Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

Motorcycle Rider Face Detection: An Experiment with R-CNN and Faster R-CNN

Muhaqiqin[1*]✉, Ridho Sholehurrohman[1]✉, Agung Pambudi[1]✉, Wartariyus[1]✉, Igit Sabda Ilman[1]✉, Muhammad Afdhaluddin[1]✉, Zulkarnaini[2]✉
[1]Department of Computer Science, Lampung University, Indonesia
[2]Faculty of Computer Science, Darmajaya Institute of Informatics and Business, Indonesia

Abstract
Motorcycle rider face detection remains a challenging task in intelligent transportation monitoring due to partial facial occlusion caused by helmets, visors, face masks, and variations in image orientation. This study aims to experimentally compare the performance of Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN in detecting and classifying motorcycle rider facial regions. A dataset comprising 2,000 motorcycle rider images was collected and divided into training, validation, and testing subsets using a ratio of 70:10:20. Rotation-based data augmentation at +10° and −10° increased the training dataset to 4,200 images. Both architectures were evaluated using the same dataset partitions with learning rates of 0.01 and 0.001 and a fixed training duration of 80 epochs. The experimental evaluation included mAP50, classification loss, localization loss, accuracy, precision, recall, and F1-score. The experimental results showed that Faster R-CNN achieved a higher mean Average Precision at an IoU threshold of 0.50 (mAP50) of 0.8467 compared with 0.7824 for R-CNN at a learning rate of 0.01. Furthermore, Faster R-CNN achieved an accuracy of 89.42%, precision of 96.35%, recall of 91.56%, and F1-score of 93.89%, outperforming R-CNN, which obtained an accuracy of 80.17%, precision of 85.66%, recall of 84.15%, and F1-score of 84.90%. These findings demonstrate that Faster R-CNN achieved better overall detection and classification performance than conventional R-CNN under the evaluated experimental conditions. Although the evaluation was limited to a single-location dataset collected under favorable daytime lighting conditions, the findings provide empirical evidence of the applicability of Faster R-CNN for motorcycle rider face detection in intelligent transportation monitoring systems.
Keywords: Motorcycle Rider Face Detection; R-CNN; Faster R-CNN; Computer Vision; Object Detection.

***Corresponding Author:**
Muhaqiqin, Lampung University, Indonesia
Email: muhaqiqin@fmipa.unila.ac.id

1. INTRODUCTION

The rapid advancement of computer vision and deep learning has significantly contributed to the development of intelligent monitoring systems in transportation, public security, and automated visual surveillance. Object detection plays an essential role in these systems by enabling the identification and localization of objects within digital images through visual feature analysis and spatial prediction [1], [2]. In transportation environments, the detection of motorcycle rider faces has become an important research area due to the increasing need for automated monitoring technologies. However, motorcycle riders exhibit considerable variations in appearance, particularly when wearing protective equipment such as helmets, visors, and face masks. These accessories may partially obscure facial features, making accurate face detection more challenging than in controlled imaging environments. Previous studies have identified facial occlusion, pose variation, and environmental conditions as important factors influencing the reliability of automated facial analysis systems [3]–[5]. Therefore, developing reliable detection approaches capable of locating motorcycle rider facial regions under varying visibility conditions remains an important challenge in computer vision applications.

Several studies have investigated the application of deep learning architectures to facial analysis and object detection. Megawan and Lestari [6] implemented Faster R-CNN with a ResNet-50 architecture

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

for face spoofing detection and reported an average training accuracy of 97.07%. Megawan et al. [7] subsequently developed a real-time face spoofing and non-spoofing detection system using Faster R-CNN on Raspberry Pi, achieving detection accuracies of 96% and 92%, respectively. Sari [8] applied a Convolutional Neural Network (CNN) to classify gender using facial images and obtained a maximum accuracy of 91.02%. Meanwhile, Alamsyah and Pratama [9] investigated fingertip detection using Faster R-CNN with an Inception V2 architecture and reported an accuracy of 91% under noise-free image conditions. These findings demonstrate the applicability of convolutional architectures to various visual recognition tasks. Nevertheless, the experimental conditions and detection targets investigated in these studies differ from motorcycle rider facial regions, where the presence of helmets, visors, and masks introduces additional variations in object appearance and visibility.

Region-based object detection has evolved through several architectural developments intended to improve localization accuracy and computational efficiency. Girshick et al. [10] introduced Region-based Convolutional Neural Network (R-CNN), which combines selective search for candidate region generation with convolutional feature extraction and object classification. Although this approach demonstrated the effectiveness of region-based object detection, its independent processing of region proposals required substantial computational resources. Girshick [11] subsequently introduced Fast R-CNN, which improved detection efficiency by sharing convolutional computations across candidate regions. Ren et al. [12] further developed Faster R-CNN by integrating a Region Proposal Network (RPN), allowing candidate object regions to be generated directly from shared convolutional feature maps. In addition, He et al. [13] proposed Residual Networks (ResNet), which facilitate the training of deep neural architectures through residual connections and have become widely used for feature extraction in object detection.

Despite these advances, detecting motorcycle rider faces remains challenging because facial features may be partially or substantially obscured by protective equipment. Different helmet configurations, visor positions, and face masks can reduce the visibility of facial characteristics required for accurate detection. Furthermore, variations in camera perspective and image orientation may alter the appearance and spatial representation of the target region. Such conditions can complicate object localization and increase the possibility of incorrect predictions. Another important consideration is the region proposal mechanism employed by the detection architecture. Conventional R-CNN generates candidate regions through selective search, whereas Faster R-CNN employs a trainable Region Proposal Network to identify potential object locations. These mechanisms differ in their feature-processing procedures and computational requirements, potentially affecting their suitability for detecting partially visible facial regions. Although data augmentation has been widely investigated as a technique for improving training diversity and model generalization [15], the effectiveness of different detection architectures cannot be determined solely through variations in training data. A comparative investigation is therefore needed to understand how region-based detection approaches respond to the visual challenges associated with motorcycle rider faces.

To address these challenges, this study experimentally compares conventional Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN for motorcycle rider face detection. Both architectures employ region-based object detection mechanisms but differ fundamentally in how candidate object regions are generated and processed. Conventional R-CNN relies on selective search to identify potential object regions before performing feature extraction and classification, whereas Faster R-CNN integrates a Region Proposal Network (RPN) to generate candidate regions through shared convolutional features. The experimental comparison evaluates the detection and classification performance of both architectures using the same motorcycle rider image dataset and evaluation framework. This approach provides quantitative evidence of their performance differences under varying facial visibility and image orientation conditions.

Previous studies in transportation-related computer vision have demonstrated the applicability of deep learning models to automated monitoring tasks. Sholehurrohman et al. [16] investigated YOLOv7 for vehicle license plate color detection and classification, highlighting the potential of object detection techniques for transportation surveillance. Similarly, Junaidi et al. [17] analyzed the robustness of YOLOv7 for motorcycle traffic violation detection under different CCTV traffic-density conditions.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

Although these studies contribute to intelligent transportation monitoring, their detection targets and architectures differ from motorcycle rider facial region detection. Previous research on motorcycle rider face classification using Faster R-CNN with ResNet-50 has primarily emphasized model implementation, hyperparameter optimization, and performance evaluation under variations in facial visibility and image orientation. However, a direct experimental comparison between conventional R-CNN and Faster R-CNN under consistent dataset partitions and evaluation conditions remains insufficiently explored. In particular, the effects of different region proposal mechanisms on detection and classification performance in motorcycle rider images require further empirical investigation. Therefore, this study addresses this research gap by experimentally comparing R-CNN and Faster R-CNN using the same image dataset and evaluation framework.

The novelty of this research lies in the experimental comparison of conventional R-CNN and Faster R-CNN for motorcycle rider face detection under varying facial visibility conditions. Unlike previous studies that primarily focused on implementing and optimizing Faster R-CNN, this research evaluates two distinct region proposal mechanisms within a comparative experimental framework. Conventional R-CNN employs selective search to generate candidate regions, whereas Faster R-CNN integrates a trainable Region Proposal Network (RPN) with shared convolutional feature extraction. Both architectures are evaluated using consistent training and testing procedures to examine differences in detection accuracy, classification performance, and localization quality. The experimental comparison provides quantitative evidence of the relative effectiveness of both architectures in detecting facial regions affected by helmets, visors, face masks, and image orientation variations. Through this approach, the study contributes empirical findings to the application of region-based object detection methods in intelligent transportation monitoring.

The primary objective of this study is to experimentally evaluate and compare the performance of conventional R-CNN and Faster R-CNN for motorcycle rider face detection. Specifically, the research examines the detection performance of both architectures using mean Average Precision (mAP), classification loss, localization loss, accuracy, precision, recall, and F1-score. The experiments investigate model performance under different learning rates and assess classification results using motorcycle rider images with variations in facial visibility and image orientation. The main contributions of this study include: (1) an experimental implementation and comparison of R-CNN and Faster R-CNN using a consistent motorcycle rider image dataset; (2) a quantitative evaluation of detection and classification performance based on actual model predictions; and (3) an empirical analysis of the differences between selective search-based region proposals and the integrated Region Proposal Network employed by Faster R-CNN. These findings provide a basis for selecting suitable region-based object detection architectures for motorcycle rider monitoring applications.

2. RESEARCH METHODOLOGY

This study adopts an experimental comparative approach to evaluate the performance of R-CNN and Faster R-CNN for motorcycle rider face detection and classification. The research methodology consists of several stages, including dataset collection, image annotation, preprocessing, data augmentation, model implementation, training, testing, and performance evaluation. Both architectures were implemented and evaluated using the same underlying image dataset and experimental framework to ensure a consistent comparison of their detection and classification capabilities. The experiments were conducted using two learning rate configurations, 0.01 and 0.001, with a fixed training duration of 80 epochs. The evaluation focused on detection and classification performance using classification loss, localization loss, mean Average Precision (mAP), accuracy, precision, recall, and F1-score. The experimental procedures were designed to examine the performance differences between the two architectures in detecting motorcycle rider facial regions under variations in facial visibility and image orientation.

2.1. Research Framework

The research framework consists of image acquisition, annotation, preprocessing, dataset partitioning, augmentation, model development, and performance evaluation. Following image

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

acquisition, bounding-box annotations were assigned to identify the facial regions and their corresponding class labels. The annotated images were subsequently processed through cropping and resizing before being divided into training, validation, and testing subsets. Rotation-based data augmentation was applied exclusively to the training subset to increase the diversity of training images.

Two region-based object detection architectures, conventional R-CNN and Faster R-CNN, were implemented as the primary experimental models. Conventional R-CNN employs selective search to generate candidate object regions [10], whereas Faster R-CNN incorporates a Region Proposal Network (RPN) to generate candidate regions through shared convolutional features [12]. Both models were trained and evaluated using the same dataset partitions and experimental evaluation framework. Their performance was compared using detection and classification metrics, including classification loss, localization loss, mean Average Precision (mAP), accuracy, precision, recall, and F1-score. The overall experimental workflow is illustrated in Figure 1.

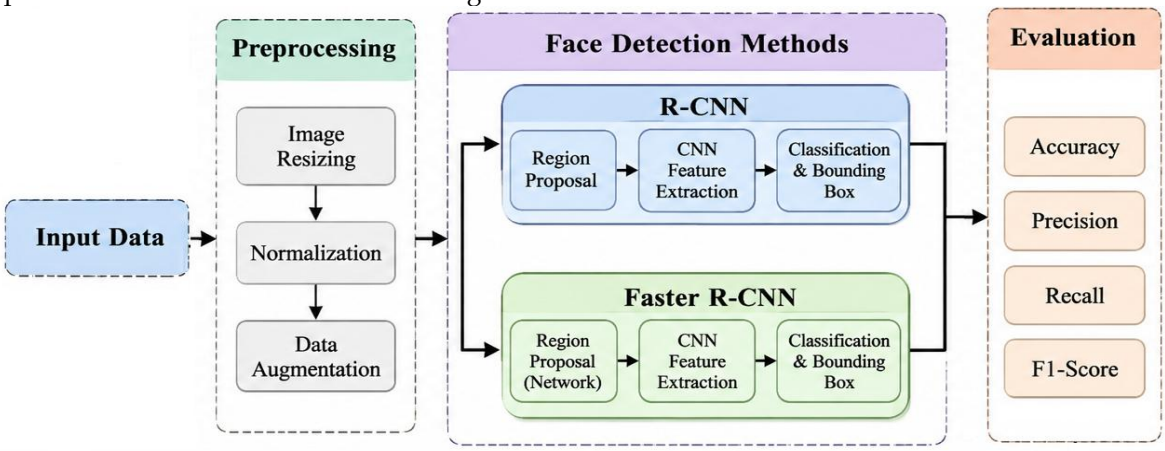


Figure 1. Experimental Research Framework

2.2. Dataset Collection and Annotation

The dataset consists of 2,000 images of individual motorcycle riders collected at the motorcycle parking area of the Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia. Images were acquired using a smartphone during daytime observations, between 10:00 a.m. and 2:00 p.m., under relatively favorable lighting conditions. The acquisition process captured motorcycle riders from different viewing perspectives to represent variations in facial appearance.

The dataset includes motorcycle riders with different facial visibility conditions based on the use of protective equipment. Five facial visibility conditions were represented in the collected images:

1. Motorcycle riders without helmets.
2. Motorcycle riders wearing helmets with fully closed double visors.
3. Motorcycle riders wearing helmets with only the inner visor closed.
4. Motorcycle riders wearing helmets and face masks.
5. Motorcycle riders wearing helmets with closed inner visors and face masks.

These conditions represent different levels of facial visibility and provide a basis for evaluating the detection capabilities of both architectures when facial regions are partially obscured. All collected images were stored in JPG or JPEG format.

Image annotation was performed by assigning bounding boxes to the relevant facial regions. Each bounding box contains spatial coordinates describing the location and size of the target region, together with its corresponding class label. The annotated dataset comprised five classes corresponding to the facial visibility conditions used in the research. These annotations served as ground-truth information for supervised object detection, model training, and subsequent performance evaluation.

2.3. Data Preprocessing and Augmentation

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

Data preprocessing was conducted to prepare the collected images for model training. The main preprocessing operations consisted of image cropping and resizing. Cropping was applied to reduce irrelevant background information and emphasize the region associated with the motorcycle rider's face. Subsequently, all images were resized to an input resolution of 600 × 600 pixels according to the experimental configuration. After preprocessing, the dataset was partitioned into training, validation, and testing subsets using image-level random sampling with a ratio of 70:10:20. The resulting subsets contained 1,400 training images, 200 validation images, and 400 testing images, respectively. Because the partitioning was performed through random sampling rather than grouping images by rider identity, images of the same rider may have been assigned to more than one subset. Consequently, the testing results should be interpreted as performance on the randomly sampled image set and should not be taken as evidence of generalization to entirely unseen riders.

Table 1. Distribution of the Motorcycle Rider Image Dataset

Dataset	Number of images	Percentage
Training	1,400	70%
Validation	200	10%
Testing	400	20%
Total	**2,000**	**100%**

To increase the diversity of the training data, rotation-based augmentation was applied using positive and negative rotations of 10°. Data augmentation is widely employed to increase sample diversity and improve the generalization capabilities of deep learning models [15]. This process generated 2,800 additional images from the original training subset, producing a total of 4,200 training images. The validation and testing subsets were kept separate from the augmented training data. Applying augmentation exclusively to the training subset after dataset partitioning helps prevent transformed images from appearing in the evaluation subsets, although potential overlap between images of the same rider across different subsets remains a consideration when interpreting model generalization.

Table 2. Distribution of Training Images After Augmentation

Image transformation	Number of images
Original training images	1,400
Rotation +10°	1,400
Rotation −10°	1,400
Total training images	**4,200**

The augmentation process was intended to increase variability in facial orientation and improve the model's ability to recognize visual patterns under different image conditions. However, the effects of augmentation on detection performance must be evaluated through the resulting model predictions rather than assumed from the increased dataset size alone.

2.4. R-CNN and Faster R-CNN Architecture

Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN are deep learning architectures designed to identify and localize objects through candidate region generation, feature extraction, and classification. Both architectures employ region-based object detection approaches but differ in their mechanisms for generating candidate regions and processing image features [2], [10], [12]. Conventional R-CNN processes candidate regions independently, whereas Faster R-CNN integrates region proposal generation and feature extraction into a shared convolutional framework [10]–[12].

Conventional R-CNN was introduced by Girshick et al. [10] to combine region proposal generation with deep convolutional feature extraction. The architecture uses selective search, developed by Uijlings et al. [18], to identify candidate object regions based on hierarchical grouping and visual similarities. In the implementation evaluated in this study, VGG16 was used as the convolutional backbone for feature extraction. Each candidate region was resized and processed by the VGG16 network to obtain feature

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

representations for subsequent classification and bounding-box regression. The region proposal process is expressed in Equation (1).

$$R = S(I) \quad (1)$$

where I denotes the input image, $S(\cdot)$ represents the selective search function, and R represents the resulting set of candidate object regions. The convolutional feature extraction process for each candidate region is expressed in Equation (2).

$$f_i = F_\theta(r_i) \quad (2)$$

where r_i denotes the i-th candidate region, $F_\theta(\cdot)$ represents the convolutional feature extraction function parameterized by θ, and f_i represents the extracted feature vector. Although R-CNN established the effectiveness of deep convolutional features for object detection, its repeated processing of candidate regions introduces substantial computational overhead [10], [11].

To address this limitation, Girshick [11] introduced Fast R-CNN, which performs convolutional feature extraction on the entire image and uses region-based feature pooling to process candidate regions. Faster R-CNN subsequently improved this approach by replacing external region proposal generation with a trainable Region Proposal Network (RPN), enabling candidate regions to be predicted directly from shared convolutional feature maps [12]. This integration reduces redundant computation and enables joint optimization of region proposal generation and object detection [11], [12].

Faster R-CNN consists of a convolutional backbone, an RPN, and a detection stage responsible for object classification and bounding-box refinement [12]. In this study, ResNet-50 is employed as the feature extraction backbone. The residual learning architecture introduced by He et al. [13] facilitates the training of deep convolutional networks through shortcut connections, supporting hierarchical feature extraction. The RPN predicts objectness scores and bounding-box adjustments, while the subsequent detection stage classifies candidate regions and refines their spatial coordinates [12], [13]. The Faster R-CNN training objective combines region proposal classification, region proposal localization, object classification, and bounding-box regression losses. The overall loss function is represented in Equation (3).

$$L_{total} = L_{RPN,cls} + L_{RPN,loc} + L_{det,cls} + L_{det,loc} \quad (3)$$

where L_{total} represents the overall training loss, $L_{RPN,cls}$ and $L_{RPN,loc}$ denote the RPN classification and localization losses, respectively, while $L_{det,cls}$ and $L_{det,loc}$ denote the final detection classification and localization losses. These components provide a conceptual representation of the multi-task training objective employed in Faster R-CNN [12].

Both R-CNN and Faster R-CNN belong to the same family of region-based object detection architectures and employ a two-stage detection approach. However, their primary differences lie in the mechanisms used to generate candidate object regions and extract image features. Conventional R-CNN employs selective search and independent feature extraction [10], [18], whereas Faster R-CNN uses an integrated RPN and shared convolutional features [12], [19]. These architectural differences provide the basis for investigating the detection and classification performance of both models when applied to motorcycle rider images containing partially obscured facial regions.

For conventional R-CNN, the detection process involved candidate region generation through selective search, feature extraction using the VGG16 backbone, object classification, and bounding-box regression. In contrast, Faster R-CNN employed ResNet-50 for feature extraction and an integrated Region Proposal Network (RPN) to generate candidate regions, followed by object classification and bounding-box refinement. Therefore, the two implementations differed in both their feature extraction backbones and region proposal mechanisms. In this study, the respective components were implemented using their default configurations, including Selective Search, classifier settings, and bounding-box regression, without additional parameter modifications. Consequently, the experimental comparison evaluates the overall performance of both implemented architectures rather than isolating the contribution of individual architectural components.

2.5. Experimental Configuration

The experimental evaluation was conducted using conventional Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN architectures to compare their performance in detecting

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

and classifying motorcycle rider facial regions. Both models were trained and evaluated using the same underlying dataset partitions, consisting of 1,400 original training images, 200 validation images, and 400 testing images. Rotation-based data augmentation increased the training dataset to 4,200 images, including the original training images and images rotated by +10° and −10°.

The comparative experiments were performed using two learning rates, 0.01 and 0.001, with a fixed training duration of 80 epochs for both architectures. These configurations were used to investigate the influence of learning rate selection on model training and detection performance. The training evaluation included classification loss, localization loss, and mean Average Precision (mAP) to compare the detection capabilities of both models.

Conventional R-CNN employed selective search to generate candidate object regions, followed by convolutional feature extraction, classification, and bounding-box refinement. In contrast, Faster R-CNN incorporated an integrated Region Proposal Network (RPN) and shared convolutional features to generate and refine candidate object regions. These differences in region proposal generation and feature processing formed the basis of the experimental comparison. Following the training process, both models were evaluated using 400 testing images under three image orientation conditions: original images, images rotated by +10°, and images rotated by −10°. The testing procedure was designed to assess the classification performance of the trained architectures under variations in image orientation. The experimental evaluation produced classification performance measurements consisting of accuracy, precision, recall, and F1-score. The resulting training and testing performance metrics were used to compare R-CNN and Faster R-CNN and determine their relative effectiveness in motorcycle rider face detection under the evaluated experimental conditions. The experimental comparison was based on identical dataset partitions and evaluation metrics, while differences in model architecture and implementation settings were considered when interpreting the observed performance variations.

2.6. Performance Evaluation

The performance of R-CNN and Faster R-CNN was evaluated using accuracy, precision, recall, F1-score, and mean Average Precision (mAP). These metrics were employed to assess classification performance and object detection capabilities under the evaluated experimental conditions. Accuracy, precision, recall, and F1-score were used to measure classification performance, while mAP was employed to evaluate object detection performance based on precision–recall relationships and object localization accuracy. Classification performance was assessed using four fundamental prediction outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP represents correctly identified positive instances, TN represents correctly identified negative instances, FP represents negative instances incorrectly classified as positive, and FN represents positive instances incorrectly classified as negative. These components provide the basis for calculating classification performance metrics [20], [21].

Accuracy measures the proportion of correctly classified instances among all evaluated instances. This metric provides an overall indication of classification correctness. However, accuracy may be influenced by imbalanced class distributions, making it important to consider additional metrics when evaluating detection models [20], [21]. Accuracy is calculated using Equation (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

Precision measures the proportion of correctly predicted positive instances among all instances classified as positive. A higher precision value indicates a lower proportion of false-positive predictions, reflecting greater reliability in positive classifications. Precision has been widely applied in intelligent transportation monitoring systems to evaluate object detection and classification performance, including vehicle license plate color classification and traffic prohibition sign detection [19], [20], [22], [23]. Precision is calculated using Equation (5).

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

Recall, also referred to as sensitivity, measures the proportion of actual positive instances correctly identified by the model. This metric indicates the ability of a detection system to identify relevant target instances and is particularly important when missed detections affect monitoring reliability. Previous studies on transportation-related object detection have employed recall alongside precision to evaluate

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

detection effectiveness under different environmental conditions [20], [22], [23]. Recall is calculated using Equation (6).

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

The F1-score combines precision and recall using their harmonic mean to provide a balanced measure of classification performance. This metric is particularly useful when both false-positive and false-negative predictions need to be considered. Previous experimental studies involving vehicle license plate color classification and traffic prohibition sign detection have demonstrated the use of the F1-score in comparing object detection architectures [21], [22], [23]. The F1-score is calculated using Equation (7).

$$F1 - score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \quad (7)$$

For the calculation of the F1-score, precision and recall are expressed on the same scale, either as proportions or percentages. The classification evaluation in this study involved five classes corresponding to the motorcycle rider facial visibility conditions. A confusion matrix was generated by comparing the predicted class labels with the actual class labels. Accuracy, precision, recall, and F1-score were subsequently calculated using the default averaging configuration of the implemented evaluation procedure. This approach was applied consistently to both R-CNN and Faster R-CNN to maintain consistency in the classification performance comparison. In addition to classification performance, object detection models are evaluated using spatial localization measures. Intersection over Union (IoU) quantifies the overlap between a predicted bounding box and its corresponding ground-truth bounding box by calculating the ratio of the intersection area to the union area. A higher IoU value indicates greater spatial agreement between the predicted and actual object regions [19], [24]. Intersection over Union (IoU) is calculated using the following equation (8).

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (8)$$

Average Precision (AP) summarizes detection performance across precision–recall operating points, while mean Average Precision (mAP) represents the average AP across the evaluated object categories. In this study, detection performance was evaluated using mAP50 (mAP@0.5), which applies an IoU threshold of 0.50 to determine whether a predicted bounding box correctly matches its corresponding ground-truth bounding box. A prediction is considered a correct detection when its predicted class matches the ground-truth class and its IoU value meets or exceeds the specified threshold. The mAP50 metric was used to compare the detection performance of R-CNN and Faster R-CNN under different learning rate configurations.

The experimental evaluation of both architectures was conducted using 400 testing images under three image orientation conditions: original images, images rotated by +10°, and images rotated by −10°. Accuracy, precision, recall, and F1-score were obtained from the experimental testing procedures implemented for R-CNN and Faster R-CNN. These classification metrics were used to compare the classification performance of both architectures, while mAP50 and the reported classification and localization losses were used to evaluate their object detection performance and training outcomes.

3. RESULT AND DISCUSSIONS

This section presents the experimental results and comparative analysis of R-CNN and Faster R-CNN for motorcycle rider face detection. The evaluation covers model training performance, classification results, and differences in detection capabilities between the two architectures. Training performance is assessed using classification loss, localization loss, and mean Average Precision (mAP), while classification performance is evaluated using accuracy, precision, recall, and F1-score. The experimental findings are discussed to examine the relative performance of both architectures under the evaluated learning rate configurations and image orientation conditions.

3.1. Model Training Results

The training performance of R-CNN and Faster R-CNN was experimentally evaluated using two learning rates, 0.01 and 0.001, with a fixed training duration of 80 epochs. Both architectures were trained using the motorcycle rider image dataset to determine their detection performance under the

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

selected experimental configurations. The evaluation focused on classification loss, localization loss, and mean Average Precision (mAP). The experimental training results are presented in Table 3.

Table 3. Training performance comparison of R-CNN and Faster R-CNN

Model	Learning rate	Epochs	Classification loss	Localization loss	mAP50
R-CNN	0.01	80	0.0285	0.0142	0.7824
R-CNN	0.001	80	0.0637	0.0195	0.4216
Faster R-CNN	**0.01**	**80**	**0.0108**	**0.0061**	**0.8467**
Faster R-CNN	0.001	80	0.0412	0.0081	0.4882

As presented in Table 3, Faster R-CNN achieved better detection performance than conventional R-CNN under both evaluated learning rates. At a learning rate of 0.01, Faster R-CNN obtained an mAP of 0.8467, whereas R-CNN achieved an mAP of 0.7824. This represents an absolute improvement of 0.0643, equivalent to 6.43 percentage points. Faster R-CNN also produced lower classification and localization losses of 0.0108 and 0.0061, respectively, compared with 0.0285 and 0.0142 for R-CNN. At a learning rate of 0.001, Faster R-CNN achieved an mAP of 0.4882, compared with 0.4216 for R-CNN, representing a difference of 6.66 percentage points. Under this configuration, Faster R-CNN recorded classification and localization losses of 0.0412 and 0.0081, respectively, whereas R-CNN produced corresponding losses of 0.0637 and 0.0195. These results indicate that Faster R-CNN consistently achieved higher mAP50 values and lower reported loss values than conventional R-CNN across both learning rate configurations.

The lower classification and localization losses achieved by Faster R-CNN indicate more favorable training outcomes in both class prediction and bounding-box refinement. At a learning rate of 0.01, Faster R-CNN reduced classification loss by approximately 62.11% and localization loss by 57.04% compared with R-CNN. These differences indicate that Faster R-CNN obtained lower reported loss values under the evaluated experimental configuration. However, the loss values should be interpreted in the context of their respective model implementations because differences in loss definitions and normalization may affect direct comparisons between architectures.

The experimental findings also demonstrate the influence of learning rate selection on detection performance. Both architectures achieved their highest reported mAP values at a learning rate of 0.01. Reducing the learning rate to 0.001 resulted in a decrease in mAP from 0.7824 to 0.4216 for R-CNN and from 0.8467 to 0.4882 for Faster R-CNN. This indicates that the selected learning rate had a substantial influence on the detection performance of both architectures within the fixed training duration of 80 epochs. The performance advantage of Faster R-CNN may be associated with its integrated Region Proposal Network and shared convolutional feature extraction. In contrast, conventional R-CNN generates candidate regions using Selective Search and processes each region independently. Nevertheless, the observed differences represent the overall performance of the implemented architectures, which also use different feature extraction backbones, namely VGG16 for R-CNN and ResNet-50 for Faster R-CNN. Therefore, the results should not be interpreted as isolating the contribution of the region proposal mechanism alone.

The substantial reduction in mAP50 at a learning rate of 0.001 suggests that the lower learning rate may have limited model optimization within the fixed training duration of 80 epochs. For R-CNN, mAP50 decreased by approximately 46.12%, whereas Faster R-CNN experienced a decrease of approximately 42.34% compared with their respective results at a learning rate of 0.01. This finding demonstrates that learning rate selection influenced the detection performance of both architectures, although the specific optimization behavior cannot be determined solely from the final evaluation metrics. Overall, Faster R-CNN demonstrated better detection performance than conventional R-CNN across both evaluated learning rate configurations. The best reported performance was obtained by Faster R-CNN using a learning rate of 0.01 and 80 training epochs, with an mAP50 of 0.8467.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

3.2. Model Testing and Classification Results

The testing phase was conducted to evaluate the classification performance of R-CNN and Faster R-CNN in identifying motorcycle rider facial regions under different image orientation conditions. Both trained models were evaluated using 400 testing images under three conditions: original images, images rotated by +10°, and images rotated by −10°. These testing conditions were selected to assess the performance of both architectures when variations in image orientation were introduced.

The classification performance of R-CNN and Faster R-CNN was evaluated using accuracy, precision, recall, and F1-score. These performance metrics were obtained from the experimental testing procedures implemented for both detection models. Accuracy was used to measure overall classification correctness, precision evaluated the reliability of positive predictions, recall measured the ability to identify actual positive instances, and F1-score provided a combined measure of precision and recall. The experimental classification results are presented in Table 4.

Table 4. Classification performance comparison of R-CNN and Faster R-CNN

Evaluation metric	R-CNN	Faster R-CNN
Accuracy	80.17%	89.42%
Precision	85.66%	96.35%
Recall	84.15%	91.56%
F1-score	84.90%	93.89%

As presented in Table 4, Faster R-CNN achieved an accuracy of 89.42%, outperforming conventional R-CNN, which obtained an accuracy of 80.17%. This represents an improvement of 9.25 percentage points. The higher accuracy indicates that Faster R-CNN produced a greater proportion of correct classifications under the evaluated testing conditions. In terms of precision, Faster R-CNN achieved 96.35%, compared with 85.66% for R-CNN, representing an improvement of 10.69 percentage points. This finding indicates that Faster R-CNN demonstrated greater reliability in its positive predictions, with a lower proportion of false-positive classifications. Higher precision is particularly important in automated transportation monitoring because false-positive predictions can reduce the reliability of detection results.

Faster R-CNN also achieved a higher recall of 91.56%, compared with 84.15% for R-CNN, representing an improvement of 7.41 percentage points. This result indicates that Faster R-CNN correctly identified a greater proportion of actual positive instances under the experimental evaluation. Nevertheless, the recall values demonstrate that neither architecture achieved complete identification of all positive instances, indicating that classification errors remained. The difference between precision and recall provides additional insight into the classification performance of Faster R-CNN. Its precision of 96.35% exceeded its recall of 91.56% by 4.79 percentage points. This indicates that the model demonstrated greater reliability in positive predictions than in identifying all actual positive instances. Although this result reflects favorable classification performance, missed positive instances remain an important consideration in motorcycle rider face detection, particularly when facial regions are partially obscured by helmets, visors, or face masks.

The F1-score provides an additional measure of classification performance by combining precision and recall into a single evaluation metric. Faster R-CNN achieved an F1-score of 93.89%, whereas conventional R-CNN obtained 84.90%, representing an improvement of 8.99 percentage points. This finding indicates that Faster R-CNN achieved a more favorable balance between precision and recall in the experimental testing results. Overall, the classification evaluation demonstrates that Faster R-CNN consistently achieved higher accuracy, precision, recall, and F1-score values than conventional R-CNN. These results are consistent with the detection performance presented in Table 3, indicating that Faster R-CNN achieved stronger overall detection and classification performance under the investigated experimental conditions. The classification results also provide an overall assessment of model performance across the evaluated testing data. Since the reported metrics represent aggregated evaluation results, the comparison focuses on the overall classification capabilities of both architectures

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

rather than attributing specific prediction outcomes to individual facial visibility or image orientation conditions.

3.3. Comparative Analysis and Discussion

The experimental comparison of R-CNN and Faster R-CNN revealed differences in their detection and classification performance for motorcycle rider facial regions. Based on the results presented in Tables 3 and 4, Faster R-CNN demonstrated better overall performance than conventional R-CNN under the evaluated experimental configurations. The findings provide quantitative evidence of the relative performance of both architectures in processing motorcycle rider images containing variations in facial visibility and image orientation.

The training results presented in Table 3 indicate that Faster R-CNN consistently achieved higher mean Average Precision (mAP) values and lower classification and localization losses than R-CNN. At a learning rate of 0.01, Faster R-CNN achieved an mAP of 0.8467, compared with 0.7824 for R-CNN, representing an improvement of 6.43 percentage points. At a learning rate of 0.001, Faster R-CNN obtained an mAP of 0.4882, exceeding the R-CNN result of 0.4216 by 6.66 percentage points. These findings demonstrate that Faster R-CNN maintained higher detection performance across both evaluated learning rate configurations.

The classification results presented in Table 4 further demonstrate the performance differences between the two architectures. Faster R-CNN achieved an accuracy of 89.42%, compared with 80.17% for R-CNN, representing an improvement of 9.25 percentage points. Faster R-CNN also achieved higher precision, recall, and F1-score values of 96.35%, 91.56%, and 93.89%, respectively, whereas R-CNN obtained 85.66%, 84.15%, and 84.90%. The consistent improvements across all four classification metrics indicate that Faster R-CNN provided more favorable classification performance under the evaluated testing conditions.

The performance differences can be discussed in relation to the architectural characteristics of the two implemented models. Conventional R-CNN generates candidate object regions using selective search and processes these regions independently through convolutional feature extraction. In contrast, Faster R-CNN employs an integrated Region Proposal Network (RPN) that generates candidate regions using shared convolutional features. This integration reduces redundant feature extraction and supports a more unified detection process. Furthermore, the two implementations use different feature extraction backbones, with VGG16 for conventional R-CNN and ResNet-50 for Faster R-CNN. The combination of region proposal mechanisms and feature extraction architectures distinguishes the two implementations evaluated in this study. Therefore, the higher mAP50 and classification performance obtained by Faster R-CNN indicate the stronger overall performance of its implemented architecture, rather than establishing the independent contribution of the RPN or backbone.

The experimental results also highlight the importance of learning rate selection. Both architectures achieved their highest reported mAP values at a learning rate of 0.01 with 80 training epochs. Reducing the learning rate to 0.001 resulted in a substantial decrease in mAP for both models. These findings indicate that learning rate selection played an important role in the detection performance of both architectures under the evaluated training configurations.

Another important consideration in motorcycle rider face detection is the presence of partial facial occlusion and image orientation variations. Helmets, visors, and face masks can obscure facial characteristics and reduce the visual information available for detection and classification. Furthermore, variations in image orientation may change the spatial arrangement of facial features and influence model predictions. The experimental testing procedure included original images and images rotated by +10° and −10° to evaluate both architectures under different orientation conditions. The inclusion of these image orientation conditions provided an evaluation setting for examining model performance when variations in image orientation were introduced. The classification results, however, were reported as aggregated metrics across the evaluated testing data. Consequently, the individual effects of each image orientation and facial visibility condition cannot be quantified separately from the reported results.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

Although Faster R-CNN demonstrated better detection and classification performance, several limitations should be considered when interpreting the findings. The experiments were conducted using images collected under relatively favorable daytime lighting conditions and were limited to the selected learning rate configurations. Furthermore, the dataset was collected from a single motorcycle parking area at Universitas Lampung, which may limit the generalizability of the findings to other transportation environments with different backgrounds, camera perspectives, illumination levels, and weather conditions.

Another consideration is that the experimental comparison focused on detection and classification performance rather than directly measuring computational efficiency. Therefore, the observed improvements establish the stronger performance of Faster R-CNN in terms of the reported evaluation metrics, but do not independently demonstrate differences in inference speed or computational resource consumption. Future experiments should include standardized measurements of inference latency and computational resource consumption using identical hardware and software environments to provide a more comprehensive comparison of both architectures.

Overall, the experimental findings demonstrate that Faster R-CNN outperformed conventional R-CNN in motorcycle rider face detection under the evaluated conditions. Its higher mAP50, accuracy, precision, recall, and F1-score indicate stronger detection and classification capabilities. These findings provide empirical evidence supporting the use of Faster R-CNN for the investigated motorcycle rider face detection task. Nevertheless, further evaluation under more diverse environmental conditions is necessary to assess its applicability in real-world intelligent transportation monitoring systems.

4. CONCLUSION

This study experimentally compared conventional Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN for motorcycle rider face detection under varying facial visibility and image orientation conditions. Both architectures were evaluated using the same underlying dataset and experimental framework to assess their detection and classification performance. The experimental results demonstrated that Faster R-CNN consistently outperformed conventional R-CNN across the evaluated learning rate configurations. At a learning rate of 0.01 with 80 training epochs, Faster R-CNN achieved the highest mAP50 of 0.8467, compared with 0.7824 for R-CNN, representing an improvement of 6.43 percentage points. Faster R-CNN also achieved lower reported classification and localization losses, indicating more favorable training outcomes under the evaluated configurations.

The classification evaluation further demonstrated the superior performance of Faster R-CNN, which achieved an accuracy of 89.42%, precision of 96.35%, recall of 91.56%, and F1-score of 93.89%. In comparison, R-CNN obtained an accuracy of 80.17%, precision of 85.66%, recall of 84.15%, and F1-score of 84.90%. These results indicate that Faster R-CNN achieved better overall detection and classification performance than conventional R-CNN under the experimental conditions investigated in this study. The observed performance differences reflect the overall characteristics of the two implemented architectures. Conventional R-CNN employed Selective Search and VGG16 for region proposal generation and feature extraction, respectively, whereas Faster R-CNN integrated a Region Proposal Network (RPN) with ResNet-50 for feature extraction. Therefore, the experimental findings demonstrate the comparative performance of the complete implementations rather than the isolated contribution of individual architectural components.

The results suggest that the integrated Region Proposal Network and shared convolutional feature extraction employed by Faster R-CNN provide architectural advantages for detecting motorcycle rider facial regions. However, the study was limited to the collected image dataset, selected learning rate configurations, and aggregated classification measurements. Furthermore, the dataset was collected from a single location under relatively favorable daytime lighting conditions, which limits the extent to which the findings can be generalized to other transportation environments. In addition, computational efficiency was not directly evaluated through inference-time or computational resource measurements. Consequently, the observed performance differences should be interpreted primarily in terms of detection and classification effectiveness rather than computational efficiency.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

Journal of Artificial Intelligence and Technology Information
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 4, December 2026
https://doi.org/10.58602/jaiti.v4i4.471
Page 798-812

Future research should investigate model performance using more diverse datasets collected under different environmental conditions, including variations in lighting, camera perspectives, weather conditions, and facial visibility. Further studies may also evaluate additional training configurations and conduct standardized computational performance measurements to provide a more comprehensive comparison of region-based object detection architectures for motorcycle rider face detection.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Muhaqiqin: Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. **Ridho Sholehurrohman**: Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft, Writing – Review & Editing. **Agung Pambudi**: Project Administration, Resources, Software, Writing – Review & Editing. **Igit Sabda Ilman**: Data Curation, Investigation, Resources, Software, Supervision. **Muhammad Afdhaluddin**: Project Administration, Resources, Software, Visualization. **Zulkarnaini**: Conceptualization, Resources, Software, Validation, Writing – Review & Editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Computer Science, Universitas Lampung, for providing laboratory facilities and assistance with data analysis. The authors also thank all individuals and institutions who contributed to the data collection, analysis, and completion of this research.

FUNDING STATEMENTS

This research was funded by the Basic Research BLU, Universitas Lampung, under Receipt Number 6188/UN.26.21/TU/2026.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors utilized ChatGPT to enhance the clarity and improve the language of this manuscript. All AI-assisted content was carefully reviewed and revised by the authors, who take full responsibility for the final content of the manuscript.

REFERENCES

[1] Y. Sun, Z. Sun, and W. Chen, "The evolution of object detection methods," Engineering Applications of Artificial Intelligence, vol. 133, Art. no. 108458, 2024, doi: 10.1016/j.engappai.2024.108458.

[2] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[3] H. L. Gururaj, B. C. Soundarya, M. Priya, V. Shreyas, and F. Flammini, "A comprehensive review of face recognition techniques, trends, and challenges," IEEE Access, vol. 12, pp. 107903–107926, 2024, doi: 10.1109/ACCESS.2024.3424933.

[4] U. Jayaraman, P. Gupta, S. Gupta, G. Arora, and K. Tiwari, "Recent development in face recognition," Neurocomputing, vol. 408, pp. 231–245, 2020, doi: 10.1016/j.neucom.2019.08.110.

[5] W. Ali, W. Tian, S. U. Din, M. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: A complete review," Multimedia Tools and Applications, vol. 80, pp. 4825–4880, 2021, doi: 10.1007/s11042-020-09850-1.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/

[6] S. Megawan and W. S. Lestari, "Deteksi spoofing wajah menggunakan Faster R-CNN dengan arsitektur Resnet50 pada video," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 9, no. 3, pp. 261–267, 2020, doi: 10.22146/.v9i3.231.

[7] S. Megawan, W. S. Lestari, and A. Halim, "Deteksi non-spoofing wajah pada video secara real time menggunakan Faster R-CNN," Journal of Information System Research (JOSH), vol. 3, no. 3, pp. 291–299, 2022, doi: 10.47065/josh.v3i3.1519.

[8] Y. Sari, "Klasifikasi jenis kelamin menggunakan metode Convolutional Neural Network (CNN) dengan data wajah manusia," Undergraduate thesis, Universitas Lampung, Bandar Lampung, Indonesia, 2022.

[9] D. Alamsyah and D. Pratama, "Deteksi ujung jari menggunakan Faster-RCNN dengan arsitektur Inception v2 pada citra derau," Jurnal Sistem dan Teknologi Informasi Komunikasi, vol. 2, no. 1, pp. 6–10, 2018.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[11] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[14] N. Rizzieri, L. Dall'Asta, and M. Ozoliņš, "Diabetic retinopathy features segmentation without coding experience with computer vision models YOLOv8 and YOLOv9," Vision, vol. 8, no. 3, Art. no. 48, 2024, doi: 10.3390/vision8030048.

[15] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," Array, vol. 16, Art. no. 100258, 2022, doi: 10.1016/j.array.2022.100258.

[16] R. Sholehurrohman, K. Muludi, and J. Triloka, "Optimizing transportation surveillance with YOLOv7: Detection and classification of vehicle license plate colors," Jurnal Ilmu Pengetahuan dan Teknologi Komputer, vol. 10, no. 4, pp. 812–820, 2025, doi: 10.33480/jitk.v10i4.6260.

[17] A. Junaidi, R. Sholehurrohman, A. N. Ganandy, and F. R. Lumbanraja, "Robustness analysis of YOLOv7 for motorcycle traffic violation detection across multi-density CCTV scenarios," Integra: Journal of Integrated Mathematics and Computer Science, vol. 3, no. 2, pp. 63–69, 2026, doi: 10.26554/integrajimcs.20263252.

[18] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, 2013, doi: 10.1007/s11263-013-0620-5.

[19] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.

[20] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[21] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, Art. no. 6, 2020, doi: 10.1186/s12864-019-6413-7.

[22] R. Sholehurrohman, Muhaqiqin, I. S. Ilman, and R. Habibi, "Assessing detection and classification performance for vehicle license plate colors using YOLOv5, YOLOv7, YOLOv8, and YOLOv9," Jurnal Pepadun, vol. 7, no. 1, pp. 67–79, 2026, doi: 10.23960/pepadun.v7i1.349.

[23] R. Sholehurrohman, A. Junaidi, T. N. Dewi, R. Andrian, I. Sabda, and M. R. Habibi, "YOLOv5s for traffic prohibition sign detection in Bandar Lampung: An empirical evaluation under real-world urban conditions," International Journal of Electronics and Communications System, vol. 6, no. 1, pp. 29–40, 2026, doi: 10.24042/ijecs.v6i1.28786.

[24] R. Sholehurrohman, A. Junaidi, H. Hanif, F. R. Lumbanraja, and M. R. Habibi, "Comparison of FA-SSD: Performance of feature extractors VGG-19, MobileNetV3, and ResNet-152 for human body temperature prediction," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 12, no. 2, pp. 300–311, 2026, doi: 10.25077/TEKNOSI.v12i2.2026.300-311.

[25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in Computer Vision – ECCV 2014, Lecture Notes in Computer Science, vol. 8693, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.

Copyright © 2026, The Author(s).
This is an open access article under the CC BY-SA license
https://creativecommons.org/licenses/by-sa/4.0/