Integra: Journal of Integrated Mathematics and Computer Science 3(2) 63-69

Integra: Journal of Integrated Mathematics and Computer Science

Journal Homepage: https://integrajimcs.com/index.php/integra/index

integra
Journal of Integrated Mathematics and Computer Science

Research Paper

Robustness Analysis of YOLOv7 for Motorcycle Traffic Violation Detection Across Multi-Density CCTV Scenarios

Akmal Junaidi[1*], Ridho Sholehurrohman[1], Assyfa Naziwa Ganandy[1], Favorisen R. Lumbanraja[1]

[1] *Department of Computer Science, Faculty of Mathematics and Natural Science, Universitas Lampung, 35145, Indonesia*
[*] **Corresponding author**: akmal.junaidi@fmipa.unila.ac.id

Keywords
YOLOv7, Object Detection, CCTV, Deep Learning, Traffic Violation

Abstract
Traffic violations committed by motorcycle riders remain a significant challenge in Indonesia's transportation sector and contribute to an increased risk of traffic accidents. Common violations include riding without helmets and carrying more than one passenger. This study proposes a motorcycle traffic violation detection approach using the YOLOv7 object detection model applied to CCTV monitoring data. The dataset used in this study was obtained from CCTV recordings at Jalan Gapura DPRD, Bandar Lampung, and consists of images representing three traffic density scenarios: low, medium, and high. The research methodology includes frame extraction, image preprocessing, object labeling, model training, testing, and performance evaluation using standard object detection metrics, namely Precision, Recall, and mean Average Precision (mAP). Experimental results show that the YOLOv7 model achieves consistently high detection performance across different traffic density conditions. The best performance was obtained using a learning rate of 0.01 with 100 training epochs, achieving 99.66% mAP in low traffic conditions, 98.75% in medium traffic conditions, and 97.08% in high traffic conditions. These results indicate that the proposed approach has strong potential for implementation in CCTV-based traffic monitoring systems and may support automated traffic violation detection as part of Electronic Traffic Law Enforcement (ETLE) initiatives.

Received: 25 February 2025, Accepted: 1 July 2026
https://doi.org/10.26554/integra.20263252

1. INTRODUCTION

Motorcycles represent one of the most widely used modes of transportation in Indonesia and play a crucial role in supporting daily mobility. However, the rapid increase in motorcycle usage has also led to a higher incidence of traffic violations, such as riding without helmets and carrying more than one passenger. These violations significantly contribute to traffic accidents and pose serious risks to both riders and other road users [1]. Therefore, the development of an accurate and reliable traffic violation detection system is essential to improve road safety and strengthen law enforcement.

Recent advancements in computer vision and deep learning have opened new opportunities for intelligent traffic monitoring systems. Conventional monitoring methods still rely heavily on manual observation by traffic officers, which is subjective, resource-intensive, and difficult to perform continuously. Although CCTV systems have been widely deployed, manual monitoring of CCTV footage remains inefficient and prone to human error [2]. Consequently, automated detection systems based on deep learning have become an important solution to address these limitations.

In recent years, Convolutional Neural Networks (CNN) have demonstrated strong performance in object detection and classification tasks [3]. Among various object detection algorithms, the You Only Look Once (YOLO) framework has gained significant attention due to its capability for real-time detection with high accuracy and efficiency [4]. Several studies have applied YOLO-based approaches to traffic monitoring, including vehicle detection, violation detection, and intelligent surveillance systems [2, 5, 6, 7, 8, 9, 10, 11, 12]. Furthermore, recent developments such as YOLOv7 have shown improvements in detection accuracy and computational efficiency compared to earlier versions, making it suitable for real-time intelligent transportation systems [13].

Despite these advancements, most existing studies primarily focus on detection accuracy under relatively controlled or homogeneous traffic conditions. Limited attention has been given to evaluating the robustness of detection models under varying real-world traffic densities. In practice, traffic environments are highly dynamic, where factors such as object occlusion, scale variation, and illumination changes can significantly affect detection performance [14]. In particular, the impact of different traffic density levels—low, medium, and high—on model performance remains underexplored, especially when using real CCTV data in urban environments.

To address this gap, this study proposes the application and robustness evaluation of the YOLOv7 model for motorcycle traffic violation detection across multi-density traffic scenarios [15]. The proposed approach utilizes real-world CCTV data collected from urban road environments and evaluates model performance under low, medium, and high traffic density conditions. The evaluation is conducted using standard object detection metrics, including Precision, Recall, and mean Average Precision (mAP), to assess detection accuracy and consistency.

The novelty of this study lies in its systematic evaluation of YOLOv7 performance across varying traffic density scenarios using real CCTV data, which introduces higher visual complexity compared to controlled datasets. In addition, this study provides empirical insights into the robustness and reliability of YOLOv7 in dynamic traffic environments. The main contribution of this research is the development of a practical and scalable traffic violation detection approach that can be integrated into Electronic Traffic Law Enforcement (ETLE) systems, thereby supporting more effective and automated traffic management in Indonesia.

2. METHODS

The research methodology consists of several main stages, such as data collection, image preprocessing, data labeling, YOLOv7 model training, model testing, and performance evaluation. The system process flow is designed to obtain reliable motorcycle traffic violation detection results and is capable of operating in real-time under various traffic density conditions. This multi-stage pipeline follows standard practices in deep learning-based object detection systems to ensure systematic model development and evaluation [4, 14]. The stages in this research include various processes as shown in Figure 1.

2.1 Video Recording

The video data was obtained from CCTV recordings installed on Jalan Gapura DPRD in Bandar Lampung City. Video recordings were taken at different times to represent three conditions, which are low, medium, and high vehicle density. These conditions were chosen to test the consistency and resilience of the model's

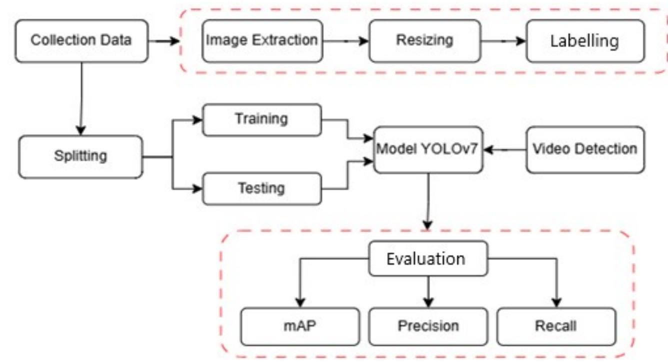

Figure 1. Research Methodology

performance in dealing with variations in the number of objects and levels. The use of real-world CCTV data is important to reflect actual traffic conditions and improve the applicability of intelligent transportation systems in real environments [2]. Each video was further processed into image frames to be used as the primary dataset for model development.

2.2 Pre-processing

During preprocessing, several processes are carried out to improve data quality before model training. The first step is to extract images from video data and save them for further processing. The extracted images are then manually cropped with the vehicle object as the center to ensure that the region of interest is clearly defined. Afterward, the images are resized to 640×640 pixels to match the input size required by the YOLOv7 model [16].

Furthermore, noise in the irrelevant background is reduced to improve image quality. Image denoising is an important step in digital image processing, as noise such as salt-and-pepper, Gaussian, and speckle can degrade image quality and negatively affect feature extraction and detection performance [17]. By reducing noise, the model is better able to focus on relevant visual features, thereby improving detection accuracy. In addition, preprocessing plays a crucial role in improving model performance by standardizing input data and reducing unnecessary variations that may affect the learning process [3]. Proper preprocessing ensures that the model focuses on relevant features, thereby improving detection accuracy and convergence stability.

2.3 Labeling

The target objects in the images are marked with bounding boxes using the Roboflow platform and labeled according to the violation class. Common violations committed by motorcyclists are not wearing a helmet and carrying more than one passenger. Thus, the classes labeled in this study are motorcyclists without helmets (no helmet) and motorcyclists carrying more than one passenger. These labeling results serve as the ground truth that will be used in the model training and testing stages [4]. Figure 2 shows the image labeling process performed by the Roboflow platform with both classes at the same time.

Figure 2. Labeling Process of an Image

2.4 Data Splitting

The labeled dataset is divided into three subsets: training, validation, and testing, with a 80%:10%:10% split. The training data is used to train the YOLOv7 model to learn the visual patterns of traffic violations, while the validation data is used to monitor class loss, objectness loss, and mAP during training to ensure model stability and convergence.

The data test is used to evaluate the final performance of the model on data that was not involved in the training process. The dataset is randomly shuffled while preserving the distribution of violation classes and traffic density variations, so that the model can be evaluated in a representative manner in accordance with the actual conditions of CCTV-based traffic monitoring [18, 19]. This strategy is widely used to avoid data bias and improve model generalization [14].

2.5 Training and Configuration

Model training was conducted using the YOLOv7 architecture on a GPU-based computing environment provided by the Kaggle platform. The training process utilized an NVIDIA GPU (e.g., Tesla T4) to accelerate computation. Several hyperparameter configurations were tested, including variations in learning rate and number of training epochs, to determine the optimal model performance. The training process was conducted using a batch size of 16 and an optimizer such as Stochastic Gradient Descent (SGD) with momentum, which is commonly used in YOLO-based models [4, 13]. During training, performance metrics including class loss, objectness loss, and mean Average Precision (mAP) were continuously monitored to ensure stable convergence and to prevent overfitting or underfitting. The use of hyperparameter tuning is essential to achieve optimal model performance and robustness.

2.6 Testing

Model testing was conducted using test data that was not involved in the training process. Model performance was evaluated by comparing the detection results with the ground truth. The evaluation metrics used included Precision, Recall, and Mean Average Precision (mAP). These metrics are commonly used in object detection tasks to measure detection accuracy, completeness, and overall performance of the model [4, 20, 21, 22]. Precision indicates the correctness of detected objects, Recall measures the ability to detect all relevant objects, and mAP provides a comprehensive evaluation across all classes [20, 23].

2.7 Evaluation

The analysis of the research results was conducted by comparing the performance of the YOLOv7 model in low, medium, and high traffic conditions. The evaluation of the model's performance was based entirely on the mean Average Precision (mAP), Precision, and Recall metrics, which were used to assess the accuracy of detection, the model's ability to recognize all violation objects, and the overall accuracy of the system. The differences in the values of these three metrics were analyzed to understand the effect of traffic density on the performance of motorcycle violation detection. This evaluation approach is important to assess model robustness under varying environmental conditions, particularly in real-world traffic scenarios [14].

Additionally, the visual detection results were analyzed to assess the model's ability to handle variations in traffic conditions and changes in lighting in CCTV images. From the analysis results, the most optimal parameter configuration was obtained to assess the feasibility of applying YOLOv7 to a real-time CCTV-based traffic monitoring system and supporting the implementation of Electronic Traffic Law Enforcement (ETLE).

Figure 3. Detection Results under Low Traffic Density

3. RESULTS AND DISCUSSION

This section presents the experimental results of the YOLOv7 model for motorcycle traffic violation detection under different traffic density conditions. The evaluation is conducted using test data derived from CCTV recordings and focuses on standard object detection metrics, including Precision, Recall, and mean Average Precision (mAP). The methodology described in Section 2 is used as the basis for model development and evaluation.

3.1 Training Performance

The YOLOv7 model was trained using labeled images extracted from CCTV recordings. The training process was conducted with several hyperparameter configurations to determine the optimal model performance. Based on the experimental results, the best performance was achieved using a learning rate of 0.01 with 100 training epochs. The training performance was monitored using class loss, objectness loss, and mean Average Precision (mAP) obtained from the validation dataset.

Model evaluation was conducted under three traffic density conditions: low, medium, and high, in order to analyze the robustness of the model under different levels of visual complexity. The detailed training results of the YOLOv7 model for each traffic density scenario are presented in Table 1.

Table 1. Results of Training with lrf 0.01 at epoch 100

Traffic Density Level	lrf	$Class_{loss}$	Obj_{loss}	mAP
Low		0.00008	0.00214	99.66
Medium	0.01	0.00001	0.00212	98.75
High		0.00219	0.00224	97.08

As shown in Table 1, the model achieves the highest performance under low traffic density, characterized by minimal loss values and maximum mAP. This indicates that the model is able to learn visual patterns effectively when the number of objects in the scene is relatively small. In contrast, increasing traffic density leads to higher loss values and a slight decrease in mAP, indicating increased detection difficulty due to higher visual complexity and object interactions.

Figure 4. Detection Results under Medium Traffic Density

3.2 Result Analysis

The analysis of model performance under different traffic density conditions is supported by visual detection results, as shown in Figures 3-5. The analysis indicates that traffic density significantly affects model performance, particularly in relation to scene complexity and object visibility.

Under low-density conditions, the model achieves near-perfect detection performance due to minimal object overlap and clearer visual features. In such scenarios, each object appears distinctly within the frame, allowing the model to accurately extract spatial and contextual features without interference from surrounding objects. As a result, the detection process becomes more reliable, leading to higher mAP and lower loss values. This is visually confirmed in Figure 3, where objects are clearly separated and detection results appear highly precise.

Figure 5. Detection Results under High Traffic Density

In medium-density conditions, the model maintains stable performance despite increased interactions between objects. Although vehicles begin to appear closer to one another, the model is still able to distinguish object boundaries and identify violation classes with relatively high accuracy. This indicates that the model has learned sufficiently robust feature representations to handle moderate levels of visual complexity. As illustrated in Figure 4, objects begin to overlap but remain distinguishable, and the model continues to detect violations accurately.

However, in high-density conditions, model performance shows a slight decrease. This decline is primarily caused by increased visual complexity, where multiple objects appear simultaneously within a single frame. In such situations, object occlusion becomes more prominent, meaning that important visual features—such as helmets or passenger counts—may be partially hidden by other vehicles. Additionally, overlapping objects reduce the clarity of object boundaries, making it more difficult for the model to accurately localize and classify targets. Variations in object scale further contribute to the reduction in detection performance. These conditions are clearly illustrated in Figure 5, where several objects are partially obscured.

Error analysis further reveals that false positives occur when background elements or partially visible passengers are incorrectly classified as violations, particularly in crowded scenes where visual ambiguity is high. On the other hand, false negatives occur when actual violations are not detected, often due to occlusion or insufficient feature visibility. These errors highlight the limitations of the model in handling complex real-world conditions, especially when key features are not fully observable. Overall, these findings indicate that increasing visual complexity directly impacts detection accuracy. The results demonstrate a clear inverse relationship between model performance and scene complexity, where higher traffic density leads to increased detection challenges. These visual observations further support the quantitative evaluation results presented in the testing results.

3.3 Testing Performance

The YOLOv7 model was evaluated using 150 test images that were not included in the training dataset. These images were extracted from CCTV recordings at the same location but from different frames to ensure independence from the training data. The frames were randomly selected to reduce temporal overlap between the training and testing datasets while maintaining representative traffic conditions. This approach ensures that the model is evaluated on unseen images while preserving the contextual characteristics of real-world CCTV monitoring.

During the testing phase, the trained YOLOv7 model configured with a learning rate of 0.01 and 100 training epochs—was applied to detect motorcycle traffic violations. The model generated bounding boxes around detected objects and classified them according to predefined violation categories, namely motorcyclists without helmets and motorcyclists carrying more than one passenger. The detection outputs were then compared with the corresponding ground-truth annotations to determine prediction accuracy.

The performance of the detection model was evaluated using standard object detection metrics, including Precision, Recall, and mean Average Precision (mAP). Precision measures the accuracy of predicted detections, Recall indicates the ability of the model to detect all relevant violation objects, and mAP provides a comprehensive evaluation of detection performance across all classes. To provide a clearer comparison of model performance across different traffic density conditions, the evaluation results are summarized in Table 2.

Table 2. Testing Performance under Different Traffic Density Conditions

Traffic Density Level	mAP (%)	Precision (%)	Recall (%)
Low	99.66	99.5	100.0
Medium	98.75	98.2	99.4
High	98.10	100.0	97.8

As shown in Table 2, the model achieves the highest performance under low-density conditions, with an mAP of 99.66%, Precision of 99.5%, and Recall of 100.0%. This indicates that the model is highly effective in detecting violations when the visual scene is less complex.

In medium-density conditions, the model maintains strong performance, achieving 98.75% mAP, 98.2% Precision, and 99.4% Recall. Despite increased object interactions, the model remains stable and capable of accurate detection. Under high-density conditions, the model achieves 98.10% mAP, 100.0% Precision, and 97.8% Recall. The slight decrease in Recall indicates that some violation instances are missed due to occlusion and overlapping objects.

To further illustrate the impact of traffic density on model performance, Figure 6 presents the relationship between traffic density and mAP values. The graph shows a gradual decrease in mAP as traffic density increases from low to high. This trend indicates that detection performance is inversely related to scene complexity, where higher object density leads to increased occlusion and overlapping, thereby reducing detection accuracy.

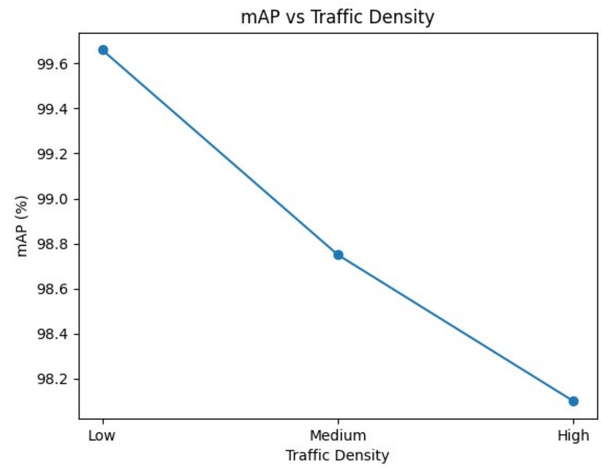

Figure 6. Relationship between Traffic Density and mAP Performance

Taken together, these results indicate that the training configuration using a learning rate of 0.01 and 100 epochs provides stable and consistent performance across different traffic density scenarios. The consistency between training and testing results demonstrates that the model generalizes well within the observed environment. However, the gradual decrease in performance highlights the impact of visual complexity on detection reliability, particularly in real-world scenarios with high object density.

3.4 Discussion

The results demonstrate that the YOLOv7 model provides strong performance in detecting motorcycle traffic violations using CCTV data. The consistency between training and testing results indicates that the selected hyperparameter configuration specifically a learning rate of 0.01 and 100 epochs—ensures stable convergence and reliable detection performance.

From both visual analysis (Section 3.2) and quantitative evaluation (Table 2), it is evident that model performance is strongly influenced by traffic density and scene complexity. As traffic density increases, detection becomes more challenging due to object occlusion, overlapping, and reduced feature visibility. This finding is consistent with previous studies in object detection, where performance degradation is commonly associated with high object density and occlusion effects [14].

Compared to previous YOLO-based studies [2, 5, 9] this study not only achieves high detection accuracy but also provides a structured evaluation of model robustness under varying traffic density conditions using real CCTV data. This constitutes a significant contribution, as many prior works primarily focus on overall accuracy without explicitly addressing performance variability under different real-world conditions.

Despite these promising results, several limitations should be acknowledged. First, the dataset used in this study is relatively limited and collected from a single location, which may restrict the generalization capability of the model across different environments with varying lighting conditions, camera angles, and traffic patterns. Second, the observed performance degradation under high-density conditions highlights the model's sensitivity to occlusion and complex visual scenes.

To address these limitations, future research should focus on expanding dataset diversity, incorporating multi-location CCTV data, and conducting comparative evaluations with other state-of-the-art object detection models. Additionally, advanced techniques such as data augmentation, domain adaptation, and model ensemble strategies may further improve detection robustness.

Taken together, these findings highlight that the YOLOv7 model demonstrates strong potential for implementation in CCTV-based traffic monitoring systems. Its ability to maintain high detection performance across varying traffic conditions supports its applicability in real-world scenarios. Furthermore, the integration of this model into Electronic Traffic Law Enforcement (ETLE) systems can contribute to improving traffic safety, enhancing law enforcement efficiency, and supporting the development of intelligent transportation systems in urban environments.

4. CONCLUSIONS

This study aims to evaluate the performance of the YOLOv7 model in detecting motorcycle traffic violations under different traffic density conditions using CCTV data. Based on the experimental results, it can be concluded that the YOLOv7 model is capable of detecting traffic violations with high accuracy across varying traffic scenarios. The evaluation results show that the model achieves strong performance with mAP values above 98%, accompanied by high Precision and Recall. The optimal model configuration was obtained using a learning rate of 0.01 and 100 training epochs, which provides stable convergence and consistent detection performance. Furthermore, the analysis indicates that traffic density significantly affects model performance. The model performs optimally under low-density conditions, while a slight decrease in performance is observed under high-density conditions due to increased occlusion and object overlap. This confirms that scene complexity is a critical factor influencing detection accuracy. In addition, the model demonstrates the ability to generate continuous detection results from CCTV video streams, indicating its applicability for real-time traffic monitoring systems. Therefore, the proposed approach has strong potential to support the implementation of Electronic Traffic Law Enforcement (ETLE) systems in improving traffic safety and enforcement efficiency. However, this study has several limitations, as the dataset used is relatively limited and collected from a single location, which may affect the generalization capability of the model. Future research is recommended to expand the dataset, incorporate more diverse traffic conditions, and evaluate the model's real-time performance in terms of computational efficiency and detection speed.

5. ACKNOWLEDGEMENT

We would like to express our gratitude to the Bandar Lampung Transportation Agency for providing traffic monitoring video data. We would also like to express our deepest appreciation to the Department of Computer Science, Universitas Lampung, for providing computing facilities through Laboratorium Komputasi Dasar.

REFERENCES

[1] R. Nurfauziah and H. Krisnani. Perilaku pelanggaran lalu lintas. *Jurnal Kolaborasi Resolusi Konflik*, 3(1):75–84, 2021 (in Indonesia).

[2] A. W. Suryanto and A. R. Kardian. Deteksi pelanggaran lalu lintas berbasis cctv. *Jurnal Teknik Komputer*, 2023 (in Indonesia).

[3] A. Santoso and G. Ariyanto. Implementasi deep learning untuk deteksi objek. *EMITOR: Jurnal Teknik Elektro*, 18(1):15–21, 2018 (in Indonesia).

[4] K. Agarwal, A. Sanyo, S. Bakshi, and M. Vinay. Performance evaluation of yolo-based object detection in real-time applications. In *Proceedings of the IEEE International Conference on Computational Intelligence*, 2023.

[5] M. R. Assidiqi. Deteksi pelanggaran lalu lintas menggunakan metode yolov5. Master's thesis, Universitas Lampung, 2023 (in Indonesia).

[6] I. Fazri. Deteksi dan pengenalan jenis pelanggaran lalu lintas menggunakan metode yolo. Master's thesis, Universitas Gadjah Mada, 2022 (in Indonesia).

[7] A. A. Halliwal. Sistem pendeteksi dan penghitung pelanggaran lalu lintas berbasis cctv. Master's thesis, Universitas Lampung, 2023 (in Indonesia).

[8] M. Ilyas. Deteksi pelanggaran berkendara menggunakan metode yolo (you only look once). Master's thesis, Universitas Komputer Indonesia, 2020 (in Indonesia).

[9] E. Maulana, C. Setianingsih, et al. Sistem deteksi pelanggaran lalu lintas berbasis deep learning. *Jurnal Teknik Informatika*, 2023 (in Indonesia).

[10] R. Sholehurrohman and B. Setiyono. Pengembangan yolov3 dengan mobilenetv2 untuk deteksi dan klasifikasi kendaraan bergerak. *Jurnal Ilmiah*, 2021 (in Indonesia).

[11] A. W. Sudjana and H. Supeno. Implementasi deep learning pada rambu lalu lintas di indonesia, 2021 (in Indonesia).

[12] J. Zophie and H. Triharminto. Implementasi artificial intelligence pada pengawasan lalu lintas. *Jurnal Penerbangan dan Antariksa*, 1(1):98–109, 2023 (in Indonesia).

[13] R. Sholehurrohman, K. Muludi, and J. Triloka. Optimizing transportation surveillance with yolov7: Detection and classification of vehicle license plate colors. *Journal of Information Technology and Knowledge*, 10(4):812–819, 2025.

[14] S. Chung, K. Shek, J. Butterfield, and A. Murphy. Current

trends in object detection for autonomous systems. In *Proceedings of IMC37 Conference*, 2021.
[15] R. Sholehurrohman, A. Junaidi, T. N. Dewi, R. Andrian, I. Sabda, and M. R. Habibi. Yolov5s for traffic prohibition sign detection in bandar lampung. *International Journal of Electronic Communication Systems*, 6(1):29–40, 2026.
[16] R. Sholehurrohman, Muhaqiqin, I. S. Ilman, and R. Habibi. Vehicle license plate color detection using yolov5, yolov7, yolov8, and yolov9. *Jurnal Pepadun*, 7(1):67–79, 2026.
[17] R. Sholehurrohman and H. M. Reza. Implementasi restorasi citra derau salt & pepper, gaussian dan speckle secara spasial dengan matlab. *Jurnal INSTEK*, 8(2):369–379, 2023 (in Indonesia).
[18] G. K. N. G., A. Kishore, and A. Krishna. Real-time traffic sign recognition using convolutional neural networks. *Multimedia Tools and Applications*, 84:43119–43154, 2025.
[19] A. Ali, A. Nadeem, N. Zafar, and M. Shiraz. Weather-aware traffic congestion prediction using deep learning. *Frontiers in Communication Networks*, 6:1666487, 2025.
[20] O. Rainio, J. Teuho, and R. Klén. Evaluation metrics for machine learning. *Scientific Reports*, 14, 2024.
[21] V. Hendriko and D. Hermanto. Performance comparison of yolov10, yolov11, and yolov12 on human detection datasets. *Brilliance: Research of Artificial Intelligence*, 5(1), 2025.
[22] R. Padilla, S. L. Netto, and E. A. B. da Silva. Performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020.
[23] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, 10(3):279, 2021.

