IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3299~3308
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3299-3308

Enhancing pedestrian detection in adverse conditions using YOLOv5s with adaptive weighted fusion efficient channel attention module

Oumayma Rachidi, Badr Bououlid Idrissi, Chafik Ed-Dahmani
Department of Electromechanical Engineering, School of Arts and Crafts, Moulay Ismail University, Meknes, Morocco

Article Info

Article history:

Received Jun 6, 2025
Revised May 19, 2026
Accepted Jun 19, 2026

Keywords:

Advanced driver assistance system
Computer vision
Efficient channel attention
Pedestrian detection
You only look once version 5

ABSTRACT

Pedestrian detection constitutes a critical task within advanced driver assistance systems (ADAS), where reliable identification of pedestrians is essential for ensuring vehicular safety. Although deep learning has substantially improved detection performance, existing state-of-the-art models continue to exhibit notable degradation in adverse weather conditions and low-light. To mitigate these challenges, this study introduces an enhanced pedestrian detection framework based on you only look one version 5 (YOLOv5s), retrained on an augmented common object in context (COCO) dataset focused on the person class. Additionally, a novel, lightweight, and adaptive attention mechanism called: the weighted fusion efficient channel attention (WF-ECA) module is incorporated into the detection architecture. The WF-ECA module selectively focusses on important features without compromising computational efficiency or inference speed. Comparative experiments demonstrate a 5% increase in mean average precision (mAP) in comparison to the baseline model, thereby demonstrating the efficacy of the proposed attention module in improving detection robustness under challenging environmental conditions. These findings highlight the potential of attention-based mechanisms to enhance pedestrian detection performance in real-world ADAS applications.

Corresponding Author:

Oumayma Rachidi
Department of Electromechanical Engineering, School of Arts and Crafts, Moulay Ismail University
Marjane 2, BP 15290, Al-Mansour, Meknes, Morocco
Email: oum.rachidi@edu.umi.ac.ma

1. INTRODUCTION

In the last few years, advanced driver assistance systems (ADAS) [1] have gained significant attention due to their crucial role in enhancing vehicle safety. These systems utilize a wide range of functionalities in order to address various challenges encountered during driving. Among these functionalities, pedestrian detection emerges as a vital component, due to the rising incidence of accidents involving pedestrians. Consequently, important research efforts have been directed toward enhancing the accuracy and speed of pedestrian detection systems.

Initially relying on traditional approaches such as histogram of oriented gradients (HOG) [2] combined with support vector machines (SVM) [3], pedestrian detection has seen notable advancements with the advent of deep learning [4]. Despite these advancements, pedestrian detection continues to present considerable challenges, particularly in low-light environments and under adverse weather conditions [5]. For example, night-time driving significantly reduces image contrast, making pedestrian silhouettes less distinguishable from the background. Heavy rain can introduce motion blur and reflective road surfaces that

distort visual features. Similarly, dense fog reduces visibility range and suppresses edge information, while strong backlighting during sunrise or sunset can cause pedestrians to appear as dark silhouettes. These scenarios often lead to missed detections (false negatives) or reduced detection confidence in conventional pedestrian detection models. Numerous studies have tried to address these limitations by proposing methods that enhance robustness and accuracy in such complex scenarios [6]–[9].

To increase the model's accuracy in low-light conditions, Nataprawira *et al.* [10] proposed a multi-spectral pedestrian detection method based on the you only look one (YOLO)v3 architecture, leveraging inputs from red, green, blue (RGB), thermal, and combined multi-spectral imaging modalities. In this approach, the three channels of the RGB image were integrated with the single channel from the thermal image to form a unified four-channel input representation. Roszyk *et al.* [11] employed the YOLOv4 architecture to develop a rapid, low-latency multi-spectral pedestrian detection system tailored for autonomous driving applications. The study explored various model architectures and fusion strategies, with feature-level fusion—referred to as YOLOv4-middle—emerging as the most effective in balancing detection accuracy and computational efficiency. Following a similar objective but adopting a transfer learning approach, Hu *et al.* [12] employed CycleGAN [13] to produce synthetic infrared images from visible-light inputs, thereby augmenting the computer vision center (CVC)-09 dataset. This augmented dataset was then used to train and evaluate YOLOv3 [14] and faster region-based convolutional neural network (R-CNN) [15] models, aiming to enhance pedestrian detection accuracy in low-light environments.

To address the challenges caused by adverse weather conditions, Wang *et al.* [16] proposed a hybrid approach combining physical modeling with deep learning techniques. The resulting model, called physics-aware and conditional (PC)-YOLO, uses two novel components: the enhanced physics-aware block (EPB) and the conditional width diversification branch block (C2f_WDBB). The effectiveness of this architecture was validated through evaluations on the RainyCityscape and HazyCityscape datasets. Liu and Lin [17] incorporated the dark channel dehazing algorithm into the YOLOv7 framework [18], and demonstrated a notable enhancement in dehazing efficiency. This integration significantly improved the performance of YOLOv7 in foggy conditions by enhancing the model's ability to detect objects in low-visibility environments. Hnewa *et al.* [19] introduced a novel framework employing cross-modality knowledge distillation (CMKD) to enhance the effectiveness of RGB-based pedestrian detection under low-light and adverse weather conditions. The proposed approach was evaluated on the "seeing through fog" dataset, thus, confirming its improved performance in challenging visual environments.

The majority of research addresses pedestrian detection under either low-light conditions or adverse weather, while few studies concurrently consider both challenges. Recent improvements in computer vision have shown that adding attention mechanisms can enhance feature representation by enabling models to focus on the most informative regions within an image. Despite these advancements, there remains limited research specifically examining the efficacy of attention modules for pedestrian detection in environments characterized by low visibility and poor weather conditions. Most studies employing attention-based approaches have primarily concentrated on improving detection performance in densely crowded scenes: Jubair and Nafea [20] proposed convolutional block attention modules (CBAM)-YOLOv8, incorporating three CBAM into YOLOv8'backbone. Experimental evaluation conducted on a composite dataset comprising 1,520 images from INRIA and ETH datasets demonstrated that the integration of CBAM leads to important improvements in both recall and the mean average precision (mAP) across the intersection over union (IoU) threshold range of 0.5 to 0.95 (mAP_0.5:0.95). Wang *et al.* [21] introduced a CCW-YOLO algorithm, which incorporated a lightweight convolutional layer utilizing GhostConv, alongside an improved C2f module aimed at boosting the detection capabilities of the network. Furthermore, the model integrated the coordinate attention mechanism to more effectively capture critical target features, thereby increasing the model's adaptability across diverse detection scenarios and further enhancing overall accuracy. Zang *et al.* [22] introduced multi-receptive field and attention mechanism for multispectral pedestrian detection (MAPD) which leverages both multi-receptive field processing and attention mechanisms. This design effectively integrates a multi-receptive field module with the CBAM to form a combined multi-receptive field and attention (MA) module. Notably, this module was built to be modular, facilitating easy integration with diverse pre-existing network architectures.

In response to limitations in prior pedestrian detection research under challenging visual conditions, this study proposes an attention-augmented YOLOv5s architecture trained on an enhanced common objects in context (COCO) dataset [23], which is augmented with synthetic transformations simulating darkness, noise, and blur. The primary objective is to evaluate whether the incorporation of a novel attention module, specifically, a weighted fusion efficient channel attention (WF-ECA) mechanism that adaptively combines max and average pooling, can improve detection robustness and accuracy for pedestrian detection under visually degraded conditions, relative to the baseline YOLOv5s model. The main contributions of this work are: i) the justification of YOLOv5s as a baseline through comparative evaluation with state-of-the-art

detectors; ii) the design of WF-ECA to dynamically balance pooling features for improved robustness; and iii) a comprehensive assessment of WF-ECA against existing attention modules, demonstrating its effectiveness in enhancing pedestrian detection under degraded visual conditions.

The structure of the paper is as: section 2 details the methodology, and provides an overview of the YOLOv5s architecture and a description of the proposed attention module. Section 3 presents the experimental results and discussions. Section 4 concludes the study.

2. METHOD

To improve pedestrian detection performance under adverse visual conditions, the YOLOv5s architecture was trained on a COCO dataset enhanced with synthetic transformations simulating darkness, noise, and blur, specifically targeting the person class. The selection of YOLOv5s is motivated by its effective compromise between detection accuracy and computational efficiency. To justify this choice, transfer learning experiments were conducted using YOLOv5s alongside other state-of-the-art detectors, including YOLOv8s [24], single shot detector (SSD) [25], and EfficientDet [26], evaluated on the augmented dataset. Starting from this baseline, we propose a novel, lightweight attention module inspired by the efficient channel attention (ECA) mechanism. The module's performance is evaluated through comparative analysis against the YOLOv5s baseline and several established attention modules to determine its impact on detection accuracy in challenging environments.

2.1. Dataset

A specialized dataset for pedestrian detection was created using the COCO dataset. From this dataset, around 40,000 images were chosen for training and 2,758 for validation, focusing on images containing the person category. To strengthen the detector's performance and handle common issues encountered in pedestrian detection, a wide range of data augmentation strategies were applied. Roboflow was used both to construct and augment the dataset. The augmentations performed on the original images included:

- Brightness adjustment: brightness was reduced by 20% to improve the model's ability to operate under low-light conditions, such as nighttime scenarios.
- Blurring: a Gaussian blur with a 2.5 pixel was added to improve the model's adaptability to camera focus issues: an important consideration in ADAS applications where both the camera and subjects (e.g., pedestrians or cyclists) may be in motion.
- Noise addition: random noise was introduced to up to 2% of the pixels, simulating sensor noise. This encourages the model to generalize better by learning to identify pedestrians even when parts of them are obscured.

Following these augmentations, the final dataset comprised 118,425 training images, 2,758 validation images, and 186 images reserved for testing. The complete dataset is hosted on Roboflow.

2.2. YOLOv5s baseline architecture

YOLOv5s [27] is a popular choice for object detection tasks because it strikes a good balance between accuracy and computational speed. Its architecture consists of three main parts: the backbone, the neck, and the head. The backbone uses C3 and convolutional modules, applying the cross-stage partial (CSP) technique to improve gradient flow and enable effective reuse of features while minimizing redundant calculations. This setup ensures efficient feature extraction, which is crucial for precise detection. The neck includes a spatial pyramid pooling fast (SPPF) layer along with additional convolution layers, which helps combine features at different scales. This not only increases the receptive field but also boosts the model's ability to understand context across various resolutions. Finally, the head component is responsible for predicting bounding box coordinates, objectness scores, and class probabilities, thus producing the final detection outputs. An overview of the YOLOv5s architecture is illustrated in Figure 1. Although alternative models such as YOLOv8s, SSD, and EfficientDet demonstrate improvements in certain performance metrics, our comparative analysis (section 3) confirms that YOLOv5s, when retrained on our augmented dataset, remains the most suitable choice for real-time pedestrian detection.

2.3. Overview of existing attention modules

Attention mechanisms have been integrated into neural network architectures to improve feature representation by directing focus toward the most informative elements, thereby improving overall performance in various vision tasks. A range of attention modules has been proposed in the literature, each employing different strategies to refine feature maps. The squeeze-and-excitation (SE) module [28] introduces channel-wise attention by first aggregating global spatial information through global average pooling to produce a channel descriptor, which is then passed through a small fully connected network to

generate channel-wise weights for feature recalibration. The CBAM [29] extends this approach by sequentially applying both channel and spatial attention. It integrates both global average pooling and max pooling to produce channel attention, followed by convolutional operations on the aggregated features to produce a spatial attention map. The ECA module [30] offers a more lightweight alternative by avoiding dimensionality reduction and instead capturing cross-channel interactions using a fast 1D convolution with a carefully selected kernel size. Despite its simplicity, ECA achieves competitive performance compared to SE and CBAM, while offering enhanced computational efficiency, which makes it well suited for real-time and resource-limited applications.

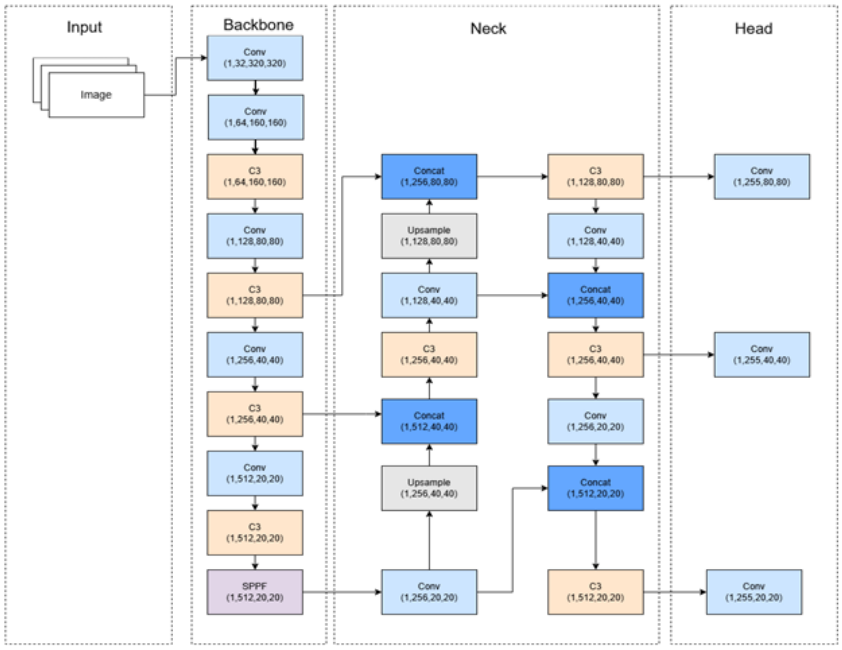


Figure 1. YOLOv5s model architecture

2.4. Improved YOLOV5s model with weighted fusion efficient channel attention

A primary objective of this study is to improve the performance of the YOLOv5s detector under challenging conditions through the integration of an effective yet lightweight attention mechanism. Building on the computational efficiency and effectiveness of the ECA module, we propose a novel attention mechanism termed WF-ECA. While the original ECA module utilizes only global average pooling to encode channel-wise information, this approach may overlook complementary features that can be captured through max pooling. It is well established that average pooling and max pooling emphasize different characteristics of feature maps: average pooling reflects the general feature distribution, whereas max pooling identifies the most dominant responses. Relying exclusively on one pooling strategy can constrain the module's capacity to adaptively focus on the most informative channel features. To address this limitation, the proposed WF-ECA module adaptively integrates both average and max pooling operations, enabling the network to learn an optimal balance between them. This fusion enriches channel-wise context modeling without compromising the lightweight nature of the ECA structure. The WF-ECA module is incorporated at the end of the backbone stage of the YOLOv5s architecture, as illustrated in Figure 2, to enhance feature extraction efficiency in a computationally economical manner.

Let $X \in R^{C.W.H}$ represent the input feature map, where the dimensions C, H, and W refer to the number of channels, height, and width, respectively. To enhance channel-wise feature representation while maintaining computational efficiency, the proposed attention mechanism first performs both global average pooling and global max pooling across the spatial dimensions. These operations yield two 1D descriptors, f_{avg} and $f_{max} \in R^C$ where each element is computed as (1) and (2).

$$f_{avg}(C) = \frac{1}{H.W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(c,i,j) \quad (1)$$

$$f_{max}(C) = \max_{i,j} X(c,i,j) \quad (2)$$

The two vectors are concatenated to form a unified descriptor $f_{concat} \in R^{2C}$.

To adaptively weigh the relative importance of average and max pooled features, a learnable parameter $\alpha \in [0,1]^C$ is introduced. This weighting vector is generated by applying a parameterized function σ typically implemented as a sigmoid to the concatenated descriptor as shown in (3).

$$\alpha = \sigma(W_\alpha . f_{concat}) \tag{3}$$

Where $W_\alpha \in R^{C.2C}$ is a learnable parameter. The adaptive attention vector is then computed as a convex combination of the two pooled vectors in (4).

$$f_{attn} = \alpha \odot f_{avg} + (1 - \alpha) \odot f_{max} \tag{4}$$

To model local channel dependencies without introducing significant computational overhead, a lightweight one-dimensional convolution is subsequently applied to the adaptive vector f_{attn}. This operation captures interactions among neighboring channels efficiently, resulting in an enhanced attention vector $f_{conv} \in R^C$. Finally, this vector is employed to recalibrate the original feature map via channel-wise multiplication, formulated as (5).

$$X'(c,i,j) = f_{conv} X(c,i,j), \qquad \forall c \in \{1,\dots,C\}, i \in \{1,\dots,H\}, j \in \{1,\dots,W\} \tag{5}$$

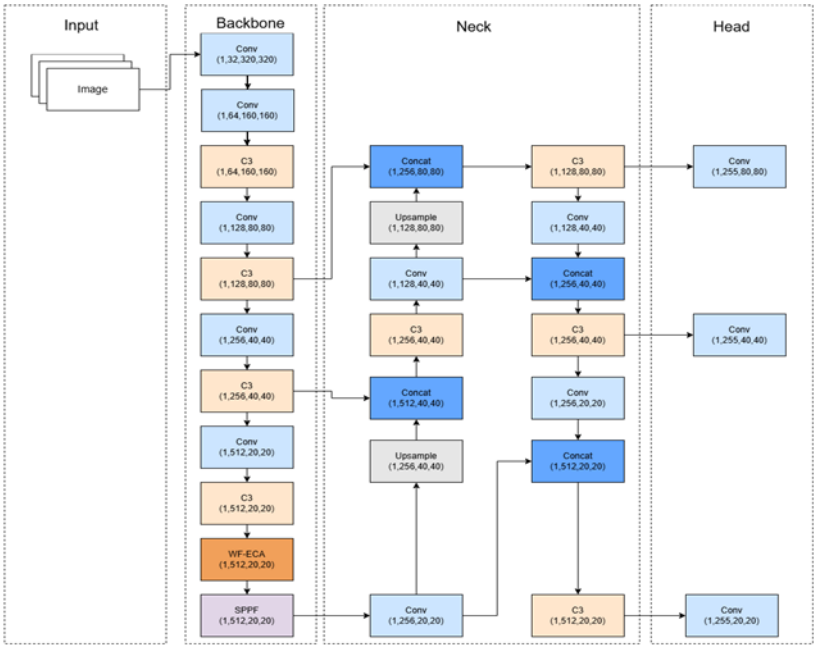


Figure 2. The improved YOLOv5s with WF-ECA architecture

3. RESULTS AND DISCUSSION

Initially, the YOLOv5s model was fine-tuned using our augmented dataset. The model training was performed using the Google Colab platform. Subsequently, we carried out a comparative analysis with several well-established object detection models. Based on the observed performance under challenging conditions, a novel attention module was incorporated to further enhance the model's detection capabilities.

3.1. Experimental setup

The selected models: YOLOv5s, YOLOv8s, SSD, and EfficientDet, were retrained using the Google Colab platform. Prior to training, the augmented dataset was converted into the appropriate format for each model using the Roboflow platform. The retraining process was performed for 300 epochs using a batch size of 16 and an input image resolution of 640×640 pixels.

The retrained models were assessed using the validation subset of the dataset. For comparative analysis, we employed two widely recognized evaluation metrics: mAP at IoU threshold 0.5 (mAP_0.5) and mAP across multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05 (mAP_0.5:0.95). The mAP_0.5 measures the average precision when the IoU is fixed at 50%, while mAP_0.5:0.95 offers a more thorough evaluation by averaging precision over a spectrum of stricter IoU thresholds.

The comparative results are illustrated in Figure 3. As depicted, both YOLOv5s and YOLOv8s significantly outperform SSD and EfficientDet in terms of detection accuracy and convergence speed. Notably, YOLOv8s achieves the highest accuracy among all tested models; however, its increased computational complexity and inference latency make it less suitable for real-time deployment on resource-constrained devices. In contrast, YOLOv5s offers a favorable compromise between accuracy and efficiency, making it particularly well-suited for real-time pedestrian detection tasks. The final mAP values are presented in Table 1.

Furthermore, Table 2 presents the inference performance of the models when deployed on a Raspberry Pi 4. The results confirm that YOLOv5s offers a balanced compromise between detection accuracy and real-time processing capabilities. This reinforcing its practicality for embedded systems and edge computing scenarios.

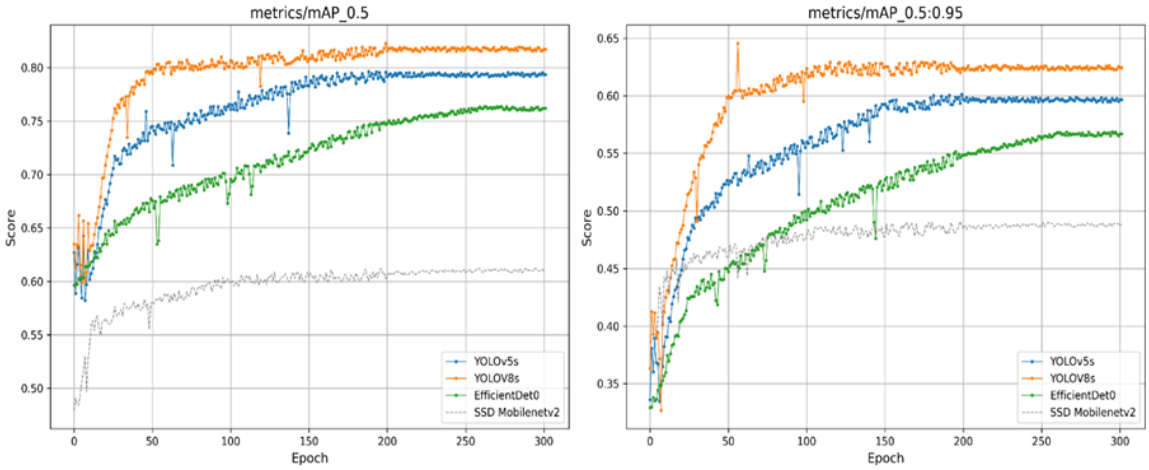


Figure 3. Comparison between modern detectors on the augmented dataset

Table 1. mAP values of the chosen models

Model	mAP 0.5	mAP 0.5:0.95
YOLOv5s	0.79	0.59
YOLOv8s	0.82	0.62
EfficientDet0	0.76	0.57
SSD	0.61	0.49

Table 2. Frames per second (FPS) of models on Raspberry Pi 4

Models	YOLOv5s	YOLOv8s	EfficientDet0	SSD
FPS	5	3.5	2	1.8

3.2. Performance of the proposed attention module

To analyze the performance of the proposed improved YOLOv5s model incorporating the WF-ECA attention mechanism, this study conducted retraining experiments using the augmented dataset. For comparison, additional retraining was performed on YOLOv5s integrated with alternative attention modules, namely SE, ECA, and CBAM. The comparative performance outcomes are summarized in Figure 4. The findings confirm that the improved YOLOv5s model with the WF-ECA module achieves a notable enhancement over the baseline YOLOv5s, with an approximately 5% increase in mAP. Furthermore, the WF-ECA variant demonstrated more rapid convergence during training and consistently higher mAP values relative to models employing the other attention modules. These findings underscore the efficacy of the WF-ECA attention mechanism in enhancing the detection accuracy and training efficiency of YOLOv5s within the context of the augmented dataset. The final mAP values are presented in Table 3.

To ensure the robustness of these results, we performed multiple independent training runs for both the baseline YOLOv5s and the WF-ECA variant (3 runs each). Table 4 reports the mean ± standard deviation of the mAP values across these runs. Statistical significance was assessed using paired t-tests, with corresponding p-values included to demonstrate that the performance improvements of WF-ECA over the baseline are consistent.

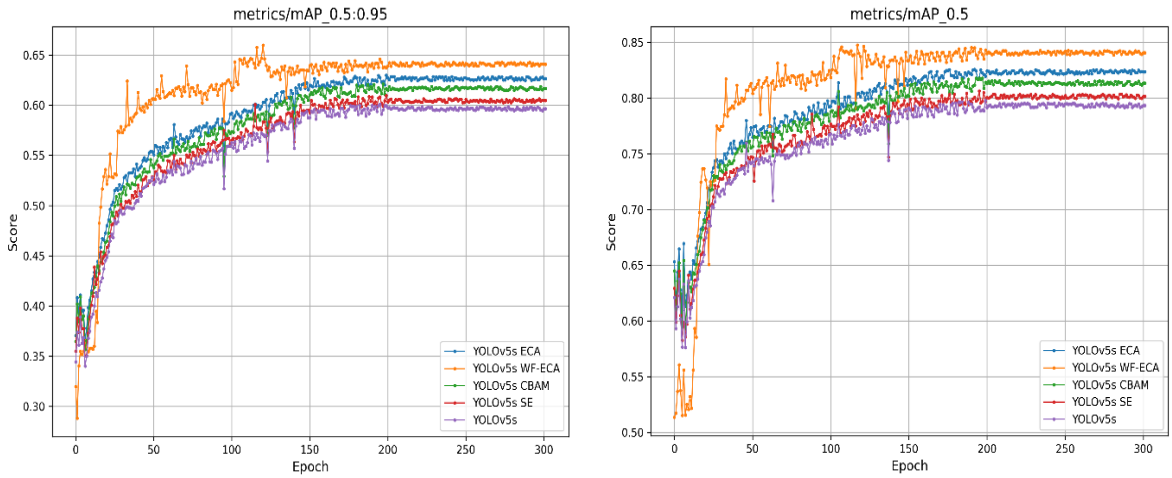


Figure 4. Comparison between different attention modules incorporated with YOLOv5s on the augmented dataset

Table 3. mAP values of the YOLOv5s with different attention modules

Model	mAP 0.5	mAP 0.5:0.95
YOLOv5s	0.79	0.59
YOLOv5s SE	0.80	0.60
YOLOv5s CBAM	0.81	0.62
YOLOv5s ECA	0.82	0.63
Ours	0.84	0.64

Table 4. mAP performance of YOLOv5s and YOLOv5s with WF-ECA (mean ±std) with statistical significance

Model	mAP 0.5 (mean ±std)	mAP 0.5:0.95 (mean ±std)	p-value vs YOLOv5s
YOLOv5s	0.79±0.02	0.59±0.02	-
YOLOv5s WF-ECA	0.82±0.03	0.63±0.03	<0.01

To further assess the robustness and generalization of the proposed WF-ECA module, we evaluated both YOLOv5s and YOLOv5s WF-ECA on FoggyCityscapes and ExDark, which present challenging conditions such as foggy urban scenes and extremely low-light environments. The results, summarized in Table 5, show that while detection performance naturally decreases under these difficult scenarios, the WF-ECA variant consistently outperforms the baseline, demonstrating its ability to enhance feature extraction and detection accuracy in adverse conditions. Another important aspect to consider is the influence of the proposed attention module on the baseline model's computational efficiency, particularly regarding inference speed. Given that our model is based on the ECA mechanism, it remains lightweight. As demonstrated in Table 6, the giga floating point operations per second (GFLOPs) metric exhibits negligible change, indicating that the addition of the attention module does not meaningfully raise the computational complexity of the baseline model.

Table 5. Performance of YOLOv5s and YOLOv5s with WF-ECA on FoggyCityscapes and ExDark datasets

Model	Foggy Cityscapes		ExDark	
	mAP 0.5	mAP 0.5:0.95	mAP 0.5	mAP 0.5:0.95
YOLOv5s	0.70	0.47	0.73	0.5
YOLOv5s WF-ECA	0.77	0.54	0.78	0.54

Table 6. Comparison between YOLOv5s and the improved one on the augmented dataset

Model	Number of layers	Number of parameters	GFLOPs
YOLOv5s	157	7,012,822	15.8
Ours	160	7,537,113	15.8

To further evaluate the performance of the improved YOLOv5s model, a qualitative visual comparison was conducted against the baseline model using a set of novel test images representing challenging

environmental conditions, including night-time, fog, and rain scenarios. As depicted in Figure 5, the improved YOLOv5s consistently exhibits superior detection performance under adverse conditions. Notably, while the standard YOLOv5s experiences a decline in performance in such scenarios, manifested primarily through an increase in false negative detections, the enhanced version effectively addresses this issue, as shown in Figure 5(a). The improved robustness of the proposed model in challenging visual environments underscores its practical applicability and efficacy in real-world settings, as shown in Figure 5(b).

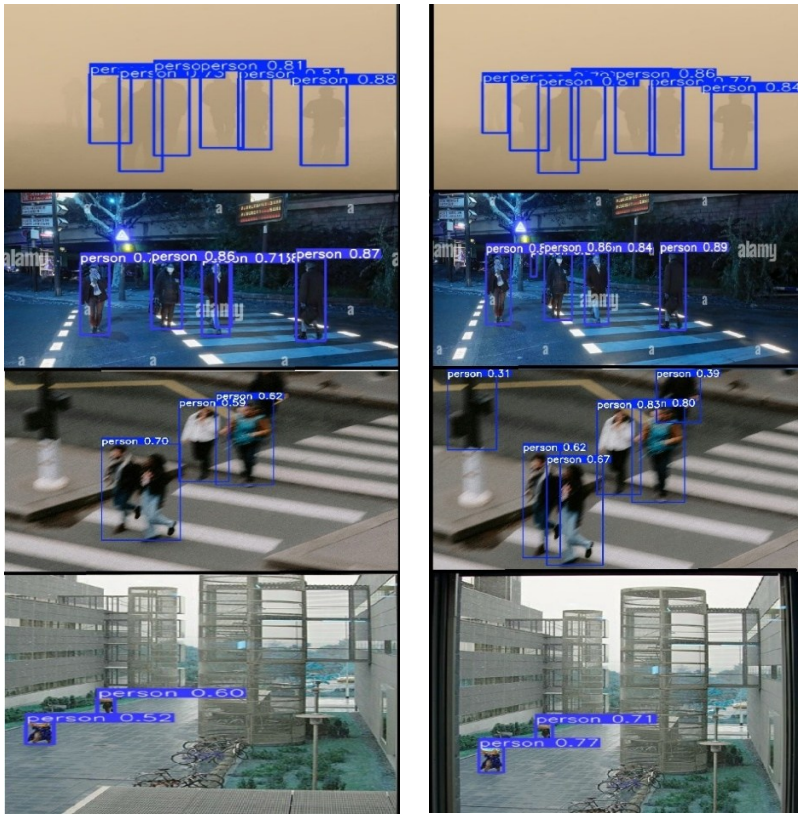


Figure 5. Side-by-side comparison of YOLOv5s and the improved detector with (a) YOLOv5s results and (b) improved model outputs

4. CONCLUSION

Maintaining safe driving conditions in scenarios involving poor illumination and adverse weather is a critical concern in ADAS systems. Pedestrian detection becomes particularly challenging under these circumstances, with even robust models such as YOLOv5 exhibiting reduced performance. This paper explored the application of an attention mechanism to address these limitations. An enhanced version of the YOLOv5s model was proposed, incorporating a novel WF-ECA module. The model was retrained on an augmented dataset specifically designed to simulate challenging conditions, including low-light and blur effects. YOLOv5s was chosen as the baseline as it offers an efficient compromise between accuracy and computational efficiency, as demonstrated through comparative evaluation with other detection frameworks. The novel model achieved a 5% increase in mAP, while maintaining the original inference speed, thus confirming the effectiveness of the proposed attention-based enhancement in difficult environmental conditions. The inclusion of the WF-ECA module enabled the model to more effectively capture critical features in degraded visual inputs. Extensive experiments confirmed that the proposed model consistently outperformed the baseline across multiple adverse scenarios. These results emphasize the potential of attention mechanisms in enhancing object detection reliability under real-world operational challenges.

ACKNOWLEDGMENTS

The authors wish to thank the University of Moulay Ismail for supporting this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Oumayma Rachidi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Badr Bououlid Idrissi		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
Chafik Ed-Dahmani	✓		✓	✓			✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [OR], upon reasonable request.

REFERENCES

[1] M. S. Gulino, G. Vichi, F. Cecchetto, L. Di Lillo, and D. Vangi, "A combined comfort and safety-based approach to assess the performance of advanced driver assistance functions," *European Transport Research Review*, vol. 17, no. 1, 2025, doi: 10.1186/s12544-024-00696-4.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[3] B. Sebastian and P. B.-Tzvi, "Support vector machine based real-time terrain estimation for tracked robots," *Mechatronics*, vol. 62, 2019, doi: 10.1016/j.mechatronics.2019.102260.

[4] K. I. Na and B. Park, "Real-time 3D multi-pedestrian detection and tracking using 3D LiDAR point cloud for mobile robot," *ETRI Journal*, vol. 45, no. 5, pp. 836–846, 2023, doi: 10.4218/etrij.2023-0116.

[5] Rendi and D. Fitrianah, "Enhancing low-light pedestrian detection: convolutional neural network and YOLOv8 integration with automated dataset," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 3, pp. 1969–1980, 2025, doi: 10.11591/eei.v14i3.8903.

[6] X. Xiong, J. Yang, Y. Jiang, and X. Hu, "A dual-stream convolutional network for visible and infrared image fusion in pedestrian detection," in *2024 International Conference on Artificial Intelligence and Autonomous Transportation*, 2025, pp. 27–34, doi: 10.1007/978-981-96-3973-1_4.

[7] F. Cao *et al.*, "SSCD-YOLO: semi-supervised cross-domain YOLOv8 for pedestrian detection in low-light conditions," *IEEE Access*, vol. 13, pp. 61225–61236, 2025, doi: 10.1109/ACCESS.2025.3557387.

[8] M. Hu *et al.*, "An enhanced feature-fusion network for small-scale pedestrian detection on edge devices," *Sensors*, vol. 24, no. 22, 2024, doi: 10.3390/s24227308.

[9] M. Jung and J. Cho, "Enhancing detection of pedestrians in low-light conditions by accentuating Gaussian–Sobel edge features from depth maps," *Applied Sciences*, vol. 14, no. 18, 2024, doi: 10.3390/app14188326.

[10] J. Nataprawira, Y. Gu, I. Goncharenko, and S. Kamijo, "Pedestrian detection on multispectral images in different lighting conditions," in *IEEE International Conference on Consumer Electronics*, 2021, pp. 1–5, doi: 10.1109/ICCE50685.2021.9427627.

[11] K. Roszyk, M. R. Nowicki, and P. Skrzypczyński, "Adopting the YOLOv4 architecture for low-latency multispectral pedestrian detection in autonomous driving," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031082.

[12] J. Hu, Y. Zhao, and X. Zhang, "Application of transfer learning in infrared pedestrian detection," in *2020 IEEE 5th International Conference on Image, Vision and Computing, ICIVC 2020*, 2020, pp. 1–4, doi: 10.1109/ICIVC50857.2020.9177438.

[13] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV),* 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[14] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, *arXiv:1804.02767*.

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

[16] Y. Wang, H. Ke, and H. Cai, "PC-YOLO: enhancing object detection in adverse weather through physics-aware and dynamic network structures," *Journal of Electronic Imaging*, vol. 34, no. 02, 2025, doi: 10.1117/1.jei.34.2.023049.

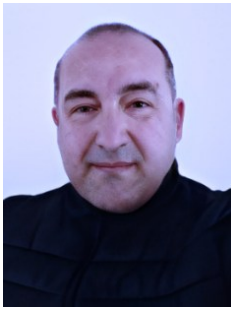
[17] X. Liu and Y. Lin, "YOLO-GW: quickly and accurately detecting pedestrians in a foggy traffic environment," *Sensors*, vol. 23, no. 12, 2023, doi: 10.3390/s23125539.

[18] W. E. Manongga, R. C. Chen, and X. Jiang, "Enhancing road marking sign detection in low-light conditions with YOLOv7 and contrast enhancement techniques," *International Journal of Applied Science and Engineering*, vol. 21, no. 3, 2024, doi: 10.6703/IJASE.202406_21(3).001.

[19] M. Hnewa, A. Rahimpour, J. Miller, D. Upadhyay, and H. Radha, "Cross modality knowledge distillation for robust pedestrian detection in low light and adverse weather conditions," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095353.
[20] M. A. Jubair and M. M. Nafea, "Attention-enhanced YOLOv8 for accurate pedestrian detection and count estimation," *Journal of Soft Computing and Data Mining*, vol. 5, no. 2, pp. 175–187, 2024, doi: 10.30880/jscdm.2024.05.02.013.
[21] Z. Wang, S. Yang, H. Qin, Y. Liu, and J. Ding, "CCW-YOLO: a modified YOLOv5s network for pedestrian detection in complex traffic scenes," *Information*, vol. 15, no. 12, 2024, doi: 10.3390/info15120762.
[22] Y. Zang, R. Cao, H. Li, W. Hu, and Q. Liu, "MAPD: multi-receptive field and attention mechanism for multispectral pedestrian detection," *Visual Computer*, vol. 40, no. 4, pp. 2819–2831, 2024, doi: 10.1007/s00371-023-02988-7.
[23] T.-Y. Lin *et al.*, "Microsoft COCO: common objects in context," in *Computer Vision – ECCV 2014*, Cham, Switzerland: Springer International Publishing, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.
[24] J. Sang, Z. Zhang, C. Xiao, H. Luo, and J. Zhang, "An improved YOLOv8s method and its application in road traffic target detection," *Hongwai yu Jiguang Gongcheng/Infrared and Laser Engineering*, vol. 53, no. 11, 2024, doi: 10.3788/IRLA20240256.
[25] W. Liu *et al.*, "SSD: single shot multibox detector," in *Computer Vision – ECCV* 2016, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[26] A. R. G.-Escalante, R. Q. F.-Aguilar, A. P.-Zubia, and E. M.-Vargas, "Automatic brake driver assistance system based on deep learning and fuzzy logic," *PLoS ONE*, vol. 19, no. 12, 2024, doi: 10.1371/journal.pone.0308858.
[27] W. Ci, L. Tian, Y. Ge, H. Hou, and J. Ma, "An absolute distance estimation method for obstacle detection based on YOLOv5s," in *ACM International Conference Proceeding Series*, 2024, pp. 6–14, doi: 10.1145/3690931.3690933.
[28] R. T. N. V. S. Chappa and M. El-Sharkawy, "Squeeze-and-excitation squeezenext: an efficient DNN for hardware deployment," in *2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020*, 2020, pp. 691–697, doi: 10.1109/CCWC47524.2020.9031119.
[29] D. Hu, Z. Zhang, and G. Niu, "Lane line detection incorporating CBAM mechanism and deformable convolutional network," *Beijing Hangkong Hangtian Daxue Xuebao/Journal of Beijing University of Aeronautics and Astronautics*, vol. 50, no. 7, pp. 2150–2160, 2024, doi: 10.13700/j.bh.1001-5965.2022.0601.
[30] B. Liu, H. Wang, Y. Wang, C. Zhou, and L. Cai, "Lane line type recognition based on improved YOLOv5," *Applied Sciences*, vol. 13, no. 18, 2023, doi: 10.3390/app131810537.

BIOGRAPHIES OF AUTHORS

Oumayma Rachidi is a third-year Ph.D. student and received an engineer's degree from the National Graduate School of Arts and Crafts, Meknes, Morocco, in 2021. She is actually working as an electromechanical engineer in the industry field. Her ongoing dissertation studies pedestrian detection in ADAS systems and explores deep learning techniques for object detection. She is particularly interested in industrial control systems, FPGA-based ADAS systems, and electrical vehicles. She can be contacted at email: oum.rachidi@edu.umi.ac.ma.



Badr Bououlid Idrissi received the Ph.D. degree from Faculté Polytechnique de Mons, Mons, Belgium, in 1997 and the engineer's degree from Ecole Nationale de l'Industrie Minérale, Rabat, Morocco, in 1992. Since 1999, he has been working at Ecole Nationale Supérieure d'Arts et Métiers (ENSAM-Meknès), Moulay Ismaïl University, Meknès, Morocco, where he is a professor in the Department of Electromechanical Engineering, in the areas of power electronics and electrical machines. His research interests are mainly electric drives, industrial control systems, DSP/FPGA-based ADAS systems and electrical vehicles. He can be contacted at email: b.bououlid@umi.ac.ma.

Chafik Ed-Dahmani received his engineering degree in Mechatronics from the University of Abdelmalek Essaadi in 2014 and his Ph.D. degree from the University of Mohamed 5-Rabat, Morocco. Currently, he is an assistant professor at the National Graduate School of Arts and Crafts-Meknès, Morocco. His research interests include renewable energy system conversion, control systems, and microgrids. He can be contacted at email: c.eddahmani@umi.ac.ma.