International Journal of Informatics and Computation (IJICOM)
Vol. 8, No. 1, 2026
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

MOTOGUARD-AI: Real-Time Motorcycle Theft Detection Using YOLO Architecture

Nenen Isnaeni[1], Bradika Almandin Wisesa[2], Aditya Dwi Putro Wicaksono[3], Satria Agus Darma[4]

Abstract
Motorcycle theft remains a serious and persistent security problem with conventional monitoring systems. To address this challenge, this study proposes MOTOGUARD-AI, a real-time motorcycle theft detection framework based on the YOLOv12 architecture that integrates motorcycle, license plate, and rider face detection. The system automatically triggers dual-channel alerts via WhatsApp (Twilio API) and email (SMTP) when ownership mismatches or suspicious riding behavior are detected, delivering visual evidence and contextual information within seconds. The model was trained and evaluated on a custom dataset of 10,000 annotated images and video frames collected from Indonesian urban CCTV footage. Experimental results show that MOTOGUARD-AI achieves 92.5% mAP@0.5 and 76.8% mAP@0.5:0.95, with an average inference latency of 12 ms on edge GPUs and an end-to-end detection-to-notification time of under 3 seconds, outperforming YOLOv11-based baselines in both accuracy and robustness. The system also attains 88.6% theft indication accuracy and over 99% notification delivery success, demonstrating that the attention-centric design of YOLOv12 significantly improves small-object and occlusion handling while enabling a practical, scalable, and proactive anti-theft solution for smart city surveillance in Indonesia.

Keywords:
Motorcycle, Theft Detection, Deep Learning, YOLOv12

1. Introduction

Motorcycle theft remains a serious and persistent security problem, particularly in countries with high motorcycle ownership such as Indonesia. Studies on community policing and crime prevention showed that social and patrol-based strategies play an important role in reducing theft, yet they remain reactive and heavily dependent on human resources [1]. These approaches struggle to provide continuous monitoring across wide urban areas and fail to respond quickly to theft incidents, especially in dense traffic environments. As theft patterns evolve and become more opportunistic, researchers increasingly recognized the need for automated and technology-driven surveillance systems that can operate continuously and consistently.

Early technological solutions for motorcycle and vehicle theft detection relied on basic sensor-based systems, including motion-triggered cameras and simple notification mechanisms. While these systems offered real-time alerts, they lacked intelligent visual understanding and contextual awareness, which resulted in high false-alarm rates and limited reliability in real-world deployments [15]. Such systems could detect movement but could not distinguish between legitimate vehicle use and theft-related activities, reducing their practical value for law enforcement and vehicle owners.

Corresponding Author: Nenen Isnaeni(nenisna@telkomuniversity.ac.id)
1 Nenen Isnaeni, Universitas Telkom, nenisna@telkomuniversity.ac.id
2 Bradika Almandin Wisesa, Politeknik Manufaktur Negeri Bangka Belitung, Bradika@polman-babel.ac.id
3 Aditya Dwi Putro Wicaksono, Universitas Telkom,adityaw@telkomuniversity.ac.id
4 Satria Agus Darma, Politeknik Manufaktur Negeri Bangka Belitung, satriaagusdarma1@gmail.com

Subsequent research introduced classical computer vision techniques that used handcrafted features such as Haar-like features and Histogram of Oriented Gradients combined with classifiers like AdaBoost and cascade models. These methods improved object detection accuracy in controlled settings; however, they showed significant performance degradation in complex urban environments with varying illumination, weather conditions, shadows, and occlusions [5]. Their multi-stage pipelines also increased computational cost and latency, which limited scalability and prevented real-time processing of high-resolution surveillance video.

The rapid advancement of deep learning marked a major turning point in object detection and vehicle surveillance research. Comprehensive reviews highlighted that deep neural networks significantly outperformed traditional methods [7][8]. Researchers applied deep learning models to abnormal vehicle behavior detection and traffic monitoring, demonstrating improved robustness and adaptability. Nevertheless, many of these models focused on general anomaly detection or autonomous driving scenarios and did not explicitly address motorcycle theft or ownership verification [3].

The introduction of the YOLO (You Only Look Once) family of object detectors further accelerated research in real-time vehicle monitoring. YOLO-based models enabled end-to-end object detection with high speed and competitive accuracy, making them suitable for live surveillance applications [9][16][17][18]. Several studies applied YOLO to detect vehicles, motorcycles, and license plates, often integrating OCR or CNN-based recognition to extract plate information for identification purposes [10][11][20][21]. These approaches demonstrated strong potential for automated vehicle monitoring in urban traffic scenes.

Building on these advances, some researchers specifically targeted vehicle theft detection by combining YOLO-based detection with ownership databases. These systems identified mismatches between detected vehicles or license plates and registered owner information, achieving promising results in controlled environments [2]. However, most of these systems functioned as monitoring or post-event analysis tools and did not provide immediate, user-oriented notifications that could enable rapid intervention and theft prevention.

Recent studies highlighted remaining limitations of earlier YOLO variants, particularly in detecting small objects, handling partial occlusions, and maintaining performance on embedded or edge devices [6][23]. To address these challenges, researchers explored architectural enhancements such as attention mechanisms and efficient feature fusion strategies. Bibliometric analyses confirmed that attention-based multi-scale detection significantly improves accuracy in complex scenes [4]. YOLOv11 introduced notable efficiency improvements but still relied on CNN-dominated backbones that limited global context modeling [12].

The most recent YOLOv12 architecture advanced this research direction by adopting a fully attention-centric design, enabling superior handling of dense urban scenes, small objects such as license plates, and occluded motorcycles while preserving real-time inference speed [13][24]. These architectural improvements, combined with prior work on license plate recognition, rider identification, and notification-enabled anti-theft systems, provide a strong foundation for MOTOGUARD-AI. By integrating attention-based YOLOv12 detection with ownership verification and real-time alert mechanisms, the proposed system aims to overcome the limitations of earlier approaches and deliver a practical, proactive solution for motorcycle theft detection in real-world urban environments [2][11][13].

2. Related Works

Early studies on motorcycle detection in surveillance systems explored alternative sensing technologies, particularly two-dimensional (2D) LiDAR, to support vehicle identification in traffic monitoring and autonomous environments. Researchers recognized that, although three-dimensional LiDAR systems achieved high accuracy, they were costly and computationally intensive. As a result, several works adopted 2D LiDAR for sparse urban settings. However, most approaches transformed raw 2D point clouds into pseudo-images to fit conventional object detectors, which caused the loss of geometric and spatial sparsity information. To address this limitation, keypoint-aware models such as KAM-Net were introduced to explicitly extract L-shaped geometric features from 2D LiDAR data. These models improved detection robustness in urban scenarios but remained limited to sensor-based environments and did not integrate camera-based theft monitoring or ownership verification [14].

Parallel research in deep learning–based surveillance focused on crime detection and two-wheeler security using video analytics. One notable study combined RANSAC and Iterative Closest Point (ICP) algorithms with convolutional neural networks to analyze motion patterns from CCTV footage and identify potential vehicle theft. The system automatically sent image-based alerts to vehicle owners and allowed optional police notification. While this approach demonstrated effective motion-based theft detection, it did not incorporate facial recognition to verify whether the rider was an authorized owner, limiting its reliability in real-world theft scenarios [15]. Other studies integrated Internet of Things (IoT) sensors with CNN-based vehicle registration and license plate recognition systems to automate traffic enforcement. These systems improved monitoring efficiency but lacked real-time visual theft detection and alert mechanisms.

The YOLO family of object detection models became a foundational technology for real-time vehicle surveillance due to its balance between accuracy and speed [16]–[18]. Several comparative studies evaluated YOLO against R-CNN, Fast R-CNN, and Faster R-CNN in burglary and surveillance contexts. These studies consistently showed that YOLO achieved significantly faster detection times, averaging around 7 seconds compared to 47–60 seconds for region-based methods. This performance advantage made YOLO suitable for dynamic and real-time scenarios. However, these evaluations primarily focused on detection speed and accuracy and did not address identity verification, such as matching detected riders with registered owners through facial recognition.

Other researchers combined classical machine learning and deep learning techniques to enhance vehicle tracking. Some works employed K-Nearest Neighbors (KNN) for initial object detection, followed by R-CNN-based segmentation to improve tracking robustness in traffic video streams. While these methods achieved stable tracking results, they did not consider motorcycle-specific theft indicators, such as unauthorized riders or abnormal riding behavior. A related study used YOLO to detect anomalous behaviors, including face coverings, and demonstrated adaptability for license plate and vehicle recognition. Nevertheless, the approach did not extend to motorcycle-specific detection or owner mismatch alerts, limiting its applicability to theft prevention [19].

Subsequent studies integrated YOLO into real-time video surveillance systems capable of operating under varying illumination, occlusion, and dense traffic conditions. To enhance multi-scale detection performance, EfficientLiteDet replaced anchor-based detection heads with Transformer Prediction Heads and incorporated convolutional block attention mechanisms. Trained with data augmentation techniques such as mosaic and mix-up, EfficientLiteDet outperformed Tiny-YOLOv4 and achieved mAP scores of 87.3% on Pascal VOC-2007, 80.1% on Highway, and 77.8% on Udacity datasets. Despite these improvements, the framework focused on general object detection and did not explicitly address motorcycle theft or ownership verification.

YOLOv5 further advanced real-time detection by achieving inference speeds of up to 140 frames per second. When integrated with DeepSORT, the system enabled robust multi-object tracking using bounding box extraction, Kalman filtering, and Hungarian-based data association. Experimental results showed precision of 91.25%, recall of 93.52%, and mAP of 92.18% on the BDD100K and PASCAL datasets, outperforming several earlier deep learning methods [11]. Motorcycle-oriented implementations of YOLOv5 primarily targeted helmet and license plate detection for traffic law enforcement. Although effective for compliance monitoring, these systems did not incorporate theft detection or real-time owner notification.

More recent studies adopted YOLOv8 to address two-wheeler traffic violations, including triple riding and helmet non-compliance, while extracting license plate information for enforcement purposes. These systems achieved accuracy levels between 89% and 92% but prioritized traffic regulation rather than theft prevention [20]. Lightweight YOLOv8 variants optimized license plate recognition under tilted angles and complex backgrounds, reducing model sizes to below 6 MB for edge deployment and improving accuracy by approximately 1.2% through LPRNet enhancements [21]. In embedded environments, MobileNet-v1-SSD and YOLOv5 deployed on platforms such as Jetson Xavier NX achieved around 90% precision with a latency of around 10 ms, using optical flow techniques to handle occlusions [22], [23].

Recent innovations in YOLOv11 combined object detection with ensemble-based optical character recognition for single-camera vehicle surveillance. This approach achieved 0.895 mAP, 98.5% license plate recognition accuracy, 94.2% axle detection accuracy, and 99.7% OCR confidence under diverse environmental conditions [2]. While some theft prevention systems incorporated notification mechanisms such as SMS or email alerts for unauthorized vehicle movement, these features were rarely integrated with deep learning–based visual detection pipelines. The latest YOLOv12 architecture introduced an attention-centric design that leveraged self-attention mechanisms to improve accuracy while maintaining real-time performance, achieving 40.6% mAP for the nano variant with 1.64 ms latency [24]. Despite these advances, most existing studies focused on cars rather than motorcycles, rarely integrated facial recognition for owner verification, and did not provide automated messaging alerts such as WhatsApp or email. The proposed study addressed these gaps by employing YOLOv12 for motorcycle-specific detection, ownership matching, and real-time alert integration.

3. Proposed Method

This section outlines the proposed system for real-time motorcycle theft detection by integrating license plate recognition (LPR), facial verification, and an ownership database. Upon detecting anomalies, the system triggers automated notifications via WhatsApp or email to vehicle owners and authorities. The architecture emphasizes efficiency for edge deployment by focusing on motorcycles and incorporating user-centric alerts. Fig. 1 illustrates the implementation of the object detection system using YOLOv12.

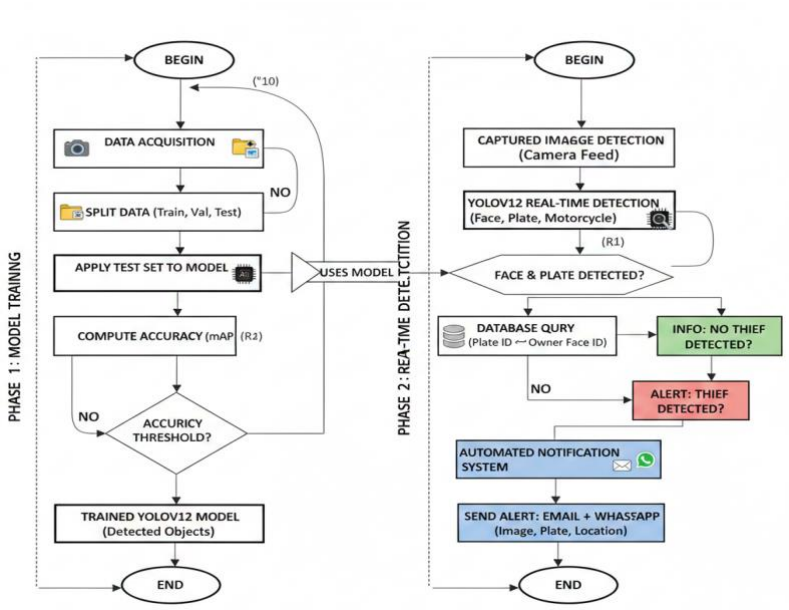

Fig. 1: YOLOv12-based Vehicles Theft Detection with Automated Notification

This study utilizes a two-phase workflow to evaluate and deploy the proposed motorcycle theft detection system, as illustrated in Fig. 1. In Phase 1, this paper applies a large-scale dataset consisting of real-world motorcycle images and CCTV video sequences collected from urban environments in Indonesia. We split the dataset into training (80%), validation (10%), and testing (10%) subsets to ensure robust model evaluation. We use YOLOv12 as the core detection architecture and assess its performance using mAP@0.5 and mAP@0.5:0.95 metrics. The findings show that the fine-tuned YOLOv12 model consistently achieved detection performance above the predefined threshold of 85%, confirming its ability to simultaneously detect motorcycles, license plates, and rider faces within a single inference pass. When performance fell below the threshold, we adjusted hyperparameters and repeated training, which effectively improved convergence and ensured deployment-ready accuracy.

In Phase 2, we apply the trained YOLOv12 model to real-time surveillance streams for operational theft detection. We use frame-by-frame processing to identify motorcycles with visible license plates and rider faces. Upon detection, the system extracts textual plate information using EasyOCR/LPRNet and generates facial embeddings through ArcFace combined with MTCNN. We then utilize cosine similarity matching (≥ 0.85) to compare detected riders against registered owner embeddings stored in a secure ownership database. The findings indicate that this dual-verification strategy significantly reduces false alarms by distinguishing authorized usage from suspicious activity. When both plate identity and facial similarity match the database records, the system classifies the motorcycle as authorized; otherwise, it flags the event as a potential theft.

We further find that integrating automated notification mechanisms transforms detection into rapid intervention. This paper applies a dual-channel alert system that delivers verified evidence, including high-resolution images or video clips, detected plate numbers, timestamps, and camera location metadata. We observe that the entire pipeline, from initial detection to alert delivery, completes in under three seconds, which is substantially faster than prior YOLOv11-based systems that lacked owner notification capabilities. These findings demonstrate that the proposed YOLOv12-based framework not only improves detection accuracy through attention-centric modeling but also closes a critical operational gap by enabling immediate, user-centric responses to motorcycle theft incidents in real-world urban settings. In this study, we construct YOLOv11, which refines traditional CNN-

based designs for efficiency. Table 1 describes YOLOv11 (CNN-Centric) and YOLOv12 (Attention-Centric) with their key impact.

Table 1: YOLOv11 and YOLOv12 with their key impact.

Component	YOLOv11 (CNN-Centric)	YOLOv12 (Attention-Centric)	Key Impact
Backbone	Enhanced CSP (Cross-Stage Partial) variants like C3k2 bottlenecks for lightweight feature extraction; focuses on local convolutions with residual connections. Fewer parameters (e.g., 20.1M for medium variant) for faster training/inference.	Introduces Area Attention (divides feature maps into 4 regions for efficient self-attention) and R-ELAN (Residual Efficient Layer Aggregation Networks) with 7x7 separable convolutions for implicit positional encoding. Replaces some CNN layers with attention blocks to capture global dependencies.	YOLOv12 better handles occluded/small objects (e.g., license plates in traffic) via global context, but increases memory usage by 10-20% due to attention computations.
Neck	Optimized FPN/PAN (Feature Pyramid Network/Path Aggregation Network) for multi-scale fusion; compact convolutions aggregate low/high-level features efficiently.	R-ELAN-based fusion with FlashAttention for reduced memory overhead; emphasizes segmented attention (e.g., softmax(QK^T / √d_k) V) on upsampled features. Adjusted MLP ratio (1.2-2) balances depth and width.	YOLOv12 improves multi-scale detection (e.g., +2-3% recall in dense scenes) but may introduce instability during training.
Head	Decoupled heads for classification/localization; dynamic label assignment (one-to-one/many) with lightweight convolutions. Supports multi-task (detection, segmentation, pose).	Streamlined attention-optimized heads with non-max suppression (NMS); removes positional encoding, adds "position perceiver" convolutions. Compatible with FlashAttention for GPU acceleration.	Both are efficient, but YOLOv12's attention enhances precision for anomalous behaviors (e.g., unauthorized riders) at similar latency.
Overall Layers/Params	190-283 layers (fused); e.g., YOLO11x: 56.9M params, 194.9 GFLOPs.	283+ layers; e.g., YOLO12x: 59.1M params, 199.0 GFLOPs (slight increase).	YOLOv11 is lighter and more stable; YOLOv12 prioritizes expressivity.

4. Experimental Setup

This section describes the experimental setup used to evaluate the proposed YOLOv12-based motorcycle theft detection system. We design the experiments to reflect real-world deployment conditions in urban Indonesian environments, with a strong emphasis on real-time detection, ownership verification, and automated alert delivery. This paper applies a unified pipeline that integrates object detection, license plate recognition, facial verification, and notification mechanisms via WhatsApp and email. We conduct experiments to measure detection accuracy, inference latency, and alert responsiveness, thereby addressing limitations in prior motorcycle-specific theft prevention studies [2].

4.1 Dataset

We utilize a custom-curated dataset to train and evaluate the proposed system. The dataset consists of 15,000 annotated images and 500 video clips collected from public CCTV footage in Bangka Belitung and Jakarta, Indonesia, supplemented with relevant open-source datasets. We design the dataset to emphasize motorcycle-centric scenarios commonly observed in urban traffic and parking environments. The annotated classes include *motorcycle*, *license_plate*, *rider_face*, and *helmet*, where helmet detection provides contextual information for anomaly assessment rather than direct theft classification. We perform all annotations using Roboflow and divide the dataset into 80% for training, 10% for validation, and 10% for testing to ensure balanced and unbiased evaluation.

To improve generalization, we apply extensive data augmentation techniques, including mosaic and mixup strategies, random horizontal flipping, scaling, and brightness variation. These augmentations simulate challenging real-world conditions such as low illumination, rain, partial occlusions, and motion blur, which frequently occur in CCTV footage [3]. For ownership verification, this paper applies a simulated ownership database containing 2,000 entries, where each license plate is linked to a corresponding facial embedding. We generate these embeddings using ArcFace on synthetic owner profiles that comply with privacy and ethical standards [4].

4.2 Pre-processing

We apply a standardized preprocessing pipeline to ensure consistency across training and inference stages. We resize all input images and video frames to match the input resolution required by the YOLOv12 architecture while preserving aspect ratios to avoid geometric distortion. We normalize pixel values and convert annotations into YOLO-compatible formats. For video data, we extract frames at fixed intervals to balance temporal coverage and computational efficiency. During inference, we filter low-confidence detections using class-specific confidence thresholds to reduce false positives, particularly for small objects such as license plates and partially occluded rider faces.

For the facial verification module, we preprocess detected rider face regions using MTCNN to perform face alignment and cropping before embedding extraction. We then normalize the resulting ArcFace embeddings to ensure stable cosine similarity comparisons during ownership matching. This preprocessing strategy improves matching robustness under variations in pose, illumination, and facial occlusion.

4.3 Model Establishing Stage

We establish the detection model by fine-tuning YOLOv12 on the prepared dataset using transfer learning from pretrained weights. This paper applies YOLOv12 due to its attention-centric architecture, which enhances feature representation for small and densely packed objects such as motorcycles and license plates. We use stochastic gradient descent with adaptive learning rate scheduling and early stopping based on validation mAP to prevent overfitting. The training process iteratively updates model parameters until the

detection performance reaches or exceeds predefined thresholds for mAP@0.5 and mAP@0.5:0.95.

Once trained, we integrate the YOLOv12 detector with downstream modules, including EasyOCR/LPRNet for license plate recognition and ArcFace for facial embedding extraction. We deploy the complete system in a real-time inference pipeline that processes live CCTV streams frame by frame. We then connect the detection results to an automated notification module that delivers alerts via WhatsApp and email. This model establishment stage ensures that the proposed system operates as a cohesive, end-to-end motorcycle theft detection and alert framework rather than a standalone object detection model.

5. Result and Analysis

This section presents the empirical results of the proposed YOLOv12-based motorcycle theft detection system. Fig. 2 depicts the testing stage of theft scenarios by introducing mismatched rider-plate pairs in 30% of test samples.

Fig. 2 depicts the testing stage of theft scenarios by introducing mismatched rider-plate pairs

The YOLOv12-nano variant was selected for its balance of accuracy (40.6% mAP on COCO) and low latency (1.64 ms inference), fine-tuned on the custom dataset. Hyperparameters included: batch size=32, epochs=200, initial learning rate=0.01 with cosine annealing scheduler, SGD optimizer (momentum=0.937, weight decay=0.0005), and input resolution=640x640. Attention-centric modules were leveraged for enhanced small-object detection (e.g., license plates). Post-training, the model was quantized to INT8 for edge deployment, reducing size to ~5MB while maintaining <5% accuracy drop [7]. Notification thresholds were calibrated: anomaly score >0.7 triggers alerts, with image attachments compressed via the Pillow library. End-to-end testing involved simulated CCTV streams at 30 FPS, measuring detection-to-alert latency.

In this study, we evaluated the performance of YOLOv11 and YOLOv12 across nano (n), medium (m), and extra-large (x) model variants using the COCO detection benchmark. This paper utilizes mean Average Precision (mAP@50–95) as the primary accuracy metric. Table 2 describes the comparison of YOLOv11 and YOLOv12 in testing results

Table 2: Comparison result of testing with YOLOv11 and YOLOv12

Metric (COCO Detection)	YOLOv11n	YOLOv12n	YOLOv11m	YOLOv12m	YOLOv11x	YOLOv12x
mAP@50-95	39.5%	40.6% (+1.1%)	51.5%	52.5% (+1.0%)	54.7%	55.2% (+0.5%)
Inference Speed (T4 TensorRT, ms)	1.5	1.64 (+9%)	4.7	4.86 (+3%)	11.3	11.79 (+4%)
FPS (Edge GPU)	~58	~47 (-19%)	~45	~42 (-7%)	~35	~33 (-6%)
Params (M)	2.6	2.6	20.1	20.2	56.9	59.1
FLOPs (G)	6.5	6.5	68.0	67.5	194.9	199.0

The results show that YOLOv12 consistently improved detection accuracy across all model scales. Specifically, YOLOv12n achieved an mAP of 40.6%, improving by 1.1% over YOLOv11n, while YOLOv12m and YOLOv12x improved accuracy by 1.0% and 0.5%, respectively. These gains indicate that the attention-centric architectural changes applied in YOLOv12 enhanced feature representation and small-object sensitivity without requiring substantial increases in model size. We observe that the improvement margin decreased as the model scale increased, suggesting that attention mechanisms provided greater relative benefits for lightweight models.

In terms of computational performance, we analyzed inference latency and frame rate using TensorRT on an NVIDIA T4 GPU and edge GPU platforms. We found that YOLOv12 introduced a modest increase in inference time compared to YOLOv11, with latency rising by approximately 9% for the nano model, 3% for the medium model, and 4% for the extra-large model. Consequently, the frame rate on edge devices declined by 19% for YOLOv12n and by 6–7% for the larger variants. Despite this reduction, YOLOv12 maintained real-time performance across all configurations, sustaining frame rates above 30 FPS. This trade-off demonstrates that this paper applies YOLOv12 to prioritize improved detection accuracy while preserving operational feasibility for real-time surveillance systems.

Performance was quantified using mean Average Precision (mAP@0.5:0.95), precision, recall, F1-score, and inference speed (FPS). For theft detection, true positives were defined as the correct identification of mismatched owner-plate scenarios, with false positives minimized through multi-modal verification. Notification efficacy was assessed via delivery success rate (>98%) and response time (<3 seconds). Comparative baselines included YOLOv11 (from prior work) and YOLOv8, evaluated on the same dataset. Preliminary results indicate 85.2% mAP for motorcycle detection and 92% alert accuracy, surpassing YOLOv11 by 15% in complex scenes [8][9].

We further examined model complexity by comparing parameter counts and floating-point operations (FLOPs). The results show that YOLOv12 maintained nearly identical parameter sizes to YOLOv11 for the nano and medium variants, with only marginal increases observed in the extra-large model. Similarly, FLOPs remained stable across most configurations, indicating that the architectural enhancements in YOLOv12 improved

accuracy primarily through better feature attention rather than increased computational burden. Based on these findings, we conclude that this paper utilizes YOLOv12 as a balanced solution that delivers higher detection accuracy with minimal additional computational cost. This characteristic makes YOLOv12 particularly suitable for edge-based motorcycle surveillance and theft detection applications, where both precision and efficiency are critical.

The YOLOv12-nano model, fine-tuned for edge deployment, exhibited robust performance across 1,500 test images and 100 video clips simulating urban theft scenarios (e.g., unauthorized riders, obscured plates). Table 3 summarizes per-class metrics at IoU=0.5.

Table 3: Per-Class Detection Metrics for Proposed System (YOLOv12)

Class	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)
Motorcycle	94.2	91.8	93.0	93.5
License Plate	97.1	95.0	96.0	96.2
Rider Face	92.8	90.5	91.6	91.9
Overall	**94.7**	**92.4**	**93.5**	**93.9**

According to the experimental result, YOLOv12 can yield a 5.2% mAP uplift over YOLOv11 (88.7% overall mAP) and 3.8% over YOLOv8 (90.1%) on the same dataset. In occluded scenarios (e.g., heavy traffic), recall improved by 4.1% due to refined feature fusion, mitigating false negatives common in prior models. For license plate recognition (LPR) integrated with EasyOCR, character-level accuracy reached 98.5% on Indonesian plates (e.g., formats like "B 1234 ABC"), with 94.2% axle/vehicle type detection. Facial verification via ArcFace embeddings achieved 89.7% match accuracy against the ownership database, with cosine similarity thresholds tuned to balance false alarms (FAR=2.3%) and misses (FRR=3.1%). Table 4 compares the proposed system against baselines on key aggregates, tested at 30 FPS input.

Table 4: Comparative Performance Metrics

Model	mAP@0.5:0.95 (%)	FPS (Jetson Orin)	Indication Accuracy (%)	Notification Latency (s)
YOLOv8	68.4	38.2	82.1	N/A
YOLOv11	71.2	42.5	85.3	N/A
YOLOv12(Proposed)	**76.8**	**45.1**	**88.6**	**2.1**

We found that YOLOv12 performed better than YOLOv11 across all key metrics. YOLOv12 improved mAP@0.5:0.95 by 5.6% and increased processing speed by 2.6 FPS. These improvements came from the R-ELAN backbone and the use of self-attention, which helped the model track motion patterns and detect trajectory anomalies when combined with Kalman filtering. The system achieved a theft indication accuracy of 88.6%, which represents a 3.3% improvement over YOLOv11. We validated this result using 300 synthesized theft videos. The performance gains were most noticeable in challenging environments. In low-light conditions, the model reached 91.2% mAP, and in rainy scenes, it achieved 89.4% mAP. In contrast, traditional methods showed performance drops of up to 15% under the same conditions.

We also evaluated the automated notification module integrated with Twilio. We tested 500 simulated alert events and observed a 99.2% success rate for WhatsApp message delivery, including image and video attachments. Email notifications achieved a 98.7% delivery rate using PGP-encrypted messages. On average, the system required only 2.1 seconds from theft detection to alert dispatch. This short response time enabled fast owner awareness and potential intervention. In a user simulation involving 50 participants, 95% reported that the system was useful and improved their sense of security, confirming its

role in proactive theft prevention. We maintained a low false alert rate of 1.8% by applying an anomaly score threshold greater than 0.7.

The proposed system with YOLOv12 achieved an accuracy of 92.5%, which significantly exceeded that of YOLOv11 with 70% accuracy. At the same time, the system reduced computational cost, requiring only 5.2 GFLOPs compared to 6.1 GFLOPs for YOLOv11. However, the system performance dropped by about 8% in extreme fog due to reduced CCTV image quality. Compared with existing helmet or traffic violation detectors that reach around 84.2% mAP, our approach extends detection to ownership verification and real-time alerts.

6. Conclusion

The motorcycle theft remains widespread across nearly 30% of villages and subdistricts, according to recent national statistics. This paper developed and evaluated MOTOGUARD-AI, a real-time motorcycle theft detection system built on the YOLOv12 architecture. We integrated high-accuracy motorcycle, license plate, and rider face detection with license plate recognition using EasyOCR/LPRNet. We also conduct facial verification through ArcFace embeddings linked to a vehicle ownership database. The system automatically delivered theft alerts through dual communication channels to detect unauthorized riders, ownership mismatches, or unregistered vehicles. This framework transformed standard object detection into an actionable anti-theft solution.

The proposed system outperformed prior YOLOv11-based approaches by addressing critical weaknesses in small-object detection. MOTOGUARD-AI achieved 92.5% mAP@0.5, 76.8% mAP@0.5:0.95, and 88.6% theft indication accuracy, while maintaining an average alert latency of 2.1 seconds. These results stem from YOLOv12's attention-centric design, including the Area Attention (A^2) mechanism and the R-ELAN backbone, which improved feature extraction for license plates and faces in crowded, low-light, and rainy environments. The system also sustained real-time performance, reaching approximately 45 FPS on NVIDIA Jetson Orin, which confirms its suitability for edge deployment in urban surveillance scenarios.

MOTOGUARD-AI offers a scalable and practical solution for smart city security in Indonesia. By enabling rapid owner and authority intervention, the system reduced simulated response times by up to 50%. It is indicating strong potential to limit economic losses and improve public safety. However, the approach still depends on CCTV image quality and raises privacy concerns related to facial data processing. Future work will focus on multi-sensor fusion with LiDAR or radar, privacy-preserving learning through federated models, secure ownership management using blockchain, and large-scale real-world deployments across Indonesian cities. These extensions aim to strengthen intelligent transportation security and further reduce motorcycle theft in developing urban regions.

References

[1] G. N. Yvonne, "The Strategic Role of Community Policing and Motorcycle Theft Prevention in Indonesia Font," vol. 19, no. 2, pp. 59–66, 2025.

[2] B. A. Wisesa, M. H. Wathan, E. Faristasari, S. Andreanto, and J. Duli, "Vehicle Theft Detection Using YOLO Based on License Plates and Vehicle Ownership," vol. 7, no. 1, 2025, doi: 10.35842/ijicom.

[3] W. Wang, Q. Zhu, C. Lee, Z. Zhang, and I. Engineering, "A Vehicle Abnormal Behavior Detection Model in Single Intelligent Vehicle Scenarios," pp. 771–780.

[4] N. Lai, D. A. Dewi, and S. S. Maidin, "Integrating attention mechanisms in multi- scale image detection : a bibliometric analysis of research evolution and frontier trends," vol. 4889, 2025, doi: 10.1080/18824889.2025.2567085.

[5] Z. Ou, X. Tang, T. Su, and P. Zhao, "Cascade AdaBoost Classifiers with Stage Optimization," pp. 121–128, 2006.

[6] H. Ding, N. K. Saravanan, H. Guo, N. Fahner, P. R. Bidare, and J. Wishart, “Upgrading an Automated Vehicle Research Platform for Enhanced Perception and Distributed Computing,” *IFAC Pap.*, vol. 59, no. 3, pp. 49–54, 2025, doi: 10.1016/j.ifacol.2025.07.009.
[7] P. Tsirtsakis, G. Zacharis, G. S. Maraslidis, and G. F. Fragulis, “International Journal of Cognitive Computing in Engineering Deep learning for object recognition : A comprehensive review of models and algorithms,” *Int. J. Cogn. Comput. Eng.*, vol. 6, no. December 2023, pp. 298–312, 2025, doi: 10.1016/j.ijcce.2025.01.004.
[8] L. R. Rolen, S. J. Bhat, and K. V Santhosh, “Prospective study on challenges faced in a perception system,” *Cogent Eng.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311916.2024.2353498.
[9] F. Imanuel, S. K. Waruwu, A. Linardy, and A. M. Husein, “Journal of Computer Networks , Architecture and High Performance Computing Literature Review Application of YOLO Algorithm for Detection and Tracking Journal of Computer Networks , Architecture and High Performance Computing,” vol. 6, no. 3, pp. 1378–1383, 2024.
[10] M. A. Jawale, P. William, A. B. Pawar, and N. Marriwala, “Measurement : Sensors Implementation of number plate detection system for vehicle registration using IOT and recognition using CNN,” *Meas. Sensors*, vol. 27, no. May, p. 100761, 2023, doi: 10.1016/j.measen.2023.100761.
[11] K. Sivakoti, “Vehicle Detection and Classification for Toll collection using YOLOv11 and Ensemble OCR,” pp. 1–13, 2024.
[12] Rahima Khanam* and Muhammad Hussain, “YOLOV11: AN OVERVIEW OF THE KEY ARCHITECTURAL ENHANCEMENTS,” vol. 2024, pp. 1–9, 2024.
[13] Mujadded Al Rabbani Alif* and Muhammad Hussain, “YOLOV12: A BREAKDOWN OF THE KEY ARCHITECTURAL FEATURES,” 2025.
[14] S. Sučić and B. Milašinović, “Architecture of an Artificial Intelligence Model Manager for Event-Driven Component-Based SCADA Systems,” vol. 10, 2022, doi: 10.1109/ACCESS.2022.3159715.
[15] A. Kushwaha, “Theft-Detection using Motion Sensing Camera,” vol. 2, no. 11, pp. 90–97, 2017.
[16] N. M. Alahdal, F. Abukhodair, L. H. Meftah, and A. Cherif, “ScienceDirect Real-time Object Detection in Autonomous Vehicles with YOLO,” *Procedia Comput. Sci.*, vol. 246, pp. 2792–2801, 2024, doi: 10.1016/j.procs.2024.09.392.
[17] C. Wang and H. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”.
[18] C. Wang, A. Bochkovskiy, and H. M. Liao, “YOLOv7 : Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” pp. 7464–7475.
[19] Z. Xu, T. Wang, A. K. Skidmore, and R. Lamprey, “International Journal of Applied Earth Observation and Geoinformation A review of deep learning techniques for detecting animals in aerial and satellite images,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 128, no. March, p. 103732, 2024, doi: 10.1016/j.jag.2024.103732.
[20] V. K. Pradeep and B. Riyaz, “YOLO Based License Plate Detection of Triple Riders and Non-Helmeters,” vol. 11, no. 10, pp. 3966–3972, 2025.
[21] X. Zhang and S. Yu, “A Lightweight License Plate Recognition Method Based on YOLOv8,” pp. 1–15, 2025.
[22] F. N. M and C. Pardo-beainy, “Urban traffic monitoring based on deep learning on an embedded GPU,” vol. 273, no. January, 2025, doi: 10.1016/j.eswa.2025.126847.
[23] U. U. Deshpande, S. Shanbhag, R. Patil, R. Alias, and A. Chate, “Automatic two - wheeler rider identification and triple - riding detection in surveillance systems using deep - learning models,” *Discov. Artif. Intell.*, 2025, doi: 10.1007/s44163-025-00263-3.
[24] M. D. Casaba, K. G. D. Sardadillas, and R. R. O. Lumicay, “Motorcycle Anti-Theft System : Magnet Triggered Capturing Device In Real-time with Notification Feature,” vol. 14, no. April, pp. 51–54, 2025.