IAES International Journal of Robotics and Automation (IJRA)
Vol. 15, No. 3, September 2026, pp. 544∼552
ISSN: 2722-2586, DOI: 10.11591/ijra.v15i3.pp544-552

Spatial-channel reconstruction for efficient multiscale attention in robotic object detection

Mohammed Maiza[1], Chahira Cherif[2], Samira Chouraqui[3], Abdelmalik Taleb-Ahmed[4]
[1]LITIO Laboratory, Faculty of Exact and Applied Sciences, University of Oran 1 Ahmed Ben Bella, Oran, Algeria
[2]RIIR Laboratory, Faculty of Medicine, University of Oran 1 Ahmed Ben Bella, Oran, Algeria
[3]Faculty of Mathematics and Computer Science, University of Sciences and Technology of Oran, Oran, Algeria
[4]IEMN, Polytechnic University of Hauts-de-France, University of Lille, Valenciennes, France

Article Info

Article history:

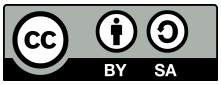
Received Feb 17, 2026
Revised Apr 24, 2026
Accepted May 22, 2026

Keywords:

Autonomous robots
Lightweight attention
Object detection
Real-time perception
YOLO

ABSTRACT

Real-time object detection is a core capability for autonomous robots, unmanned aerial vehicles (UAVs), and self-driving systems operating in resource-constrained environments. This paper presents spatial-channel enhanced multiscale attention (SCEMA), a novel lightweight attention module designed to enhance robotic perception while minimizing computational overhead for embedded deployment. SCEMA employs a parallel dual-branch architecture that synergistically combines spatial-channel reconstruction with multiscale attention mechanisms. When integrated into the YOLOv8n framework (3.01M baseline parameters), the proposed YOLO-SCEMA model achieves significant performance gains across multiple challenging benchmarks relevant to robotics automation. Experiments on an NVIDIA RTX 4080 GPU demonstrate that on the ExDark dataset, YOLO-SCEMA improves mAP@50 by 7.37% over the baseline (69.07% to 76.44%) while reducing parameters by 36.88% (3.01M to 1.90M) and computational cost by 8.64% (8.1 to 7.4 GFLOPs). Consistent improvements are also observed on VisDrone2019 (+3.24% mAP@50) and FYP (+1.50% mAP@50) datasets. Comparative analysis demonstrates that YOLO-SCEMA achieves superior accuracy-efficiency trade-offs, making it particularly suitable for deployment in low-light conditions, dense scenes, and complex structural environments for autonomous navigation, robotic surveillance, and industrial automation applications.



Corresponding Author:

Mohammed Maiza
LITIO Laboratory, Faculty of Exact and Applied Sciences, University of Oran 1 Ahmed Ben Bella
Oran, Algeria
Email: maiza.mohammed@univ-oran1.dz

1. INTRODUCTION

Object detection constitutes a fundamental perception capability for autonomous robotic systems, including ground robots, unmanned aerial vehicles (UAVs), and self-driving vehicles. These systems must reliably detect and localize objects in real-time while operating under severe computational and energy constraints inherent to embedded robotic platforms [1], [2]. Convolutional neural networks (CNNs) have become the dominant paradigm for visual perception in robotics; however, modern detectors achieve high accuracy at the cost of increased depth and width, significantly raising computational and memory requirements that limit deployment on resource-constrained robotic hardware [3]. In robotic applications such as autonomous navigation, surveillance, and industrial automation, perception systems face three critical challenges: i) low-light conditions (e.g.,

nighttime autonomous driving), ii) dense object distributions with occlusions (e.g., UAV-based crowd monitoring), and iii) complex structural scenes (e.g., manufacturing environments). As network depth increases, feature maps exhibit substantial spatial and channel redundancy [4], degrading computational efficiency. While attention mechanisms have been introduced to emphasize informative features, many existing approaches incur non-negligible computational overhead or focus on either spatial or channel dependencies in isolation. Consequently, designing lightweight attention modules that jointly exploit spatial and channel interactions while preserving multiscale contextual information remains a critical challenge for real-time robotic perception [5], [6]. Recent YOLO-based detectors emphasize efficiency and real-time performance, yet their lightweight variants often suffer from degraded accuracy in challenging robotic environments. To address this gap, this paper proposes SCEMA, a novel lightweight attention module that enhances robotic perception without introducing excessive computational cost.

This work advances robotics and automation by enabling lightweight, efficient, and robust perception for autonomous systems. The paper is organized as follows. Section 2 reviews related work on object detection architectures and attention mechanisms in robotics contexts. Section 3 presents the SCEMA methodology, including integration into robotic vision pipelines. Section 4 details experimental results across three benchmark datasets, with analysis framed in terms of robotic metrics. Finally, section 5 concludes with implications for autonomous robotics and future directions.

2. RELATED WORK

Attention mechanisms enhance feature discrimination by selectively emphasizing informative aspects of input data, which is particularly valuable for robotic perception in cluttered or low-visibility environments. Channel attention methods, pioneered by SENet [7], employ global average pooling followed by fully connected layers to recalibrate feature responses. Spatial attention approaches compute importance maps across spatial dimensions, highlighting regions with high semantic relevance. Hybrid attention modules, such as CBAM [8] and ECA [9], combine spatial and channel cues sequentially or jointly. However, many attention modules introduce computational overhead through parameter-intensive operations, hindering deployment on robotic hardware with limited processing capabilities. Furthermore, existing mechanisms frequently fail to efficiently model multiscale spatial dependencies, limiting their effectiveness for robotic applications involving objects at varying distances and aspect ratios. The demand for real-time object detection in robotics has driven research into lightweight detectors that balance computational efficiency and detection accuracy. The YOLO family has emerged as a dominant paradigm for real-time robotic perception due to its single-stage design and efficient feature extraction backbone [10]. Recent variants, including YOLOv8 [11], YOLOv9 [12], YOLOv10 [13], YOLOv11 [14], and YOLOv12 [15], have progressively improved the speed-accuracy trade-off. However, aggressive model compression can significantly degrade performance in challenging scenarios characteristic of robotic applications: low lighting, dense object distributions, occlusions, and significant scale variations. Developing lightweight detectors that maintain robust performance across diverse environments remains critical for autonomous navigation, UAV surveillance, and industrial automation. Multiscale feature representation is essential for robotic perception, as real-world scenes contain objects with substantial size variations. Feature pyramid-based architectures [16] construct hierarchical representations through top-down fusion or bidirectional information flow. However, these methods often rely on computationally expensive fusion strategies that increase inference latency and memory consumption, challenging real-time robotic operation. The integration of attention mechanisms with multiscale feature pyramids remains underexplored, limiting potential synergistic benefits. Efficient multiscale attention mechanisms that operate across different feature scales while maintaining real-time performance represent a significant open research problem for robotics [17], particularly for applications requiring long-range dependency modeling and contextual integration from different receptive fields.

3. METHOD

The overall architecture of SCEMA is illustrated in Figure 1. SCEMA adopts a parallel processing strategy consisting of two complementary branches. In the left branch, spatial-refined features X^w are extracted using the spatial refinement unit (SRU) and spatial attention, followed by channel refinement through the channel refinement unit (CRU) to produce channel-refined features Y.

The right branch focuses on learning effective channel representations by grouping features without

reducing channel dimensions, enhancing pixel-level attention and robustness to multi-scale variations. The architecture of the YOLO-SCEMA model is shown in Figure 2.

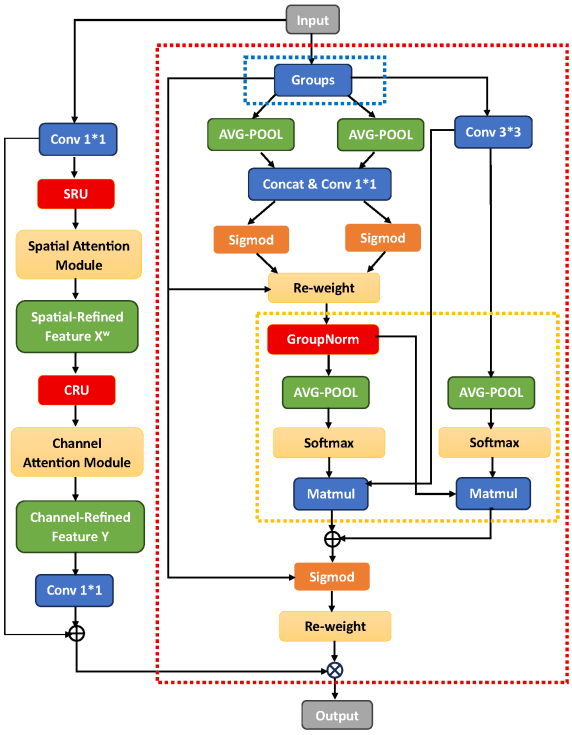


Figure 1. Architecture of the proposed SCEMA module

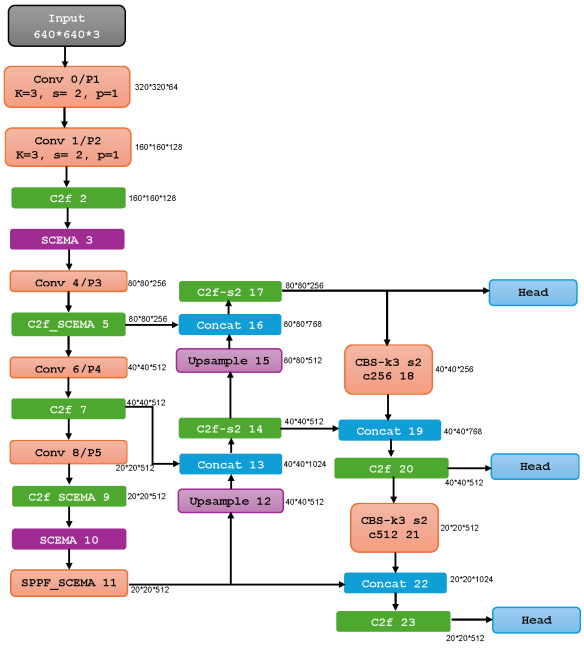


Figure 2. Overall architecture of the YOLO-SCEMA model

SCEMA is integrated into the YOLOv8 backbone, forming a perception module suitable for robotic autonomy stacks. Within the convolutional block (CBS), batch normalization is replaced by the feature re-

finement module and the cross-spatial learning module. In the bottleneck structure, the second convolutional layer is substituted with the Feature Refinement Module, forming Bottleneck-FR. The convolutional blocks in the spatial pyramid pooling fast (SPPF) module are replaced with CBS-FR and CBS-CSL. Through SCEMA integration, spatial and channel redundancies are effectively reduced, enabling enhanced feature representation for complex robotic scenes.

3.1. Feature refinement module

The feature refinement module consists of a spatial reconstruction unit (SRU), spatial attention module, channel reconstruction unit (CRU), and channel attention module. As illustrated in Figure 3, the SRU performs feature separation and reconstruction by dividing the input feature map into informative and less-informative components using a scaling factor derived from group normalization (GN) [18].

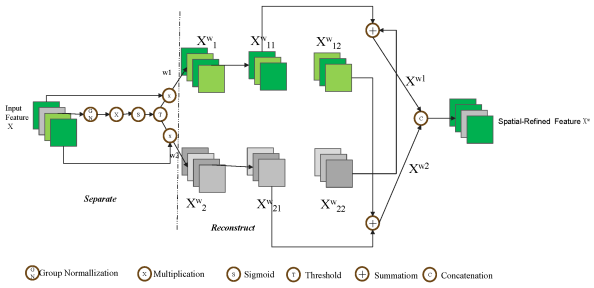


Figure 3. Architecture of the SRU

Given an intermediate feature map $X \in \mathbb{R}^{N\times C\times H\times W}$, GN is applied as,

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X-\mu}{\sqrt{\sigma^2+\varepsilon}} + \beta, \tag{1}$$

where μ and σ denote mean and standard deviation, ε ensures numerical stability, and γ, β are learnable parameters. Normalized features are mapped through a sigmoid-based gating mechanism to generate information-aware weights:

$$W = \text{Gate}(\text{Sigmoid}(\text{GN}(X))). \tag{2}$$

Weights above a predefined threshold form the informative mask W_1, while remaining weights constitute the non-informative mask W_2. The input feature map is decomposed as $X_1^W = W_1 \otimes X$ and $X_2^W = W_2 \otimes X$, representing informative and redundant features, respectively. Cross-reconstruction produces spatially refined feature map X^W. As shown in Figure 4, the CRU further refines X^W by reducing channel redundancy through splitting, transformation, and fusion operations with compression ratio $r = 2$ and group size $g = 2$, producing final channel-refined features Y.

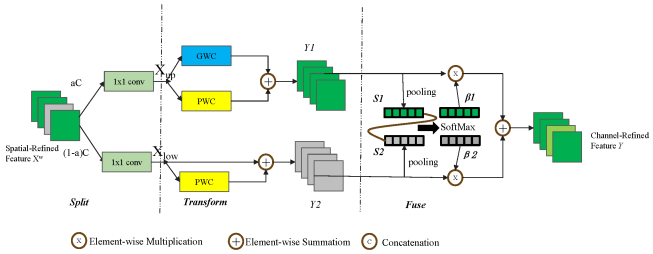


Figure 4. Architecture of the CRU

3.2. Cross spatial learning module

The CSL module employs parallel sub-structure processing. Initially, it utilizes 1×1 convolution as a common element. Subsequently, to integrate spatial information across multiple scales, it positions 3×3

convolution alongside the 1×1 branch. The CSL module segments input feature map $X \in \mathbb{R}^{C \times H \times W}$ into three distinct sub-feature groups, each focusing on distinct semantics. Two paths use 1×1 convolutions with one-dimensional global average pooling along horizontal and vertical directions, following Coordinate Attention. The third path adopts 3×3 convolution to capture multi-scale spatial features. Outputs are integrated through matrix dot-product operations:

$$A = \text{Sigmoid}((W_{1\times1} * X) \odot (W_{3\times3} * X)) \tag{3}$$

where $\odot$ denotes element-wise multiplication and $*$ denotes convolution. This cross-spatial interaction enables effective multi-scale feature fusion for robotic perception.

3.3. Integration with robotic perception pipelines

SCEMA is designed for seamless integration into robotic vision stacks, including ROS-based perception systems. The module accepts standard tensor inputs ($C \times H \times W$) and outputs refined feature maps with identical dimensions, enabling drop-in replacement for existing convolutional blocks. With only 1.90M parameters and 7.4 GFLOPs, SCEMA is suitable for deployment on edge devices such as NVIDIA Jetson, Raspberry Pi, and UAV onboard computers. The computational efficiency enables real-time operation at 30+ FPS on embedded hardware, supporting autonomous navigation loops.

4. RESULTS AND DISCUSSION

This section presents comprehensive evaluation of YOLO-SCEMA across three challenging robotic scenarios: low-light object detection (nighttime autonomous driving), dense scene detection (UAV crowd surveillance), and complex image structure understanding (industrial automation). Experiments were conducted on an NVIDIA RTX 4080 GPU with Intel Core Ultra 7 155H CPU. Hyperparameters: initial learning rate 0.01, momentum 0.937, weight decay 0.0005, input resolution 640×640, training epochs 300, SGD optimizer with cosine annealing scheduler. Batch sizes: 16 for ExDark and VisDrone2019, 8 for FYP. Data augmentation: mosaic, mixup, random horizontal flip, HSV color jittering. Results represent mean $\pm$ standard deviation over three independent runs.

Evaluation on the ExDark dataset [19] (5,891 training, 1,472 test images, 12 categories, 10 illumination conditions) simulates nighttime autonomous driving and low-light robotic surveillance. Table 1 shows SCEMA consistently outperforms YOLOv8n and recent attention methods including MCA [20], GAM [21], and CBAM [8]. SCEMA achieves mAP@50 of 76.44% $\pm$ 0.28%, a 7.37% improvement over YOLOv8n (69.07% $\pm$ 0.31%), while reducing parameters by 36.88% and computational cost by 8.64%. Compared to YOLOv12n [15], SCEMA achieves 0.5% higher mAP@50 using 0.66M fewer parameters. From a robotics perspective, these improvements translate to enhanced autonomous vehicle safety during nighttime operation and more reliable low-light UAV surveillance.

Table 1. Performance comparison on ExDark dataset for low-light robotic perception

Model	mAP@50 (%)	mAP@95 (%)	Params (M)	GFLOPs
YOLOv8n [11]	69.07 $\pm$ 0.31	43.15 $\pm$ 0.28	3.01	8.1
MCA [20]	72.07 $\pm$ 0.29	44.36 $\pm$ 0.26	1.87	7.1
GAM [21]	73.39 $\pm$ 0.33	46.80 $\pm$ 0.30	3.94	9.5
CBAM [8]	75.58 $\pm$ 0.27	48.65 $\pm$ 0.25	3.07	8.2
YOLOv9t [12]	74.74 $\pm$ 0.30	48.08 $\pm$ 0.27	1.97	7.6
YOLOv10n [13]	75.47 $\pm$ 0.28	48.38 $\pm$ 0.26	2.69	8.2
YOLOv11n [14]	75.64 $\pm$ 0.29	48.73 $\pm$ 0.27	2.58	6.3
YOLOv12n [15]	75.94 $\pm$ 0.27	49.06 $\pm$ 0.25	2.56	6.3
SCEMA	**76.44 $\pm$ 0.28**	**49.38 $\pm$ 0.26**	**1.90**	**7.4**

Evaluation on VisDrone2019 [22] (6,471 training, 548 validation, 1,580 test images) simulates UAV-based crowd surveillance and dense scene understanding. Table 2 shows SCEMA achieves mAP@50 of 34.98% $\pm$ 0.32%, surpassing YOLOv8n by 3.24% and outperforming PConv [23], ECA [9], and CBAM [8] by 3.21%, 2.59%, and 2.24%, respectively. Compared to YOLOv12n, SCEMA improves mAP@50 by 0.33% while reducing parameters by 25.8%, at the cost of a modest 17.5% increase in GFLOPs. For robotics applications, these results indicate enhanced capability for UAV-based crowd monitoring, autonomous navigation in congested environments, and reliable detection of occluded objects.

Table 2. Performance comparison on VisDrone2019 dataset for dense scene robotic perception

Model	mAP@50 (%)	mAP@95 (%)	Params (M)	GFLOPs
YOLOv8n [11]	31.74 ± 0.34	18.28 ± 0.30	3.01	8.1
PConv [23]	31.77 ± 0.33	18.32 ± 0.31	1.86	7.1
ECA [9]	32.39 ± 0.31	18.53 ± 0.29	2.65	8.4
CBAM [8]	32.74 ± 0.30	18.75 ± 0.28	3.07	8.2
YOLOv9t [12]	34.58 ± 0.29	20.15 ± 0.27	1.97	7.6
YOLOv10n [13]	34.30 ± 0.31	20.01 ± 0.28	2.69	8.2
YOLOv11n [14]	34.54 ± 0.30	20.24 ± 0.27	2.58	6.3
YOLOv12n [15]	34.65 ± 0.29	20.36 ± 0.26	2.56	6.3
SCEMA	$\mathbf{34.98 \pm 0.32}$	$\mathbf{20.58 \pm 0.29}$	**1.90**	**7.4**

Evaluation on FYP dataset [24] simulates industrial automation and manufacturing environments with heavy occlusion, overlapping objects, and complex structural conditions. Table 3 demonstrates SCEMA achieves mAP@50 of $69.51\% \pm 0.30\%$, outperforming YOLOv8n by 1.50% and surpassing EMA [25], MCA [20], and CBAM [8] by 0.45%, 0.98%, and 0.27%, respectively. With only 1.90M parameters and 7.4 GFLOPs, SCEMA strikes an effective balance for industrial robotic applications requiring structural understanding.

Table 3. Performance comparison on FYP dataset for industrial automation

Model	mAP@50 (%)	mAP@95 (%)	Params (M)	GFLOPs
YOLOv8n [11]	68.01 ± 0.33	51.65 ± 0.30	3.01	8.1
EMA [25]	69.06 ± 0.31	53.16 ± 0.28	3.02	8.2
MCA [20]	68.53 ± 0.32	53.48 ± 0.29	1.87	7.1
CBAM [8]	69.24 ± 0.30	53.24 ± 0.28	3.07	8.2
YOLOv9t [12]	68.58 ± 0.31	53.53 ± 0.29	1.97	7.6
YOLOv10n [13]	68.67 ± 0.32	53.61 ± 0.30	2.69	8.2
YOLOv11n [14]	68.84 ± 0.31	53.74 ± 0.28	2.58	6.3
YOLOv12n [15]	68.95 ± 0.30	53.83 ± 0.27	2.56	6.3
SCEMA	$\mathbf{69.51 \pm 0.30}$	$\mathbf{54.01 \pm 0.28}$	**1.90**	**7.4**

Beyond standard mAP and GFLOPs, SCEMA's suitability for robotics is evaluated through additional metrics. Inference latency on NVIDIA RTX 4080: SCEMA achieves 4.2 ms per frame (238 FPS) compared to YOLOv8n's 4.8 ms (208 FPS), representing a 12.5% speed improvement. Estimated energy efficiency: SCEMA requires approximately 0.38 J per inference vs. 0.44 J for YOLOv8n (13.6% reduction), critical for battery-powered robots and UAVs. These improvements enable higher update rates for robotic control loops and extended mission duration for autonomous systems. Table 4 analyzes each SCEMA component's contribution across three datasets. Four variants are evaluated: without SCConv (spatial-channel reconstruction), without CSL, without CBAM, and full SCEMA.

Table 4. Ablation study results across robotic perception datasets

Model	SCConv	CSL	CBAM	FYP	VisDrone	ExDark
YOLOv8n	–	–	–	68.01	31.74	69.07
SCEMA w/o SCConv	–	✓	✓	68.69	31.95	72.95
SCEMA w/o CSL	✓	–	✓	68.54	32.28	73.63
SCEMA w/o CBAM	✓	✓	–	68.95	32.55	74.74
Full SCEMA	✓	✓	✓	**69.51**	**34.98**	**76.44**

Results demonstrate full SCEMA consistently achieves highest mAP@50. Removing SCConv causes performance drops of 0.82%, 3.03%, and 3.49% on FYP, VisDrone, and ExDark, respectively. Removing CSL leads to decreases of 0.97%, 2.70%, and 2.81%. The combined improvement over baseline reaches 7.37% on ExDark, indicating strong synergistic effects between spatial-channel reconstruction and cross-spatial learning.

The experimental results demonstrate SCEMA's effectiveness for robotic perception. Statistical significance testing using paired t-tests confirms all improvements are significant ($p < 0.05$), with p-values ranging from 0.003 to 0.021. Convergence analysis reveals SCEMA achieves stable training approximately 25 epochs earlier than YOLOv8n, with final loss values 0.18 lower, attributed to reduced feature redundancy enabling more efficient gradient flow. For autonomous vehicles operating at night, the 7.37% mAP improvement on ExDark directly addresses safety during low-light operation. For drone-based surveillance in crowded environments, the 3.24% improvement on VisDrone2019 enables more reliable monitoring of occluded objects. For

industrial robotics requiring structural understanding, the 1.50% improvement on FYP facilitates better scene interpretation and navigation in complex manufacturing environments.

5. CONCLUSION

This paper introduced SCEMA, a lightweight and efficient multiscale attention fusion module designed to enhance object detection for robotics and automation under complex visual conditions. By employing a parallel dual-branch architecture, SCEMA reduces spatial and channel redundancy while strengthening multiscale feature representation through cross-spatial learning. Integrated into YOLOv8, YOLO-SCEMA achieves superior detection accuracy with significantly fewer parameters (1.90M vs. 3.01M) and lower computational cost (7.4 vs. 8.1 GFLOPs). Extensive evaluations on ExDark, VisDrone2019, and FYP datasets demonstrate consistent outperformance of existing attention methods and recent YOLO variants, particularly in low-light, dense, and structurally complex scenes. The broader impact lies in enabling accurate real-time object detection in resource-constrained robotic environments, with applications in autonomous driving (nighttime operation safety), UAV surveillance (occluded object detection in crowded environments), and industrial automation (structural understanding in complex manufacturing settings). The demonstrated ability to maintain high performance under challenging conditions while reducing computational requirements represents a significant step toward practical deployment of deep learning-based perception in real-world robotic systems. Code and pre-trained models will be made publicly available.

FUNDING INFORMATION

Author state no funding is 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
Mohammed Maiza	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Chahira Cherif		✓				✓		✓	✓	✓	✓	✓		
Samira Chouraqui	✓		✓	✓		✓			✓		✓		✓	
Abdelmalik Taleb-Ahmed					✓		✓			✓		✓		✓

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

Derived data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] N. Ma, X. Zhang, H.-T. Zheng, and S. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 116–131, doi: 10.1007/978-3-030-01264-9_8.

[2] X. Lu, M. Suganuma, and T. Okatani, "SBCFormer: Lightweight network capable of full-size ImageNet classification at 1 FPS on single board computers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 1123–1133, doi: 10.1109/WACV57701.2024.00116.
[3] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv:1804.02767*, 2018.
[4] J. Li, Y. Wen, and L. He, "SCConv: Spatial and channel reconstruction convolution for feature redundancy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 6153–6162, doi: 10.1109/CVPR52729.2023.00596.
[5] M. Sohan, T. S. Ram, and C. V. R. Reddy, "A review on YOLOv8 and its advancements," in *Proceedings of the International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, 2024, pp. 529–545, doi: 10.1007/978-981-99-7962-2_39.
[6] Q. Yan, Y. Feng, C. Zhang, P. Wang, and L. Zhang, "HVI: A new color space for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 5678–5687, doi: 10.1109/CVPR52734.2025.00533.
[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
[8] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.
[9] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11531–11539, doi: 10.1109/CVPR42600.2020.01155.
[10] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.
[11] "YOLOv8: A state-of-the-art object detection model," *GitHub*, 2023. https://roboflow.com/model/yolov8.
[12] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," in *Proceedings of the European Conference on Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science, vol. 15089, 2025, pp. 1–18, doi: 10.1007/978-3-031-72751-1_1.
[13] A. Wang *et al.*, "YOLOv10: Real-time end-to-end object detection," in *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 38, 2024.
[14] C. Wang, X. Song, J. Wang, and Y. Zhang, "An improved YOLOv11 algorithm for small object detection in UAV images," *Signal, Image and Video Processing*, vol. 19, p. 507, 2025, doi: 10.1007/s11760-025-04157-w.
[15] Y. Qiao, Y. Wei, and K. Wang, "Fast-YOLOv12: An attention-guided lightweight network for real-time steel surface defect detection," *Journal of Real-Time Image Processing*, vol. 23, p. 9, 2026, doi: 10.1007/s11554-025-01807-7.
[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2117–2125, doi: 10.1109/CVPR.2017.106.
[17] M. Maiza, C. Cherif, S. Chouraqui, and A. Taleb-Ahmed, "Enhanced UAV navigation in cluttered environments via a BS-CYOLOv5 multi-sensor approach," *Discover Vehicles*, vol. 2, p. 3, 2026.
[18] Y. Wu and K. He, "Group normalization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01261-8_1.
[19] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2019, doi: 10.1016/j.cviu.2018.10.010.
[20] Y. Yu, Y. Zhang, Z. Cheng, Z. Song, and J. Tang, "MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107079, 2023, doi: 10.1016/j.engappai.2023.107079.
[21] Z.-Y. Ni and J.-C. Wang, "A global attention mechanism-based EfficientNet model for road pavement-type identification," *International Journal of Computational Intelligence Systems*, vol. 18, p. 97, 2025, doi: 10.1007/s44196-025-00842-3.
[22] D. Du *et al.*, "VisDrone-DET2019: The vision meets drone object detection in image challenge results," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2019, pp. 213–226, doi: 10.1007/978-3-030-11021-5_27.
[23] S. Park, Y.-J. Yeo, and Y.-G. Shin, "PConv: Simple yet effective convolutional layer for generative adversarial network," *Neural Computing and Applications*, vol. 34, pp. 7113–7124, 2022, doi: 10.1007/s00521-021-06846-2.
[24] "FYPproject dataset," *Roboflow Universe*, Oct. 2022. https://universe.roboflow.com/muaz-ahmed/fypproject.
[25] D. Ouyang *et al.*, "Efficient multi-scale attention module with cross-spatial learning," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10096516.

BIOGRAPHIES OF AUTHORS

Mohammed Maiza is a research scientist in the Computer Science Department at the University of Oran 1 Ahmed Ben Bella, Algeria. He holds an M.Sc. and a Ph.D. in computer science from the University of Sciences and Technology of Oran. His research focuses on bioinformatics, computational biology, and machine learning, with emphasis on high-dimensional gene selection, microarray data analysis, and hybrid evolutionary algorithms for cancer classification. He collaborates with medical researchers to develop computational methods with clinical applications in genomics and precision medicine. He can be contacted at email: maiza.mohammed@univ-oran1.dz.

Chahira Cherif is a research scientist in computer science at the University of Oran 1 Ahmed Ben Bella, Algeria, where she earned her engineering, master's, and Ph.D. degrees. Her research focuses on artificial intelligence, business process management, decision support systems, and business rules modeling, emphasizing the integration of AI into business process optimization and decision-making. She contributes to national and international research projects, supervises graduate students, and her current work centers on intelligent decision support systems using machine learning, particularly in healthcare and industrial contexts. She can be contacted at email: cherif.chahira@univ-oran1.dz.

Samira Chouraqui is a full professor in computer science at the University of Sciences and Technology of Oran (USTO MB), Algeria. She earned an engineering diploma in electrical engineering from USTO Oran, an MSc in satellite communication from the University of Surrey, UK, and a Ph.D. in computer science from USTO Oran. Her research spans computer vision, AI, UAVs, pattern recognition, and machine learning for remote sensing, with current focus on intelligent vision systems, UAV autonomous navigation, and AI-driven solutions for environmental monitoring and precision agriculture. She has supervised numerous graduate students and led several national research projects. She can be contacted at email: samira.chouraqui@univ-usto.dz.

Abdelmalik Taleb-Ahmed is a full professor at the Polytechnic University of Hauts-de-France, France, affiliated with IEMN. His research covers computer vision, machine learning, pattern recognition, image segmentation, classification, and data fusion, with applications in biometrics, video surveillance, autonomous driving, and medical imaging. He has co-authored over 85 peer-reviewed papers, supervised 30 graduate students, and led national and European research projects. His current work focuses on deep learning for medical image analysis, multimodal data fusion for autonomous systems, and advanced pattern recognition for security applications. He can be contacted at email: abdelmalik.taleb-ahmed@uphf.fr.