ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

Enhancing The Accuracy of Small Object Detection In Traffic Safety Attributes Using Yolov11 And Esrgan

Tuahta Hasiholan Pinem[1], Muhammad Haris[2]

[1,2] Ilmu Komputer, Universitas Nusa Mandiri Margonda, Indonesia

[1*]14220020@nusamandiri.ac.id, [2]muhammad.uhs@nusamandiri.ac.id
*: *Penulis korenspondensi* (*corresponding author*)

Informasi Artikel

Received: June 2025
Revised: June 2025
Accepted: June 2025
Published: October 2025

Abstract

This study aims to detect motorcycle rider attributes, specifically helmets and side mirrors, using a deep learning approach combining YOLOv11 and ESRGAN models. The proposed model addresses challenges in attribute detection under real-world conditions, such as low-resolution images, varying angles, and uneven lighting. The dataset comprises images of motorcycle riders captured by surveillance cameras (CCTV), which underwent preprocessing, augmentation, and resolution enhancement using ESRGAN to improve input quality.

The results demonstrate that ESRGAN significantly enhances the performance of YOLOv11, particularly for high-resolution images. The YOLOv11 + ESRGAN model with 300 epochs achieved the best performance, with precision of 75.8%, recall of 69.1%, and an F1-score of 0.7 during testing. During validation, the model reached a precision of 0.826 and recall of 0.797, indicating good generalization capabilities. Compared to the YOLOv11 model without ESRGAN, this combination significantly improved accuracy, especially in detecting small attributes such as side mirrors.

This study suggests further exploration with larger and more diverse datasets and fine-tuning to enhance detection accuracy. Additionally, integrating the model into real-world systems based on edge computing can accelerate real-time inference and reduce reliance on cloud-based servers. With broader implementation, this

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

model has the potential to improve the efficiency and safety of AI-powered traffic monitoring systems.

Keywords: Rider attribute detection, YOLOv11, ESRGAN, Deep learning, Surveillance camera
Kata kunci: Model Machine Learning, Algoritma Voting Classifier, Pengelompokan Fitur

Abstrak

Penelitian ini bertujuan untuk mendeteksi atribut pengendara sepeda motor, khususnya helm dan spion samping, dengan menggunakan pendekatan *deep learning* yang menggabungkan model YOLOv11 dan ESRGAN. Model yang diusulkan ini mengatasi tantangan dalam deteksi atribut di bawah kondisi dunia nyata, seperti gambar beresolusi rendah, sudut pandang yang bervariasi, dan pencahayaan yang tidak merata. Dataset yang digunakan terdiri dari gambar pengendara sepeda motor yang diambil oleh kamera pengawas (CCTV), yang telah menjalani pra-pemrosesan, augmentasi, dan peningkatan resolusi menggunakan ESRGAN untuk meningkatkan kualitas input.

Hasil penelitian menunjukkan bahwa ESRGAN secara signifikan meningkatkan kinerja YOLOv11, terutama untuk gambar dengan resolusi tinggi. Model YOLOv11 + ESRGAN dengan 300 epoch mencapai kinerja terbaik, dengan presisi 75,8%, *recall* 69,1%, dan *F1-score* sebesar 0,7 pada pengujian. Selama validasi, model ini mencapai presisi 0,826 dan *recall* 0,797, yang menunjukkan kemampuan generalisasi yang baik. Dibandingkan dengan model YOLOv11 tanpa ESRGAN, kombinasi ini secara signifikan meningkatkan akurasi, terutama dalam mendeteksi atribut kecil seperti spion samping.

Penelitian ini menyarankan eksplorasi lebih lanjut dengan dataset yang lebih besar dan lebih beragam serta penyesuaian untuk meningkatkan akurasi deteksi. Selain itu, mengintegrasikan model ini ke dalam sistem dunia nyata yang berbasis edge computing dapat mempercepat inferensi waktu nyata dan mengurangi ketergantungan pada server berbasis cloud. Dengan implementasi yang lebih luas, model ini memiliki potensi untuk meningkatkan efisiensi dan keselamatan sistem pemantauan lalu lintas yang didukung oleh AI.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

1. Introduction

Road safety is a critical issue that requires special attention, particularly for motorcycle riders. Safety attributes such as helmets and mirrors play a vital role in protecting riders from serious injuries. Monitoring of these safety attributes is conducted through CCTV cameras installed by the Transportation Department, as shown in Figure 1.

Road safety is an important aspect that affects public health and safety. In Indonesia, traffic accidents remain one of the leading causes of death, especially among motorcycle riders. Small object attributes, such as helmets and mirrors, play a crucial role in safety by protecting riders from serious injuries during accidents.

Figure 1. CCTV Image Data from Bogor City

Therefore, enforcing rules regarding the use of safety attributes is essential to reduce the risk of accidents. The Transportation Agency of Bogor City has installed CCTV systems at various strategic points to monitor traffic and ensure compliance with safety regulations. However, the current detection system still faces challenges in terms of accuracy and detection speed of safety attributes. Conventional technology often struggles to detect small objects quickly and accurately, thus affecting the effectiveness of surveillance.

To address this issue, a more advanced and efficient approach is required. Artificial intelligence technologies, such as YOLO (You Only Look Once) YOLOv11 with computational requirements (GFLOP), and the YOLOv11x model, show the most consistent performance. These models have successfully achieved high accuracy, fast processing times, low power consumption, and efficient disk space usage [1], while ESRGAN offers a potential solution for image quality enhancement [2]. YOLOv11 is known as one of the best object detection models capable of quickly and accurately detecting objects in images. Meanwhile, ESRGAN can be used to generate more realistic training data, thus improving the performance of detection models.

This research aims to develop and implement a more accurate and efficient vehicle safety attribute detection model using a combination of YOLOv11 and ESRGAN on the CCTV system of the Bogor City Transportation Agency. The primary focus of this study is to improve the detection accuracy of small safety attribute objects, such as helmets and mirrors, as well as to accelerate the detection processing. By achieving this goal, it is expected that traffic safety monitoring can be carried out more effectively, thereby contributing to improving traffic safety in Bogor City.

The research involves several stages—specific stages related to the method include collecting and categorizing image data from CCTV recordings under various lighting conditions and

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

viewpoints, with data gathered on December 4, 10, 16, 23, 24, and 30, 2024. The process includes annotation, augmentation, as well as exploration and validation of the dataset using Cvat, Albumentations, and FiftyOne to enhance the efficiency and quality of data used for training the ESRGAN (RealESRGAN_x2plus) model to improve and refine the quality of training data to be more representative and realistic. The results from the accuracy of the model are developed using the YOLOv11 detection model.

This research is based on the understanding that the accuracy and effectiveness of detecting small safety attribute objects are key to improving traffic monitoring and rider safety. Using the YOLOv11 model with ESRGAN can enhance detection performance through improved training data quality. The use of Cvat for annotation and Albumentations for augmentation will be highly useful in improving the quality and diversity of the dataset, with FiftyOne serving as a tool for data exploration, validation, and model results.

Vehicle safety attributes are mandatory equipment for motorcycle riders to reduce the risk of serious injury in accidents. Detecting the use of these attributes is crucial to understanding public awareness of personal safety. The completeness of safety attributes is greatly influenced by several factors, including wearing a helmet and protective clothing [3].

Image processing classification is a technique in the field of image processing and computer vision aimed at grouping images or parts of images based on specific features extracted from the images. Classification networks can significantly reduce detection time. This process involves preprocessing [4], providing a theoretical framework for understanding overfitting and underfitting. Underfitting can be caused by insufficient features or too few parameters [5]. Overfitting is often addressed using techniques such as regularization, dropout, and cross-validation [6]. Some of the processes performed in image preprocessing include annotation using Cvat (Computer Vision Annotation Tool), augmentation using Albumentations, and exploration and validation with FiftyOne.

Deep learning is a branch of machine learning that mimics the way the human brain processes information, using artificial neural networks composed of many layers. YOLO, which stands for "You Only Look Once," is an important concept found in various fields such as object detection, target identification, and localization [7]. YOLOv11 sets a new benchmark in object detection, where the backbone serves as the primary feature extractor, utilizing convolutional neural networks (CNN) to transform raw image data into multi-scale feature maps rich in spatial and contextual information [8]. The Neck strengthens feature representations by combining information from multiple scales. This component uses approaches like Feature Pyramid Networks (FPN) and Path Aggregation Networks (PANet) to ensure that spatial and contextual information from the backbone is efficiently processed. The Head improves bounding box regression accuracy and class prediction through a multi-scale approach. By supporting object detection at various scales, this head ensures the model can effectively handle objects ranging from small to large sizes.

ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) is a deep learning model based on artificial neural networks, a revolutionary form of artificial intelligence designed to generate high-resolution images from low-resolution ones with exceptional detail and quality. In this model, the generative model G captures the data distribution, while the

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

discriminative model D estimates the probability that a sample originates from the training data rather than from G [9].

Figure 2 illustrates the basic architecture of the Generative Adversarial Network (GAN) used for image resolution enhancement tasks, as seen in the Enhanced Super-Resolution GAN (ESRGAN) model.

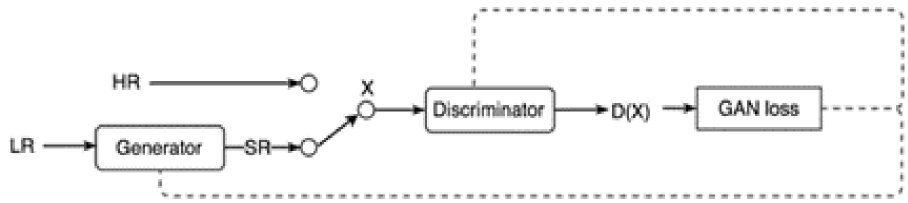


Figure 2. Network design for the generator and discriminator
Source: Long, Xinyue et al. [9]

The process begins with a low-resolution (LR) image, which is given as input to the Generator. The Generator's function is to convert the image into a high-resolution (SR) image.

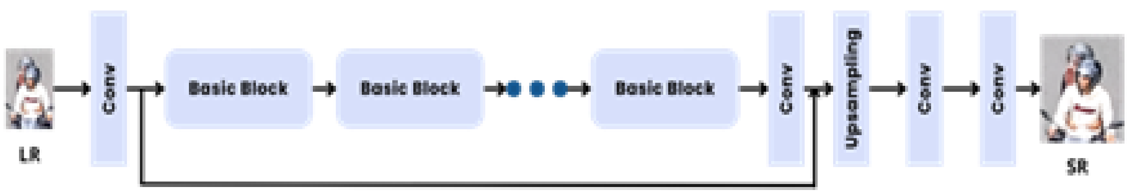


Figure 3. ESRGAN Architecture

Figure 3 illustrates the detailed architecture of the Generator and Discriminator networks in the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) model. At the top, the Generator network receives a low-resolution (LR) image as input, which is then processed through the first convolutional layer with the PReLU activation function. The image passes through several residual blocks (Residual Blocks - RB), which help preserve image details. Each residual block is equipped with skip connections to maintain the flow of information. The image is then passed through convolutional layers, batch normalization (BN), and pixel shuffler techniques to gradually upscale the image resolution. This process generates a high-resolution image (Super-Resolved Image - ISR) as the output of the generator network [10].

The confusion matrix is a tool used to evaluate the performance of classification models, applicable for both binary and multi-class classification problems. The metrics used, along with the underlying formulas, help assess the model's accuracy, precision, recall, and other performance characteristics.

Accuracy measures how often the model makes correct predictions. It is calculated by dividing the number of correct predictions by the total number of predictions made, expressed as a percentage. The formula is:

$$Accuracy = \frac{TP+TN}{TN+TP+FN+FP} \quad (1)$$

This metric is useful for assessing the overall performance of a classification model.

Precision measures the proportion of true positive predictions out of all positive predictions made by the model. It indicates how accurate the model is in identifying positive cases. The formula is:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

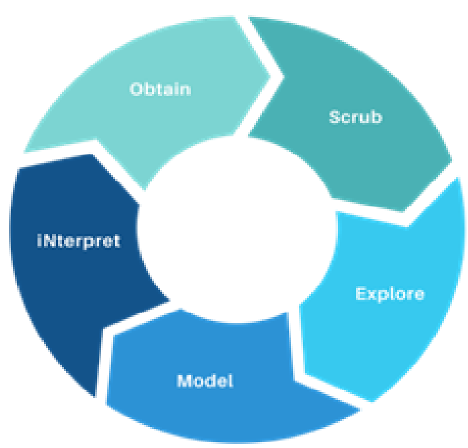
A high recall indicates that the model successfully identifies most of the positive data. Recall measures the proportion of true positive predictions out of all actual positive cases. The formula is:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score measures the harmonic mean of precision and recall, providing a balance between the two. It is especially useful when there is an uneven class distribution. The formula is:

$$F1 - Score = 2\,x\,\frac{Recall\ x\ Precission}{Recall+Precission} \quad (4)$$

The OSEMN Framework is a structured approach designed to guide the data science process. OSEMN consists of five main stages: Obtain (Collect Data), Scrub (Clean Data), Explore (Explore Data), Model (Modeling), and iNterpret (Interpret Results). The OSEMN process is iterative, meaning each stage can be repeated if necessary, especially if issues or shortcomings are found in the previous stage [11].



Gambar 4. OSEMN Life Cycle

Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 22, No. 3, Februari 2025, pp.117-131
ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

Several studies have been conducted related to attribute detection and vehicle detection, where researchers have created a variety of classification methods to address these issues. Various studies have been carried out to develop helmet detection systems using different deep learning and computer vision techniques. One of the methods used is the Single Shot Detector (SSD). In their 2023 study titled "Motorcycle Rider Helmet Violation Detection System Using SSD Algorithm," it was shown that the SSD algorithm simplifies helmet violation detection for motorcyclists. The system works with a camera that detects passing motorcycles and sends violation notifications to the police. The system achieved a mAP@50IOU of 79.2% and AR@100 of 61.4% [12].

Another study by Bambang Widodo et al. (2021) [13] used the Convolutional Neural Network (CNN) method to detect project helmet use. Testing on 90 images covering 9 scenarios yielded an F1-score of 0.79, indicating that the CNN method holds significant potential for helmet detection. This study highlighted CNN's ability to capture visual patterns more deeply than traditional methods.

Furthermore, Sri Dianing Asri et al. (2022) [14] applied a combination of the Circle Hough Transform (CHT) and Support Vector Machine (SVM) methods for vehicle wheel detection. The results showed a training data accuracy of 94.76%, a validation accuracy of 86.7%, and a test data accuracy of 67.6%.

Based on the previous studies outlined, the novelty of this research lies in the implementation of YOLOv11 and ESRGAN for vehicle attribute detection in Indonesia, focusing on the detection of small objects such as helmets and mirrors. This study proposes a technological solution using YOLOv11 and ESRGAN as a deep learning-based object detection approach.

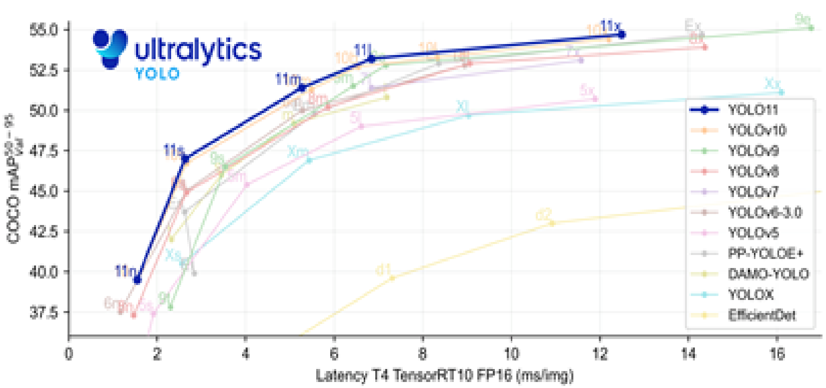


Image 5. YOLOv11 Performance
Source: Ultralytics YOLO Docs [15]

YOLOv11 shows significant improvement in object detection performance on the dataset when compared to previous YOLO versions and several other models.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

2. Methode

This study will apply the experimental method as the primary foundation. The series of experimental methods includes a sequence of stages, starting from the research workflow, image processing, model formulation, to the research tools used. Below is the experimental plan using OSEMN for the research on developing vehicle attribute detection in Indonesia using the deep learning algorithms YOLOv11 and ESRGAN.

2.1. Obtain (Collecting and Understanding Data)

The first step in this research is to collect and understand the data relevant to the problem that needs to be solved. The main issue faced is the difficulty in detecting small attributes on riders, such as helmets and mirrors, using CCTV cameras under various weather and lighting conditions.

2.2. Scrub (Cleaning the Data)

The researchers performed data cleaning to address noise and other disturbances using techniques such as outlier removal and filtering. Data from various sources were combined into a single comprehensive dataset through integration. Data transformation was carried out to meet the model's requirements, while annotation using CVAT helped label objects such as helmets and mirrors. Data augmentation using Albumentations was applied to enhance the diversity and quality of the dataset.

2.3. Exploring Data

Data exploration using FiftyOne was performed to visualize the dataset, evaluate annotations, and identify issues such as mislabeling or inconsistencies. The focus was on class distribution, small object attributes (helmets and mirrors), as well as variations in lighting, camera angles, and weather conditions. This step ensures that the dataset is representative and ready to support model training with high accuracy.

2.4. Model (Building the Model)

The modeling phase aims to develop a vehicle attribute detection system using ESRGAN and YOLOv11. RealESRGAN_x2plus [16] was selected to enhance the quality of CCTV images by improving visual details, enabling more accurate detection of small attributes such as helmets and mirrors. This supports more efficient traffic surveillance and provides high-quality data for further analysis.

The YOLOv11x model [15] was chosen for its ability to detect small objects with high precision, even in complex conditions. Through full fine-tuning, this model was adapted to a specific dataset that includes small attribute annotations under various lighting and weather conditions. With real-time speed and accuracy, the combination of ESRGAN and YOLOv11x is expected to produce an effective surveillance system that enhances road safety.

2.5. Interpreting Results

Evaluation was carried out by measuring key metrics, namely accuracy, precision, recall, and F1-score. These metrics were used to assess how accurately and consistently the model detected small objects based on the test data.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

3. Results And Discussion

This research utilized the Google Colab GPU platform based on Google Compute Engine for deploying the ESRGAN + YOLOv11 model, with system specifications of 12.7 GB RAM, 15 GB GPU RAM, and 112.6 GB disk. The model used was the pretrained YOLOv11 version yolov11x.pt, optimized to detect helmet and mirror attributes. Experiments were conducted on model configurations with 200, 300, and 400 epochs to evaluate the best balance between training, validation, and testing results.

Table 1. Data Label

No	Label	Number of Label Data
1	Helmet	648
2	Mirror	923

Table 1 shows the distribution of data based on the object labels used in this study. The dataset consists of two main categories: helmets and mirrors, with 648 data points for the helmet label and 923 for the mirror label.

The dataset is prepared for hyperparameter tuning, with the data split in a 70:15:15 ratio—165 training data, 35 validation data, and 36 testing data—to ensure balanced data proportions in each subset. The hyperparameter configuration for YOLOv11 applied by the researchers is shown in Table 4.4.

Table 2. Training Parameter Configuration

Parameter	Nilai
Model	Yolo11x.pt
Data Path	Yaml
Epochs	200,300 and 400
Batch Size	8
Initial Learning Rate (lr0)	0.001
Device	Cuda
Workers	8
Image Size (imgsz)	640
Augmentations	Active (True)
Learning Rate Scheduler	Cosine Annealing (True)
Save Model	Active (True)
Plots	Active (True)

In this study, the researchers performed training, validation, and testing on the model without ESRGAN and the model with ESRGAN and YOLOv11. The model without ESRGAN was trained with 200 epochs, while the model with ESRGAN and YOLOv11 was trained with 200, 300, and 400 epochs.

The performance evaluation results of the object detection model without ESRGAN based on YOLOv11 are presented in a table that summarizes three evaluation stages: training with 200 epochs, split validation with confidence threshold = 0.5, and split testing with confidence threshold = 0.5. The model's performance is measured using two key metrics, precision and recall, applied to three object classes: All, Helmet, and Mirror, as shown in Table 3.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

Table 3. YOLOv11 Epoch 200 Results

Class	Image	Instances	Precision	Recall
all	35	255	0,75	0,768
Helmet	34	98	0,756	0,765
Mirror	33	157	0,745	0,771
Split Val (conf=0.5)				
all	35	255	0,771	0,773
Helmet	34	98	0,791	0,794
Mirror	33	157	0,751	0,752
Split Test (conf=0.5)				
all	36	248	0,722	0,653
Helmet	36	109	0,738	0,615
Mirror	36	139	0,738	0,691

Based on the table above, the performance decline observed in the testing phase can be interpreted as a challenge in the model's generalization when applied to new data. This result still indicates that the model is capable of detecting attributes with a relatively high level of accuracy and sensitivity, especially after the training and validation processes.

In addition to evaluation through training, validation, and testing results, the performance analysis of the YOLOv11 model was also conducted using a confusion matrix for the training, validation, and testing results in Table 4.

Table 4. YOLOv11 Epoch 200 Confusion Matrix Results

Stage	Class	Helmet	Mirror	background	total
Training	Helmet	83	2	40	125
	Mirror	2	134	87	223
	Background	13	21	-	34
Validation	Helmet	74	1	19	94
	Mirror	3	118	30	151
	Background	21	38	-	59
Testing	Helmet	67	1	27	95
	Mirror	2	97	31	130
	Background	40	41	-	81

Based on the table, the evaluation results show that the YOLOv11 model has fairly good detection capability during training and validation stages, with a significant improvement in recall from training to validation. The performance drop in the testing phase indicates that the model still struggles with generalizing to new data. Misclassification of the background class remains a major challenge at all evaluation stages.

The use of ESRGAN has proven to improve the overall performance of the model. The results from various epoch configurations show that the high resolution produced by ESRGAN helps

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

the model detect small objects such as helmets and mirrors more effectively, especially in the testing data.

At epoch 200, the ESRGAN + YOLOv11 model showed performance improvement compared to the baseline (without ESRGAN). Precision and recall during training were recorded at 0.789 and 0.75, respectively, while in testing, precision and recall decreased to 0.728 and 0.675.

At epoch 400, the model's performance showed improvement in training and validation, with precision reaching 0.79 and 0.817, and recall at 0.779 and 0.761. However, in testing, precision decreased to 0.737 and recall to 0.65, indicating signs of overfitting.

The ESRGAN + YOLOv11 model at epoch 300 showed the best results, with performance evaluation analysis presented in Table 5.

Table 5. ESRGAN + YOLOv11 Epoch 300 Results

Class	Image	Instances	Precision	Recall
all	35	255	0,772	0,783
Helmet	34	98	0,769	0,837
Mirror	33	157	0,776	0,729
Split Val (conf=0.5)				
all	35	255	0,826	0,797
Helmet	34	98	0,813	0,824
Mirror	33	157	0,84	0,77
Split Test (conf=0.5)				
all	36	248	0,758	0,691
Helmet	36	109	0,733	0,706
Mirror	36	139	0,783	0,783

Overall, the ESRGAN YOLOv11 model with 300 epochs showed good performance in detecting helmet and mirror objects, especially during the validation phase, where the model's generalization capability appeared better. In addition to evaluation based on training, validation, and testing results, the performance analysis of the ESRGAN + YOLOv11 model with 300 epochs was also conducted using a confusion matrix for the training, validation, and testing results in Table 6.

Table 6. ESRGAN + YOLOv11 Epoch 300 Confusion Matrix Results

Stage	Class	Helmet	Mirror	background	total
Training	Helmet	84	4	34	122
	Mirror	2	129	56	187
	Background	12	24	-	36
Validation	Helmet	77	1	23	101
	Mirror	2	111	26	139
	Background	19	45	-	64
Testing	Helmet	77	2	26	105
	Mirror	1	95	24	120
	Background	31	42	-	73

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

Overall, the ESRGAN + YOLOv11 model with 300 epochs showed the best performance in detecting helmet and mirror objects during training, validation, and testing, with recall remaining relatively stable at each stage.

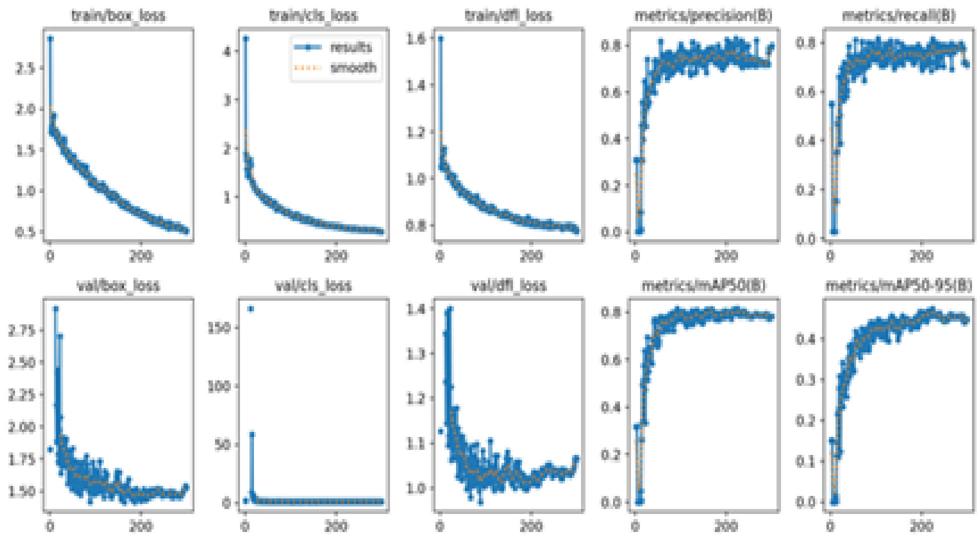


Figure 6. Results Train ESRGAN + YOLOv11 epoch 300

The model with ESRGAN at epoch 300 performed better, with improved image quality visible in theFigure 7.

The model showed the best performance in two key metrics, mAP50(B) and mAP50-95(B). For mAP50(B), the highest value of 0.81471 was achieved at epoch 92 with precision of 0.76392 and recall of 0.7877, demonstrating the model's ability to minimize false positives and detect attributes with high sensitivity. Meanwhile, for mAP50-95(B), the best result of 0.47432 was obtained at epoch 225, with precision of 0.73859 and recall of 0.81418, reflecting a balance between prediction accuracy and detection coverage at a stricter intersection-over-union (IoU) threshold. These results indicate that the combination of ESRGAN and YOLO provides reliable performance in detecting small attributes, such as helmets and mirrors, on images with varying quality.

Figure 7. Comparison of Helmet and Mirror Detection Results
(a) Without ESRGAN, (b) With ESRGAN

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

The baseline results show that YOLOv11 without ESRGAN has lower accuracy on the test data, making ESRGAN an essential component in improving the quality of input to maximize the model's performance, as seen in the output data of Figure 7.

4. Conclusion

This research has successfully implemented the detection of motorcyclist attributes, namely helmet and mirror, using deep learning methods by utilizing YOLOv11 and ESRGAN. Based on the evaluation results across various configurations, the ESRGAN + YOLOv11 model with 300 epochs was chosen as the best configuration due to its balanced performance between training, validation, and testing. This model achieved a precision of 75.8% and a recall of 69.1% on the test data, reflecting good generalization capabilities without overfitting.

The use of ESRGAN as a resolution enhancement method has proven to contribute significantly to improving YOLOv11 detection accuracy, particularly for small visual attributes like mirrors. Compared to the model without ESRGAN, ESRGAN + YOLOv11 showed consistent performance improvements at all evaluation stages. The baseline results without ESRGAN indicated that the low resolution of input data affected the model's ability to detect small attributes, which was addressed with the use of ESRGAN.

Evaluation across various epochs showed a performance improvement trend up to epoch 300, where precision and recall in training, validation, and testing remained stable. However, at epoch 400, the model began to show signs of overfitting, where training and validation performance increased, but the results on the test data decreased. This indicates that selecting the right number of epochs is a key factor in maintaining a balance between training accuracy and the model's generalization ability.

Furthermore, challenges in model implementation, such as higher inference time due to the use of ESRGAN, were addressed by optimizing the model using quantization techniques. Initial testing showed that the system was able to process 15-20 images per second with accuracy consistent with the evaluation results, supporting the application of the model in real-world scenarios.

Overall, this research demonstrates that the combination of ESRGAN and YOLOv11 can be effectively used to detect helmet and mirror attributes in vehicle images. With an optimal configuration, this model not only provides good accuracy but also has the potential for implementation in various AI-based surveillance applications, such as traffic security systems or vehicle analysis. This research opens up opportunities for further development, such as integration with edge computing for real-time inference or adapting the model to detect other attributes.

Future research is recommended to expand the dataset scope by using CCTV cameras from various locations and angles, as well as adding variations in conditions such as weather, time, and camera resolution to improve the model's generalization ability. A multi-label classification approach could be applied to detect multiple attributes simultaneously, improving system efficiency. Additionally, testing the system in real-world environments, such as highway cameras, will provide further insights into its performance under dynamic conditions. The research could also explore additional technologies, such as combining

YOLOv11 with Transformer-based models, to detect more complex attributes simultaneously with higher accuracy and efficiency.

Daftar Pustaka

[1] N. Jegham, C. Y. Koh, M. Abdelatti dan A. Hendawi, "Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors," *arXiv,* vol. 1, pp. 1-20, 31 Oktober 2024.

[2] Y. Xu, W. Luo, A. Hu, Z. Xie, X. Xie dan L. Tao, "TE-SAGAN: An Improved Generative Adversarial Network for Remote Sensing Super-Resolution Images," *Remote sensing,* vol. 14, pp. 1-17, 18 Mei 2022.

[3] R. Anggraini, A. Alvisyahri dan S. Sigiarto, "Persepsi Keselamatan Berkendara Pengguna Sepeda Motor di Kota Banda Aceh terhadap Pelanggaran Lalu Lintas dan Kelengkapan Atribut," *Jurnal Teknik Sipil: Jurnal Teoritis dan Terapan Bidang Rekayasa Sipil,* vol. 28, no. 3, pp. 329-336, 23 September 2022.

[4] E. A. Aldakheel, M. Zakariah dan A. H. Alabdalall, "Detection and identification of plant leaf diseases using YOLOv4," *Frontiers in Plant Science,* vol. 15, pp. 1-22, 22 April 2024.

[5] E. S. Budi, A. N. Chan, P. P. Alda dan M. A. F. Idris, "Optimasi Model Machine Learning untuk Klasifikasi dan Prediksi Citra Menggunakan Algoritma Convolutional Neural Network," *RESOLUSI : Rekayasa Teknik Informatika dan Informasi,* vol. 4, no. 5, pp. 502-509, Mei 2024.

[6] SkillPlus, "Underfitting dan Overfitting Model," 2 Agustus 2019. [Online]. Available: https://skillplus.web.id/underfitting-dan-overfitting-model/. [Diakses 15 Desember 2024].

[7] H. Luo, F. Gao, K. Hai, M. Shaodan dan H. V. Poor, "YOLO: An Efficient Terahertz Band Integrated Sensing and Communications Scheme with Beam Squint," *arXiv,* vol. 3, pp. 1-16, 6 Februari 2024.

[8] A. F. Rasheed dan M. Z. M. Zarkoosh, "YOLOv11 Optimization for Efficient Resource Utilization," *arXiv,* vol. 2, pp. 1-12, 21 Desember 2024.

[9] X. Long dan M. Zhang, "An Overview of Generative Adversarial Networks," *Journal of Computing and Electronic Information Management,* vol. 10, no. 3, pp. 31-36, 2023.

[10] X. Wang, L. Xie, C. Dong dan Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops,* pp. 1905-1914, 2021.

[11] F. T. Br Sitepu, V. A. Prada Sirait dan R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Prophet Facebook," *Paradigma,* vol. 23, no. 1, pp. 99-105, Maret 2021.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i3.xxxx

[12] F. Fuadi, C. Setianingsih dan M. W. Paryasto, “Sistem Deteksi Pengendara Sepeda Motor Tanpa Helmet Menggunakan Algoritma SSD,” *e-Proceeding of Engineering,* vol. 10, no. 1, pp. 823-830, Februari 2023.

[13] B. Widodo, H. Armanto dan E. Setyati, “Deteksi Pemakaian Helmet Proyek Dengan Metode Convolutional Neural Network,” *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION,* pp. 23-29, 2021.

[14] S. D. Asri, D. Ramayanti, A. D. Putra dan Y. T. Utami, “DETEKSI RODA KENDARAAN DENGAN CIRCLE HOUGH TRANSFORM (CHT) DAN SUPPORT VECTOR MACHINE (SVM),” *JURNAL TEKNOINFO,* vol. 16, no. 2, pp. 427-434, Juli 2022.

[15] Ultralytics, “Ultralytics YOLO11,” 30 September 2024. [Online]. Available: https://docs.ultralytics.com/models/yolo11/. [Diakses 17 Desember 2024].

[16] xinntao, “Real-ESRGAN,” Github, 20 09 2022. [Online]. Available: https://github.com/xinntao/Real-ESRGAN. [Diakses 15 Desember 2024].