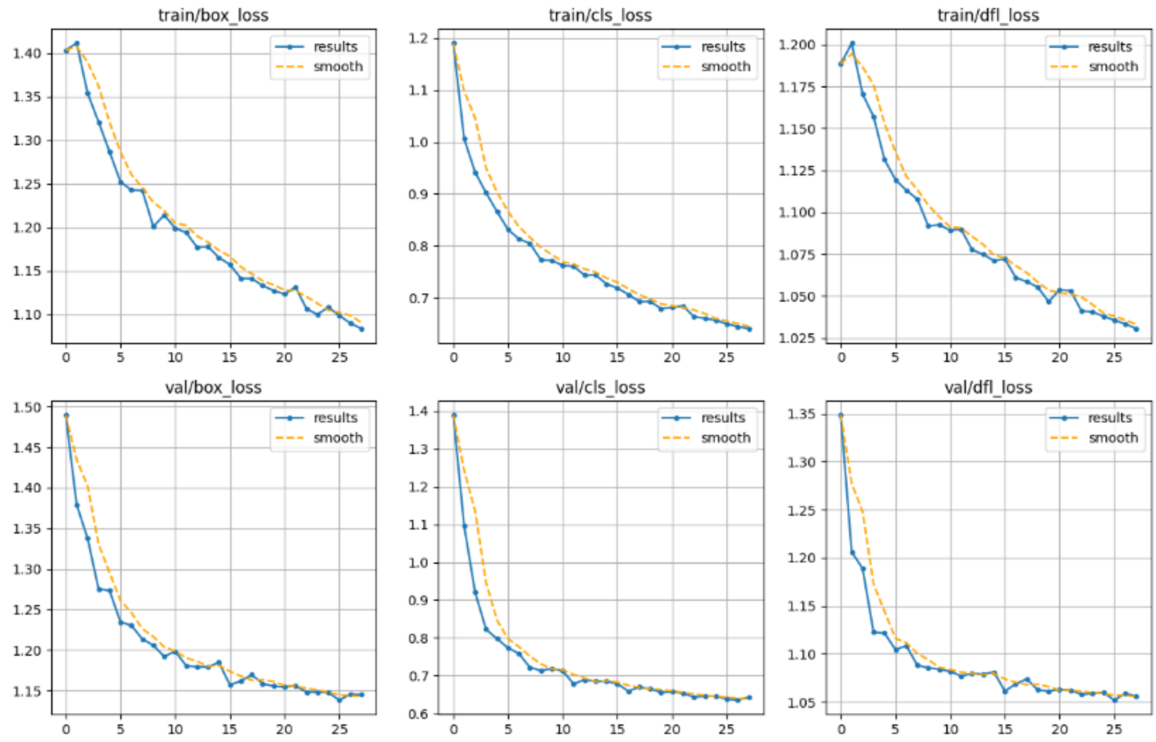

Figure 5. *Results of loss components on training and testing sets*.

Figure 6 illustrates the improvement of evaluation metrics such as Precision, Recall, and mAP over training time. The final results achieved were Precision 89.8% and Recall 72.7% in daytime conditions, with Recall reaching a remarkable 90.1% at night. This confirms the model's ability to strike a good balance between accurate object recognition and minimizing misses in environments with high vehicle density and complex lighting conditions.

Combining the analysis of the gradual decrease in the loss function (Figure 5) and the steady growth of performance indicators (Figure 6) demonstrates the validity of the proposed method before incorporating it into the ByteTrack tracking algorithm.

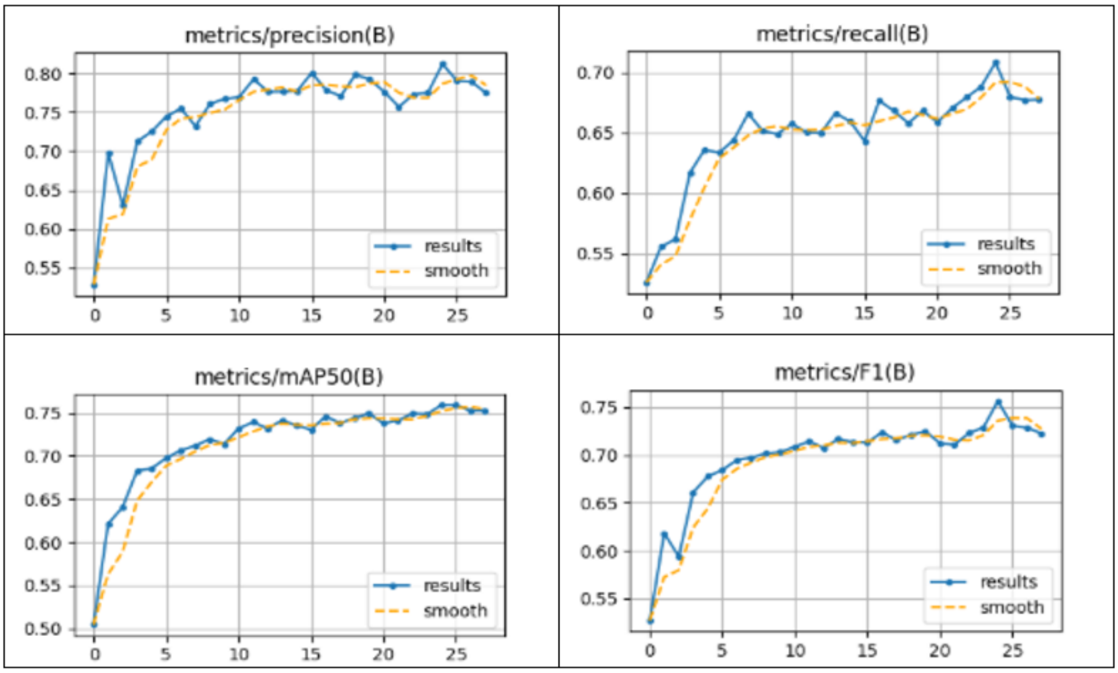

Figure 6. *The graph shows the changes in the performance metrics of the training model.*

4.3. Analyzing, comparing and evaluating experimental results

Experimental results show that the multi-object detection and tracking system based on the combination of YOLOv9c and ByteTrack performs well in congested traffic scenarios. The YOLOv9c detection model demonstrates the ability to accurately identify vehicles in complex environmental conditions, including changes in lighting, rotation angles, and high object density. High Precision and Recall values indicate that the model strikes a relatively good balance between accurate object detection and minimizing omissions, especially in daytime and nighttime scenes.

For multi-object tracking, ByteTrack effectively maintains object identities across video sequences. High IDF1 and MOTA scores indicate stable identity assignment and reduced tracking errors, even under severe occlusion and object overlap, demonstrating the suitability of the proposed method for congested traffic environments.

Table 7. *Comparison of tracking performance between SORT and ByteTrack.*

Tracker	IDF1 (%)	MOTA (%)
SORT	72.62	61.41
ByteTrack	**84.30**	**69.90**

This study focuses on IDF1 and MOTA for identity and tracking accuracy. The IDF1 has higher signals and fewer identity switches with ByteTrack. Table 7 shows that ByteTrack improves IDF1 and MOTA under the same settings, confirming better identity consistency in crowded scenes.

Table 8. *Results of experimental comparison with several other methods.*

Data sets	Method	Detector/Tracker	Evaluation (mAP@0.5, MOTA)
BDD100K	YOLO-based detectors and trackers (ADAS) [16]	YOLOv8-L	mAP@0.5 = 36.66%
Video	Object Tracking Alg Comparison (multitracker) IJSCI [9]	ByteTrack	MOTA = 54.70%

KITTI, EuroCity, BDD100K	YOLOv9-DeepSORT (pedestrian, generic) [17]	YOLOv9 + DeepSORT	MOTA = 68.70%
BDD100K, Vietnam Traffic Dataset (Video)	**Ours**	**YOLOv9c and ByteTrack**	**mAP@0.5 = 78.00% MOTA = 69.90%**

Based on the results in Table 8, the proposed model outperforms the selected comparative studies on the same task. The integration of YOLOv9c with ByteTrack improves detection accuracy and tracking stability, particularly in congested and occluded traffic scenarios. Compared with earlier YOLO versions and conventional tracking methods, the proposed approach achieves a better balance between performance and processing speed, making it suitable for real-time applications. These results highlight the importance of combining modern detection architectures with efficient tracking strategies and confirm the practical applicability of the proposed model.

5. Conclusion and future developments

This paper presents an integrated approach for multi-object detection and tracking in congested traffic by combining the YOLOv9c detector with the ByteTrack tracking algorithm. YOLOv9c ensures high detection accuracy and real-time performance, while ByteTrack improves identity consistency by exploiting both high- and low-confidence detections, making the method effective in high-density and heavily occluded traffic scenes. Experiments on the BDD100K dataset and real-world traffic data from Vietnam demonstrate improved tracking stability and better identity preservation compared to conventional methods, while maintaining real-time processing. In the future work includes integrating ReID or multi-task learning models and expanding datasets to enhance robustness and scalability for large-scale traffic applications.

Acknowledgments

The authors would like to thank the Faculty of Information Technology, Ho Chi Minh City University of Industry and Trade which is sponsor of this research. We also thank anonymous reviewers for their helpful comments on this paper.

Conflict of Interest

The authors declare no conflict of interest in this article.

REFERENCES

[1] D. Stadler and J. Beyerer, "Modeling ambiguous assignments for multi-person tracking in crowds," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2022.

[2] S. E. Bekhouche *et al*., "Driver drowsiness detection in video sequences using hybrid selection of deep features," *Knowledge-Based Systems*, vol. 252, art. no. 109436, 2022, doi: 10.1016/j.knosys.2022.109436.

[3] X. Zhang *et al*., "SkyNet: A hardware-efficient method for object detection and tracking on embedded systems," in *Proc. Conf. Machine Learning and Systems (MLSys)*, 2020.

[4] X. Lin *et al*., "CCTSDB dataset enhancement based on a cross-augmentation method for image datasets," *Intelligent Data Analysis*, vol. 28, no. 5, pp. 1151–1169, 2024, doi: 10.3233/IDA-230075.

[5] VTIS Homepage. Accessed: Feb. 2025. [Online]. Available: https://vtis.vn/

[6] R. M. Alamgir *et al*., "Performance analysis of YOLO-based architectures for vehicle detection from traffic images in Bangladesh," in *Proc. 25th Int. Conf. Computer and Information Technology (ICCIT)*, 2022.

[7] Z. Liang *et al*., "Vehicle and pedestrian detection based on improved YOLOv7-Tiny," *Electronics*, vol. 13, no. 20, art. no. 4010, 2024, doi: 10.3390/electronics13204010.

[8] C. Rana, "Artificial intelligence-based object detection and traffic prediction by autonomous vehicles: A review," *Expert Systems with Applications*, vol. 255, art. no. 124664, 2024, doi: 10.1016/j.eswa.2024.124664.

[9] Y. Zhang *et al*., "ByteTrack: Multi-object tracking by associating every detection box," in *Proc. Eur. Conf. Computer Vision (ECCV)*. Cham, Switzerland: Springer, 2022.

[10] H. S. Sim and H. C. Cho, "Enhanced DeepSORT and StrongSORT for multicattle tracking with optimized detection and re-identification," *IEEE Access*, vol. 13, pp. 19353–19364, 2025, doi: 10.1109/ACCESS.2025.3535092.

[11] J. Cao *et al*., "Observation-centric SORT: Rethinking SORT for robust multi-object tracking," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023.

[12] B. C. Krishna *et al*., "Real-time object detection using YOLOv9c and Flask web application," in *Proc. IEEE Delhi Section Flagship Conf. (DELCON)*, 2024.

[13] H. T. Nguyen *et al*., "YOLOv9c: A robust framework for insect detection," in *Proc. Int. Conf. Multi-disciplinary Trends in Artificial Intelligence*. Cham, Switzerland: Springer, 2024.

[14] Y. Wang and V. Y. Mariano, "A multi-object tracking framework based on YOLOv8s and ByteTrack," *IEEE Access*, vol. 12, pp. 120711–120719, 2024, doi: 10.1109/ACCESS.2024.3450370.

[15] H. Zhao *et al*., "A fish appetite assessment method based on improved ByteTrack and spatiotemporal graph convolutional network," *Biosystems Engineering*, vol. 240, pp. 46–55, 2024, doi: 10.1016/j.biosystemseng.2024.02.011.

[16] P. Azevedo and V. Santos, "Comparative analysis of multiple YOLO-based target detectors and trackers for ADAS in edge devices," *Robotics and Autonomous Systems*, vol. 171, art. no. 104558, 2024, doi: 10.1016/j.robot.2023.104558.

[17] J. Pang *et al*., "Quasi-dense similarity learning for multiple object tracking," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021.

Nguyen Thi Thanh Thuy was born in 1981, graduated with a degree in Information Technology from the University of Science, Vietnam National University Ho Chi Minh City (VNUHCM-US) in 2003. Research fields: Data Mining, Machine Learning and Deep Learning.
Email: thuyntt@huit.edu.vn. ORCID: https://orcid.org/0000-0002-1764-3948

Ho Van Luc was born in 1984, graduated in The University of Science, Viet Nam National University Ho Chi Minh City (VNUHCM-US or HCMUS) in 2010. Research interests: Software Engineering, Information System.
Email: hovanluc@gmail.com. ORCID: https://orcid.org/0009-0000-6096-888X

Nguyen Thi Thai An was born in 1983, graduated in Vietnam National University HCM - University of Information Technology (VNUHCM - UIT) in 2005. Research interests: Machine Learning, Deep Learning.
Email: thaiantl@gmail.com. ORCID: https://orcid.org/0009-0000-6461-4261

Phung The Bao was born in 1985, received the Ph.D. degrees in Information Technology from Irkutsk National Research Technical University (IrNITU), Irkutsk City, Russia, in 2014. He is currently the lecturer of Faculty of Information Technology, Ho Chi Minh City University of Industry and Trade, Vietnam. His research interests include the Mathetical model, applications of artificial intelligence, deep learning, machine learning, data mining.
Email: baopt@huit.edu.vn. ORCID: https://orcid.org/0009-0008-3138-7849

Nguyen Thi Dinh was born in 1983, graduated in Pedagogy Informatics Ho Chi Minh City University of Education in 2006, and received a Master's degree in industry Data transmission and computer network at Ho Chi Minh City Institute of Post and Telecommunications Technology Ho Chi Minh City in 2011. In 2023, she received a PhD degree in Computer Science from the University of Science, Hue, Vietnam. Field research: image processing, image retrieval, and information system.
Email: dinhnt@huit.edu.vn. ORCID: https://orcid.org/0000-0003-3428-3101