Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Enhancing YOLOv5s with Attention Mechanisms for Object Detection in Complex Backgrounds Environment

Ali Impron[1], Dina Lestari[2], Linda Sutrian[3], Syadza Anggraini[4], Randi Rizal[2,5]

[1,3,4]*Departement of Informatics, Faculty of Engineering and Agriculture, Universitas Muhammadiyah Sampit, Central Kalimantan, Indonesia*
[2]*Department of Informatics, Faculty of Engineering, Universitas Siliwangi, Kota Tasikmalaya, Indonesia*
[5]*Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Melaka, Malaysia*

[1]*ali.impron@gmail.com*, [2]*lestaridina096@gmail.com*, [3]*lindasutriani19@gmail.com*, [4]*anggrainisyadza@gmail.com*, [5]*randirizal@utem.edu.my*

ARTICLE INFORMATION

Article History:
Received: September 11, 2025
Last Revision: November 16, 2025
Published *Online*: November 30, 2025

KEYWORDS

Attention Mechanism,
C3CBAM,
Complex Environment,
Object Detection,
YOLOv5s

CORRESPONDENCE

Phone: 081347595678
E-mail: ali.impron@gmail.com

ABSTRACT

Enhancing performance for object detection in complex environments is essential for real-world applications that involve challenges such as overlapping or stacked objects within the same scene. Existing object detection models still face difficulties when dealing with complex backgrounds, as accuracy often decreases when the target objects are small or occluded by others. Therefore, this study proposes an improvement method to enhance detection performance in complex environments using the YOLOv5s algorithm. The optimization involves integrating a CBAM (Convolutional Block Attention Module) with the C3 layer (C3CBAM) in the backbone and adding a P2 feature map in the head of the YOLOv5s architecture. The proposed modifications yield promising results, with precision increasing by 1.6%, mAP@0.5 improving by 1.4%, and mAP@50–95 increasing by 0.1%, proving that the applied enhancements effectively improve model performance. However, the addition of the attention mechanism also increases computational load. Future work should focus on reducing computation, for example, through knowledge distillation, where a lightweight model is trained to mimic the behavior of a larger one. This study contributes to improving object detection performance under real-world occlusion and background complexity, enabling more reliable deployment in practical visual recognition systems.

1. INTRODUCTION

Object detection in environments with complex backgrounds remains a challenging task in computer vision [1], [2]. A complex environment occurs when target objects are surrounded by various irrelevant items, diverse colors, and inconsistent lighting [2]. Such conditions are commonly found in real-world applications, including autonomous vehicles, smart surveillance systems, robot navigation, and highway CCTV. [3]. Variations in viewing angles, distances, and object occlusion often reduce the precision of object detection [4], making this issue still an active research topic.

Recent studies have enhanced detection performance in complex environments through architectural modifications of the YOLO family models [5]. For instance, NLE-YOLO improved the receptive field and feature extraction capability of YOLOv5, showing strong results under low-light conditions [6]. Another work enhanced YOLOv5 by integrating a novel attention mechanism and modifying the backbone and head, which improved detection of small, distant object [7]. A two-stage meta-learning model based on YOLO was also proposed to reduce detection errors in multi-capture scenarios [8].

The attention mechanism a module that adaptively focuses on informative spatial and channel features while suppressing irrelevant background noise, it's become a widely used enhancement strategy. For example, YOLOv5 integrated with the C3GAM attention module achieved 90.29% mAP on the High-Spatial-Resolution remote sensing dataset [9]. Similarly, the addition of Coordinate Attention (CA) blocks and modified CSP bottlenecks improved YOLOv5's accuracy by 3.2% and 1.7% for mAP50 and mAP50-95, respectively [10] . However, some models still exhibit limitations such as slower inference speed and reduced robustness. Further enhancement, such as integrating attention in YOLOv8n for small object

refinement with Soft-NMS, improved results but remained sensitive to occlusion [11].

Based on these findings, this study proposes an enhanced YOLOv5s architecture that integrates a C3CBAM module with P2 feature maps to strengthen small-object detection performance, particularly under occlusion and cluttered backgrounds. The City Persons dataset [12], obtained from the Roboflow platform, is used for model training and validation due to its complex urban scenes and varying illumination conditions. In addition, model generalization is evaluated using benchmark datasets including COCO2014, PASCAL VOC 2007, BDD100K, and UAV. Unlike previous approaches that applied attention mechanisms only to deep feature layers, this study focuses on enriching low-level spatial representations through the C3CBAM–P2 integration, thereby enhancing the model's ability to detect small and partially occluded objects while maintaining high computational efficiency.

In summary, this research modifies the YOLOv5s architecture, embeds the proposed attention module, and evaluates its performance across multiple datasets to demonstrate improvements in both accuracy and robustness under complex background conditions.

2. RELATED WORK

Object detection has become one of the core problems in computer vision, aiming to locate and classify objects within an image or video through bounding box predictions. The evolution of deep learning–based detectors has significantly improved detection accuracy and speed, especially after the introduction of the You Only Look Once (YOLO) algorithm by Joseph Redmon in 2015 [13].

Over time, YOLO underwent substantial technical evolution to improve its efficiency and precision. YOLOv2 introduced batch normalization and anchor boxes, while YOLOv3 added multi-scale prediction using a Feature Pyramid Network (FPN). YOLOv4 incorporated advanced optimization and augmentation techniques (Mosaic, Mish activation, and CSPDarknet backbone), which boosted performance across various tasks. YOLOv5, developed by Ultralytics in 2020 using the PyTorch framework, further emphasized modularity, lightweight deployment, and faster training with better balance between accuracy and inference time.Successive versions such as YOLOv6–YOLOv8 focused on industrial deployment, introducing dynamic heads, anchor-free detection, and improved feature extraction via BiFPN and PAN-FPN structures. These enhancements significantly improved inference speed, reduced computational overhead, strengthened multi-scale object representation, and optimized real-time performance across edge devices and large-scale production environments [14].

Several studies from 2021–2024 have specifically explored enhancing YOLOv5 to address weaknesses in small-object detection, occlusion handling, and background complexity. For instance, some works modified the backbone and neck structures by inserting attention modules such as CBAM (Convolutional Block Attention Module) [15], ECA (Efficient Channel Attention) [16], and CA (Coordinate Attention) [17] to refine feature extraction. Others introduced dynamic convolution blocks, cross-stage feature fusion, or multi-head detection branches to increase robustness and speed in real-time detection tasks. These efforts consistently demonstrated that targeted architectural modifications could improve mAP50–95 scores while maintaining lightweight model parameters.

The attention mechanism itself plays a critical role in enhancing YOLO's ability to focus on meaningful visual cues. It can be categorized into several types (channel attention, spatial attention, coordinate attention) [18]. Channel Attention, which reweights feature channels based on their importance, improving the network's sensitivity to relevant semantic information. Spatial Attention, which emphasizes significant regions within the image, helping the model distinguish target objects from cluttered backgrounds. Coordinate Attention, which embeds precise positional encoding into feature maps, allowing YOLO to better capture long-range dependencies and object layouts.

Between 2021 and 2024, numerous researchers proposed various improvements to YOLOv5 for different application domains and environmental conditions. The study [19] proposes an improved YOLOv5 architecture for real-time object detection in vehicle-mounted camera scenarios by integrating a GS-FPN structure combining Bi-FPN, CBAM, and GSConv to enhance multi-scale feature extraction. Experimental results show up to a 12.2% mAP improvement on small traffic sign detection, demonstrating higher accuracy and robustness in complex driving environments. [20] SE (Squeeze-and-Excitation) channel-attention module into a backbone built on the lightweight ShuffleNetV2 network within the YOLOv5 framework to improve line selection accuracy for small-current grounding faults under noisy conditions. They report a detection accuracy of 93.6% and real-time inference speed of 122 fps (image resolution 640×640) even with limited real-fault data and under noisy environments.

These collective findings indicate that integrating attention modules into YOLOv5 significantly boosts detection performance, yet many approaches still struggle with the trade-off between accuracy, inference efficiency, and robustness under illumination changes or occlusion. Integrating these attention modules into YOLO's backbone improves hierarchical feature extraction, while insertion into the neck enhances feature fusion between layers. Such integration leads to better generalization in scenarios involving occlusion, lighting variation, or complex urban scenes.

3. METHODOLOGY

This research aims to improve the performance of the YOLOv5s model on object detection tasks in environments with complex backgrounds, enhancement is done using attention mechanism. It can be seen in Figure 1 for the initial step taken is to conduct a literature study on research related to the topic raised, then search for or collect datasets that will be used to train the model. Furthermore, the process of developing the YOLOv5s model by adding an attention mechanism layer. Next, the model that has been developed can be evaluated using the model evaluation

matrix parameters. Finally, the enhanced model can be used to detect objects in complex environments.

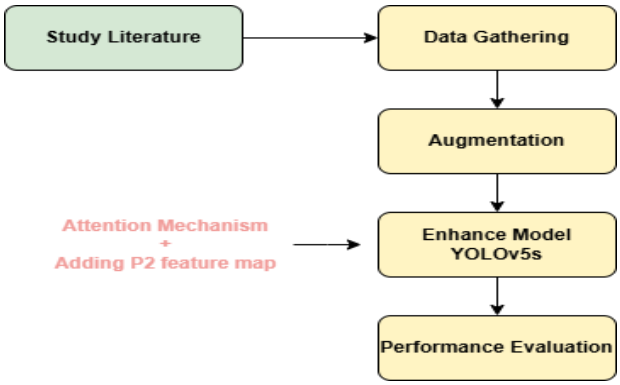

FIGURE 1. RESEARCH METHOD

3.1 Studi Literature

The first thing that is done in this research is a literature study. Searching, reading, and understanding relevant topics, namely, how to improve with attention mechanisms and other improvement methods for object detection in complex environments. The references taken come from accredited journals, thus producing quality knowledge as a strong foundation for this research. This step is very useful for conducting research, because by knowing various studies with relevant topics can provide broad insight into the things that will be done during the research.

3.2 Data Gathering

In making an object detection model, it is essential to prepare a dataset that will be used to train and evaluate the model. In this study, the dataset used is called CityPersons, obtained from the open-source platform Roboflow. The dataset contains a total of 3,475 images, which are divided into 80% (2,792 images) for training, 10% (342 images) for validation, and 10% (341 images) for testing. CityPersons was chosen because it provides diverse urban street scenes with various levels of occlusion, scale variation, and complex backgrounds, which closely resemble real-world conditions encountered in pedestrian detection tasks. The dataset includes challenging instances such as partially visible pedestrians, overlapping objects, and varying lighting conditions, making it suitable for evaluating the robustness of object detection models in complex environments. In Table I, the dataset's class distribution is presented, consisting of three labels: ignore, none, and ped and the detailed division of training, validation, and test sets is provided.

TABLE 1. DATASET SUMMARY

Class	Image Size	Data Division	Proportion	Amount
Ignore	640x640px	Train	80%	2792
None	640x640px	Validation	10%	342
Ped (Pedestrian)	640x640px	Test	10%	342

To ensure optimal model performance, several training configurations and hyperparameters were carefully selected during the model development stage. These parameters were determined based on empirical testing to achieve a balance between detection accuracy, computational efficiency, and convergence stability. The detailed configuration used for training the enhanced YOLOv5s model is presented in Table II.

TABLE II. TRAINING CONFIGURATION

Parameter	Value/Setting
Epochs	50
Batch Size	32
Momentum	0.937
Learning Rate	0.01
Input Image Size	640
GPU	NVIDIA A100-SXM4-40GB

3.3 Augmentation

To minimize the risk of overfitting during model training, an augmentation phase was performed on the dataset. The augmentation techniques used can be seen in Table III. The application of augmentation techniques aims to enrich the distribution of training data so that the model can recognize objects more accurately on test data that has never been seen before, thereby mitigating overfitting in the model.

TABLE III. AUGMENTATION TECHNIQUE

Technique	Description
Outputs per training example: 2	Each training image produces two augmented versions.
90° Rotate	Rotates the image 90° clockwise or counterclockwise.
Rotation	Randomly rotates the image between -45° and +45°.
Grayscale	Converts 15% of images to black-and-white.
Brightness	Adjusts brightness between -50% (darker) and +50% (brighter).
Blur	Applies a blur effect of up to 2.5 pixels.

3.4 Enhance Model YOLOv5s

This stage is the core part of the research, namely development of the model to be improved. Model creation using YOLOv5s with integrated C3 layer and attention mechanism module. For the integrated attention mechanism is CBAM (Convolutional Block Attention Module). In Fig. 2, the outline of the YOLOv5s structure can be seen consisting of backbone, neck, and head. The backbone includes various modules, namely Conv, C3, and SPPF. The outline of the YOLOv5s structure consists of backbone, neck, and head. The backbone includes various modules, namely Conv, C3, and SPPF. By combining the Conv and C3 modules, the model is increasingly optimal in extracting features. The SPPF module utilizes several small-sized pooling cores in sequence to replace the single large-sized pooling core in the previous SPPF module, significantly improving the feature extraction capability.

This improvement allows the model to detect objects of various sizes and speed up its processing performance. In the neck section, the YOLOv5 model adopts the PANet structure and adds a bottom-up enhancement path to the top-down feature pyramid, which strengthens the feature fusion capability in the network. The head section consists of three detection layers, each corresponding to three feature map sizes generated from the neck. Based on the feature map sizes, a grid is divided, and each grid is equipped with three anchors with different aspect ratios for the purposes of target prediction [24].

Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

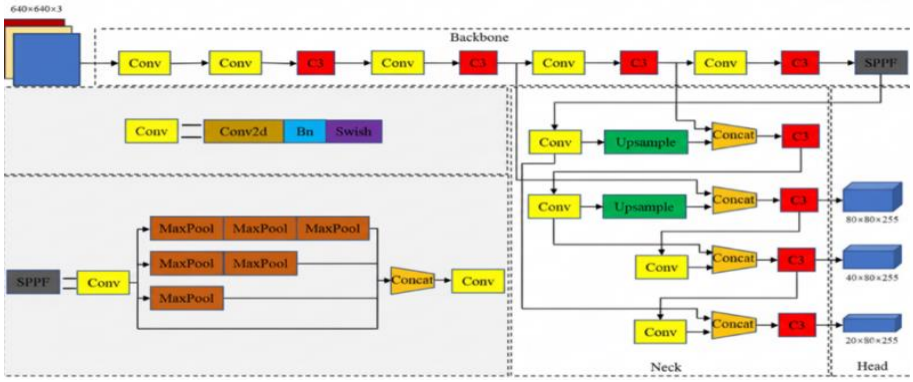

FIGURE 2. YOLOV5 ARCHITECTURE [24]

In the field of computer vision, attention mechanism plays an important role. This mechanism allows neural networks to automatically generate masks, so that the network can learn and focus on relevant areas. By performing weighted iterations based on the mask score, the influence of the focus area can be increased, while the weight of irrelevant information is reduced, which ultimately optimizes the performance of the network model. Based on the location of the mask generation, attention mechanisms are generally divided into three types: channel attention mechanism, spatial attention mechanism, and mixed-domain attention mechanism. In this study, we use a module developed with the mixed-domain attention mechanism technique as shown in Fig. 3, namely the CBAM module [25]. The CBAM module includes channel attention mechanism (CAM) and spatial attention mechanism (SAM). The working process of the CBAM module is with input features FϵR^(C*H*W) processed by the CAM module to produce M_c ϵR^(C*1*1) which is the channel weight vector. Then, M_c multiplied by F, thus producing weighty features F'. Furthermore F' entered the SAM module to produce a spatial weight matrix M_s ϵR^(1*H*W). Then, lastly M_s multiplied by F^', generate spatial feature weights F^".

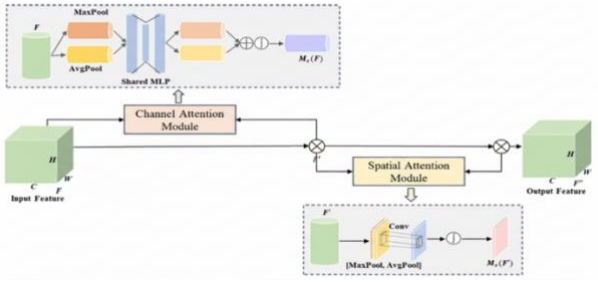

FIGURE 3. CBAM MODULE STRUCTURE CBAM [24]

The CAM module focuses on important channel features and ignores unimportant ones, allowing the model to pay more attention to effective spatial information. By using parallel pooling operations and MLP processing, the CAM module can produce more accurate and effective output results. The SAM module in the model can focus more on important areas in the feature map. By using parallel pooling operations and convolution layers, the SAM module can produce more accurate and effective output results in capturing relevant spatial information. By using the CBAM module, the network can focus more on important features and the C3 and CBAM modules can improve the network's ability to capture relevant features and improve detection performance.

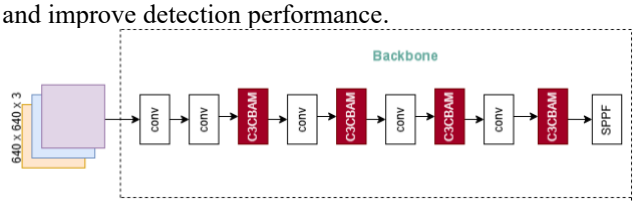

FIGURE 4. ARCHITECTURE YOLOV5S + C3CBAM IN BACKBONE

In developing its model, this study changed the C3 layer on Yolov5s with C3CBAM in the backbone section, the aim being to extract deeper features in the data. In Fig. 4, the YOLOv5s architecture has been modified. In addition to changing the backbone, to answer the challenge of the difficulty of detecting small objects due to the object being far from view, the P2 feature map was added to the head of the YOLOv5s architecture. In Fig. 5, the architecture has been added with a new feature map, namely P2.

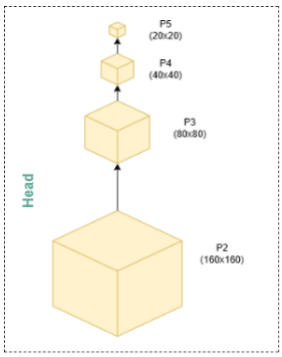

FIGURE 5. P2 FEATURE MAP

3.5 Performance Evolution

This study aims to improve the YOLOv5s model, and to test the effectiveness of the model, several evaluation metrics commonly used in the field of object detection are used, such as precision, recall, and mAP. By using these metrics, the performance of the model can be assessed and compared with other models. True positive (TP) is when the model successfully detects an object that is in the image. False positive (FP) is when the network incorrectly detects an object that is not in the image. False negative (FN) is when the network fails to detect an object that is in the image.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{c} \quad (2)$$

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

4. RESULT AND DISCUSSION

4.1 Quantitative and Evaluation

In order to comprehensively assess the effectiveness of the proposed approach, comparative experiments were conducted. Specifically, to evaluate whether the integration of an attention mechanism can enhance the performance of the YOLOv5s model, a comparison was made between the baseline YOLOv5s and the modified version incorporating the attention mechanism layer. Table IV presents the evaluation results of both models using the same dataset, namely CityPersons.

TABLE IV. COMPARISON MODEL YOLOV5S WITH YOLOV5S+C3CBAM

Model	Precision	Recall	mAP50	mAP50-95	Latency	FPS
YOLOv5s	77.5%	**53.8%**	63.2%	36.1%	**7.9** ms	**102**
YOLOv5s C3CBAM + P2 **(ours)**	**79.1%**	52.3%	**64.6%**	**36.2%**	9.8 ms	126

A comprehensive evaluation of the architectural modification integrating P2 in the Head and C3CBAM into YOLOv5s reveals notable changes in detection performance, as summarized in Table X. Precision increased from 77.5 % to 79.1 % (+1.6 %), indicating that the model produces more accurate predictions with fewer false positives. However, recall decreased by 1.5 %, suggesting a trade-off where some true objects were missed as the model became more selective. mAP@50 improved by 1.4 %, reflecting better overall object detection accuracy, while mAP@50–95 showed only a marginal improvement of 0.1 %.

Architecturally, integrating P2 in the detection head provides higher-resolution spatial information from shallow backbone layers, enhancing sensitivity to small objects and contributing to the gain in mAP. Meanwhile, C3CBAM, as an attention mechanism, strengthens feature selectivity by emphasizing the most informative spatial and channel features, which explains the rise in precision. However, the slight drop in recall occurs because CBAM's selective weighting tends to suppress weaker activations from partially occluded or low-contrast objects. This means the model focuses more narrowly on prominent targets while missing subtle detections an inherent trade-off between precision enhancement and detection coverage.

From a computational perspective, the modified YOLOv5s shows a minor latency increase of 1.9 ms, indicating slightly longer processing per frame. Interestingly, the frame rate increased by 24 FPS, suggesting that despite the added architectural complexity, the optimized P2 integration and C3CBAM's efficient feature refinement improve throughput in batch or continuous inference pipelines. In practice, this small latency penalty is acceptable given the significant gain in overall inference efficiency and precision, which is particularly beneficial for real-time detection applications.

4.2 Cross-Dataset Generalization

The comparison in Table VI demonstrates that integrating P2 in the Head and C3CBAM into YOLOv5s produces dataset-dependent gains: notable improvements are observed on CityPersons and UAV datasets, particularly in precision and mAP, where small-object detection and occlusion are dominant. However, performance declines on large-scale and heterogeneous datasets such as COCO 2014, PASCAL VOC 2007, and BDD100K, revealing a trade-off between specialized accuracy and generalization capability across diverse environments.

TABLE VI. COMPARISON USING BENCHMARK DATASETS

Dataset	YOLOv5s				YOLOv5s with P2 in Head + C3CBAM (ours)			
	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)
CityPersons	77.5	53.8	63.2	36.1	79.1	52.3	64.6	36.2
COCO 2014	34.2	33.0	56.0	37.2	42.4	2.5	2.6	1.19
PASCAL VOC 2007	47.3	45.0	79.0	51.0	48.0	43.2	42.8	19.6
BDD100K	62.3	30.5	49.3	26.0	65.8	31.3	36.3	18.9
UAV	70.6	69.8	72.9	40.0	74.1	69.8	73.6	40.4

Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

4.3 Qualitative Analysis

Overall improved performance, the model shows a significant increase in performance during the training process. To see a more real comparison, see Fig. 6 which is the prediction result of each model. The prediction results of the modified yolov5s by adding C3CBAM to the backbone and head that adds P2 are more 'sensitive' to objects that look small because of the long distance and in a complex environment, with objects that block the target object (pedestrian class). For objects that are far from the reach of our model, it detects with a higher confidence value compared to the usual YOLOv5s model.

The modified YOLOv5s, enhanced by integrating the C3CBAM module into both the backbone and head and by adding the P2 detection layer, exhibits higher sensitivity to small-scale objects caused by long-distance scenarios and complex environments with occlusions, particularly for the pedestrian class. Moreover, for objects located at farther distances, the proposed model achieves higher confidence scores compared to the standard YOLOv5s, indicating improved feature representation.

FIGURE 6. COMPARISON OF EXPERIMENTAL RESULTS FROM THE TWO MODELS

5. CONCLUSIONS

This research demonstrates that integrating the C3CBAM attention mechanism into the backbone and adding the P2 feature map to the head of YOLOv5s can effectively improve object detection performance in complex environments. Precision increased by 1.6%, mAP@0.5 by 1.4%, and mAP@50–95 by 0.1%, indicating more accurate and consistent detections. Moreover, the modified model shows better capability in detecting small and partially occluded objects, as evidenced by its ability to identify more targets with higher confidence in visual comparisons.

REFERENCES

[1] H. Da, "Complex Environment Road Object Detection Algorithm Based on Improved YOLOv5s," in *2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, 2024, pp. 625–630. doi: 10.1109/DOCS63458.2024.10704511.

[2] J. Zhong, Q. Cheng, X. Hu, and Z. Liu, "YOLO Adaptive Developments in Complex Natural Environments for Tiny Object Detection," *Electronics (Switzerland)*, vol. 13, no. 13, Jul. 2024, doi: 10.3390/electronics13132525.

[3] J. Ruan, H. Cui, Y. Huang, T. Li, C. Wu, and K. Zhang, "A review of occluded objects detection in real complex scenarios for autonomous driving," *Green Energy and Intelligent Transportation*, vol. 2, no. 3, p. 100092, 2023, doi: https://doi.org/10.1016/j.geits.2023.100092.

[4] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020, doi: 10.1109/ACCESS.2020.2983149.

[5] C. Baoyuan, L. Yitong, and S. Kun, "Research on Object Detection Method Based on FF-YOLO for Complex Scenes," *IEEE Access*, vol. 9, pp. 127950–127960, 2021, doi: 10.1109/ACCESS.2021.3108398.

[6] D. Peng, W. Ding, and T. Zhen, "A novel low light object detection method based on the YOLOv5 fusion feature enhancement," *Sci Rep*, vol. 14, no. 1, p. 44, 2024, doi: 10.1038/s41598-024-54428-8.

[7] W.-Y. Hsu and W.-Y. Lin, "Adaptive Fusion of Multi-Scale YOLO for Pedestrian Detection," *IEEE Access*, vol. 9, pp. 110063–110073, 2021, doi: 10.1109/ACCESS.2021.3102600.

[8] X. Ren, W. Zhang, M. Wu, C. Li, and X. Wang, "Meta-YOLO: Meta-Learning for Few-Shot Traffic Sign Detection via Decoupling Dependencies," *Applied Sciences*, vol. 12, no. 11, 2022, doi: 10.3390/app12115543.

[9] F. Cao *et al.*, "An Efficient Object Detection Algorithm Based on Improved YOLOv5 for High-Spatial-Resolution Remote Sensing Images," *Remote Sens (Basel)*, vol. 15, no. 15, Aug. 2023, doi: 10.3390/rs15153755.

[10] Y. Li, M. Zhang, C. Zhang, H. Liang, P. Li, and W. Zhang, "YOLO-CCS: Vehicle detection algorithm based on coordinate attention mechanism," *Digit Signal Process*, 2024, doi: 10.1016/j.dsp.2024.104632.

[11] Q. Su and J. Mu, "Complex Scene Occluded Object Detection with Fusion of Mixed Local Channel Attention and Multi-Detection Layer Anchor-Free Optimization," *Automation*, vol. 5, no. 2, pp. 176–189, Jun. 2024, doi: 10.3390/automation5020011.

[12] C. Conversion, "Citypersons Dataset," Dec. 2022, *Roboflow*. [Online]. Available: https://universe.roboflow.com/citypersons-conversion/citypersons-woqjq

[13] J. R. Terven and D. M. Cordova-Esparza, "A COMPREHENSIVE REVIEW OF YOLO ARCHITECTURES IN COMPUTER VISION: FROM YOLOV1 TO YOLOV8 AND YOLO-NAS PUBLISHED AS A JOURNAL PAPER AT MACHINE LEARNING AND KNOWLEDGE EXTRACTION."

[14] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions," Mar. 2025, [Online]. Available: http://arxiv.org/abs/2411.00201

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.06521

[16] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.03151

[17] Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.02907

[18] M.-H. Guo *et al.*, "Attention Mechanisms in Computer Vision: A Survey," Nov. 2021, doi: 10.1007/s41095-022-0271-y.

[19] Z. Ren, H. Zhang, and Z. Li, "Improved YOLOv5 Network for Real-Time Object Detection in Vehicle-Mounted Camera Capture Scenarios," *Sensors*, vol. 23, no. 10, May 2023, doi: 10.3390/s23104589.

[20] S. Hao, W. Li, X. Ma, and Z. Tian, "SSE-YOLOv5: a real-time fault line selection method based on lightweight modules and attention models," *J. Real-Time Image Process.*, vol. 21, no. 4, May 2024, doi: 10.1007/s11554-024-01480-2.

AUTHORS

Ali Impron

He is a lecturer and researcher at Universitas Muhammadiyah Sampit, Indonesia. His work focuses on IoT, machine learning, smart mining, and cloud computing, with publications spanning modern IT

analysis, IoT-enabled smart mining, energy consumption modeling, ERP systems, and AI applications.

Dina Lestari
She is a researcher at Universitas Siliwangi, Indonesia, with academic interests in informatics and intelligent systems. She has published severalstudies in national journals, focusing on the development and application of intelligent computing to support advancements in computer science and digital innovation

Linda Sutriani
She is a lecturer at Universitas Muhammadiyah Sampit, Indonesia. Her academic interests span IoT, smart mining, artificial intelligence, data warehousing, and information security. She has co-authored works on IoT-enabled wastewater management in coal mining, AI applications in deforestation and sustainability, sentiment analysis on palm oil plantations, and ETL data warehouse performance.

Syadza Anggraini
She is a lecturer at Universitas Muhammadiyah Sampit with expertise in information retrieval, text mining, and computer vision. Her works include research on multi-document summarization, lexical and semantic similarity for Arabic documents, sentiment analysis on palm oil plantations, and bibliometric studies on AI and agribusiness.

Randi Rizal
He is an academic affiliated with Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Melaka, Malaysia. His research interests include Digital Forensics, Information Security, Cybersecurity, and AI Forensics, with a growing citation record of over 320 citations and an h-index of 9 since 2020. He has co-authored several notable works such as Network Forensics for Detecting Flooding Attack on IoT Devices and Enhanced Readiness Forensic Framework for IoT Investigation Based on Artificial Intelligence.