Object detection for indoor mobile robot: deep learning approaches review

Hind Messbah, Mohamed Emharraf, Mohammed Saber
Smart Information, Communication and Technologies Laboratory, National School of Applied Sciences, Mohammed First University, Oujda, Morocco

Article Info

Article history:

Received Oct 24, 2024
Revised Jul 19, 2025
Accepted Aug 6, 2025

Keywords:

Deep learning
Indoor robot navigation
Object detection
Real time
Sensor fusion
You only look once

ABSTRACT

Efficient object detection is crucial for enabling autonomous indoor robot navigation. This paper reviews current methodologies and challenges in the field, with a focus on deep learning-based techniques. Methods like you only look once (YOLO), region-based convolutional neural networks (R-CNN), and Faster R-CNN are explored for their suitability in real-time detection in dynamic indoor environments. Deep learning models are emphasized for their ability to improve detection accuracy and adaptability to varying conditions. Key performance metrics such as accuracy, speed, and scalability across different object types and environmental scenarios are discussed. Additionally, the integration of object detection with navigation systems is examined, highlighting the importance of accurate perception for safe and effective robot movement. This study provides insights into future research directions aimed at advancing the capabilities of indoor robot navigation through enhanced deep learning-based object detection techniques.

Corresponding Author:

Hind Messbah
Smart Information, Communication and Technologies Laboratory, National School of Applied Sciences
Mohammed First University
BP 669 Bd Mohammed VI, Oujda 60000, Morocco
Email: h.mesbahi@ump.ac.ma

1. INTRODUCTION

The field of robotics, particularly indoor navigation, has evolved significantly over the past decade, with a critical emphasis on developing robust object detection systems that enhance the ability of robots to navigate complex environments. This literature review explores key advancements in object detection technologies and their implications for indoor robot navigation, drawing insights from relevant studies. The development of autonomous mobile robots (AMRs) has revolutionized various industries, including humanitarian assistance, automotive, agriculture, education, and healthcare [1]. AMRs are designed to operate in unpredictable and partially unknown environments, requiring them to navigate complex spaces while avoiding obstacles. However, one of the main challenges confronting AMRs is their ability to perceive and interact effectively with their surroundings [2]. Object detection plays an essential role in AMR's vision systems, empowering robots to carry out intricate tasks and navigate various challenges [3]. For instance, grasp detection is essential for robots to collect objects in front of them, whereas dynamic obstacle detection is vital for real-time navigation [4]. To achieve accurate detection, AMRs rely on a combination of sensors, including navigation, localization, and detection systems. Current research indicates that sensor technology, including sensor fusion and the use of multiple sensors, can significantly impact the quality of information perceived by AMRs [5]. Computer vision is essential for numerous applications in automation and robotics, particularly in object detection. Furthermore, explainability is a critical requirement for algorithms in robotics applications, as it aids in identifying and resolving potential issues [6]. Object detection techniques like face, pedestrian, and obstacle detection rely on supervised learning in artificial intelligence (AI), typically deep learning methods. The methods employed include single-stage detectors like you only look once (YOLO), single shot multibox detector (SSD), and RetinaNet, as well as two-stage detectors, such as convolutional neural networks (CNN) and Faster region-based convolutional neural networks (R-CNN) [7], [8]. The performance of sensors and deep learning algorithms in AMRs remains a topic of ongoing discussion.

Contemporary research in deep learning has significantly influenced the design of object detection systems for indoor AMRs, particularly in terms of accuracy, adaptability, and deployment efficiency. A refined approach targeting small object detection under cluttered and complex scenes, known as YOLOv8-QSD, was proposed in [9], addressing key challenges in autonomous indoor systems. YOLOv8's capabilities were further extended through its application to light detection and ranging (LiDAR) point cloud data, demonstrating improved spatial precision for object detection in three-dimensional environments [10]. An optimized implementation of YOLOv8 was proposed to balance speed and computational constraints without sacrificing detection reliability [11]. Multi-scale feature fusion techniques have been introduced to improve the detection of variably sized objects in cluttered environments by leveraging semantic information across different spatial resolutions [12]. Depth-awareness has also been integrated into object detection pipelines to enhance performance in scenarios with occlusions and varying object distances [13]. Lightweight detectors such as EfficientDet and MobileNet offer efficient trade-offs between accuracy and processing requirements, making them suitable for embedded systems [14]. Finally, self-supervised learning strategies have been employed to improve model generalization in indoor environments while reducing the reliance on annotated datasets-an essential step for scalable deployment of autonomous robots [15].

This study seeks to evaluate the performance and the detection accuracy of deep learning techniques applied to AMRs. The literature review and results analysis are discussed in detail, providing insights into the current state of object detection techniques in AMRs. This study follows a structured approach consisting of three sections. Section 1 introduces the concept of object detection in AMRs. Section 2 analyzes the current state of object-detection techniques in AMRs. Section 3 provides insights into the challenges and opportunities facing the development of object detection techniques in AMRs. Finally, section 4 concludes the discussion.

2. METHOD

Advancements in object detection technologies have significantly impacted various fields, particularly in remote sensing. A key focus has been on detecting small objects within vast images, which presents unique challenges due to factors like resolution and object orientation [16]. Recent developments in deep learning techniques, such as the YOLO series and SSD [17], have notably improved the performance of these detection algorithms [18]. Small object detection is categorized into multiple strategies, including multi-scale predictions and enhanced feature resolutions [16]. Researchers are also addressing irregularities in remote sensing images that complicate detection efforts. Gaining insight into these methodologies not only improves the effectiveness of object detection but also highlights potential future research avenues in high-resolution environments, where detecting small objects remains a significant challenge [19].

2.1. Traditional computer vision techniques

Traditional computer vision methods have been employed for object detection in various fields, including remote sensing. These traditional techniques rely on image processing and feature extraction to identify objects within images [20]. Table 1 provides an overview of the most widely adopted traditional computer vision techniques employed in object detection tasks.

Table 1. Traditional computer vision methods overview [21]–[25]

Technique	Description	Examples	Limitations
Edge detection	Identifies object edges within an image.	Sobel operator, Canny algorithm	Affected by image noise and illumination variations.
Template matching	Compares a given image with stored templates to detect objects.	Cross-correlation, normalized correlation coefficient (NCC)	Computationally expensive, poor performance with large databases.
Feature extraction	Extracts relevant features from an image to describe objects (texture, color, shape).	Histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT)	Computationally expensive with large image databases.
HOG and SIFT	HOG describes object texture, while SIFT describes object shape and orientation for object recognition.	HOG for pedestrian detection, SIFT for object tracking	May not perform well on large datasets, and high computational cost.

2.2. Deep learning approaches

Recent advancements leverage CNNs and deep learning paradigms. Notable frameworks include YOLO, SSD, and Faster R-CNN, which have enabled real-time object detection with significant accuracy [26].

2.2.1. Single-stage models

Single-stage object detectors, such as YOLO and SSD, perform detection in a single forward pass. They are optimized for speed and are suitable for real-time applications within indoor environments [27]: i) YOLO known for its ability to perform real-time object detection. It segments the image into a grid and predicts bounding boxes along with corresponding class probabilities directly. YOLO's fast inference time makes it suitable for applications requiring real-time decisions, such as indoor robotics [28] and ii) SSD similar to YOLO, SSD performs object detection in a single forward pass, but it divides the image at multiple scales. This enhances its accuracy, particularly for smaller objects, which is crucial in indoor environments [29].

2.2.2. Two-stage models

Two-stage models, such as Faster R-CNN, first propose regions of interest and then classify those regions. These models offer higher accuracy but generally require more computational resources [30]: i) CNNs have become the foundation for object detection models due to their ability to learn features from images automatically. Architectures such as AlexNet, VGG, and ResNet have set the stage for more sophisticated detection models [31] and ii) R-CNN their variants (Fast R-CNN, Faster R-CNN) first generate region proposals and then perform classification and bounding box regression. These models are known for high accuracy but are computationally intensive, which may limit their real-time applicability [32].

2.3. Attention mechanism and transfer learning

A recent advancement in object detection and recognition technologies is the use of transfer learning combined with attention mechanisms. This cutting-edge approach enhances the model's ability to focus on important parts of an image by dynamically weighting the significance of different features, enabling more accurate and efficient object detection [33]. Transfer learning allows models pre-trained on large datasets to be fine-tuned for specific tasks, significantly reducing training time and improving performance, especially in limited data scenarios [34]. The integration of attention mechanisms within these architectures represents a major leap forward, improving the precision and speed of recognition systems in both real-time and complex environments [35]. Object detection refers to the ability to identify and locate objects within an image. Traditional methods often relied on hand-crafted features; however, with the advent of deep learning, CNNs have become the standard approach due to their ability to automatically learn features from data.

3. RESULTS AND DISCUSSION

Computer vision has evolved dramatically, moving from conventional approaches to the advanced techniques of deep learning. Traditional computer vision, which involves teaching computers to understand images through specific, programmed rules, has been largely replaced by deep learning techniques that enable computers to learn from large amounts of data, while traditional methods relied on detecting specific features like edges, shapes, or textures using algorithms, deep learning has revolutionized image interpretation by empowering computers to learn from data and adjust to a broad range of tasks. The key differences between traditional computer vision and deep learning lie in their data dependency, computational power, flexibility, and accuracy, as shown in Figure 1. Deep learning models, which are based on neural networks, can handle complex patterns and large-scale image data, and can automatically adjust to diverse tasks without being explicitly programmed for each new problem. In contrast, traditional methods require more human guidance to define features and are often less accurate than deep learning for complex vision tasks, such as image recognition and object detection in varied conditions [36].

Traditional computer vision and deep learning are not mutually exclusive, but rather complementary fields that can inform and enhance each other [37]. By studying traditional computer vision techniques, one can gain a deeper understanding of the fundamental principles of image processing and feature extraction, which are essential for deep learning models. Conversely, knowledge of deep learning can provide new insights and techniques for improving traditional computer vision methods. Ultimately, the intersection of traditional computer vision and deep learning can lead to more effective and efficient computer vision solutions, making one a more skilled and versatile expert in the field.

Modern approaches to object detection can be categorized into single-stage and two-stage detectors [7]. Single-stage detectors, such as YOLO and SSD, perform detection in a single step, combining classification and bounding box regression. These models are typically faster and simpler, making them suitable for real-time applications, but they may compromise on accuracy due to their streamlined architecture [38]. In contrast, two-stage detectors, like R-CNN and Faster R-CNN, operate in two phases: first, they generate region of interest (RoI) proposals, and then these regions are further classified and refined. While this method provides higher accuracy, particularly through RoI pooling, it comes at the cost of increased computational time and complexity [30].

Object detection is defined as identifying object instances from predefined categories within a given region, as discussed by [39]. This approach emphasizes detecting a wide variety of natural objects, avoiding limitations to specific categories like faces, trees, or vehicles. Despite the range of potential objects, research efforts have largely focused on highly structured objects (e.g., faces, airplanes) and articulated objects such as animals. Object detection supports various applications, including facial recognition, autonomous driving, and behavior analysis [40].

In large-scale surveillance systems, accurate object tracking relies on effective motion estimation and compensation techniques, as noted by [41]. The study proposed a hardware architecture incorporating real-time motion detection, estimation, and compensation, utilizing a Kogge-Stone adder to enhance operational speed. Although the method projected a 4.21% false detection rate, experimental results indicated an 11.91% rate. Additionally, Zheng *et al.* [42] proposed a cost-effective, integrated robotic system using Cartesian and articulated configurations for object detection in agricultural environments. However, the design faces challenges due to limited accuracy, necessitating human collaboration to achieve optimal performance.

Table 2 provides a summary of the advantages, disadvantages, and examples of single-stage, two-stage detectors, and transfer learning with attention; highlighting their respective trade-offs in speed, accuracy, and computational cost, particularly in indoor navigation tasks. While both approaches have their respective drawbacks, two-stage detectors typically offer superior accuracy. On the other hand, single-stage detectors are generally faster, as they avoid the complexity of multiple stages. The improved accuracy of two-stage detectors can be attributed to the inclusion of region proposal networks (RPN) or RoI pooling.

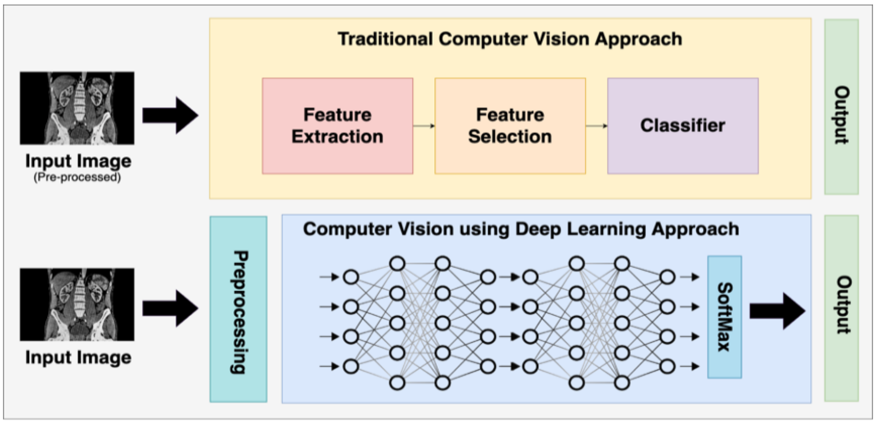

Figure 1. Deep learning vs. traditional computer vision approaches [37]

Table 2. Comparison between single-stage and two-stage detector [43], [44]

Type	How it works	Advantages	Disadvantages	Examples
Single-stage detector	A single-layer feed-forward network that performs object classification and regression to the bounding boxes.	Simpler and faster for detection.	May have reduced computational accuracy.	YOLO, YOLOv3, SSD, RetinaNet
Two-stage detector	Uses two networks. The first generates a sparse RoI, followed by classification and regression.	Offers improved accuracy through the use of RoI pooling.	Increased computational time due to multiple stages.	R-CNN, Cascade R-CNN, Faster R-CNN
Transfer learning with attention	Utilizes a pre-trained model, fine-tuned on a specific task, combined with attention mechanisms to highlight important regions of an image.	Reduces training time and requires fewer data; attention enhances focus on key image areas for improved accuracy.	May still require substantial computational resources; complex architecture.	Detection transformer (DETR), EfficientDet

3.1. Challenges

Indoor environments present unique challenges for object detection due to the following factors [45]:

– Occlusions: objects may be partially hidden behind other objects, making detection difficult.

- Varying lighting conditions: Indoor lighting can change dramatically based on time of day, artificial light sources, and shadowing effects.
- Dynamic objects: objects in motion, such as people or robots, create additional challenges for detection algorithms, particularly in environments where robots must avoid collisions.
- Real-time processing: deploying advanced models on mobile robots requires optimization for limited computational resources without sacrificing accuracy.

In summary, while deep learning approaches such as CNN-based architectures (e.g., YOLO, RCNN) have revolutionized object detection, they come with the drawback of significant computational complexity. This demand for high processing power makes their deployment on embedded systems, often used in indoor AMRs, impractical due to limited hardware resources. As a result, a more feasible solution is the integration of classical methods, such as feature extraction, with deep learning techniques. This combination enables the use of object detection models on resource-constrained embedded systems, providing a balance between performance and efficiency while overcoming hardware limitations. Addressing these challenges requires models capable of high generalization, as well as the use of sensor fusion techniques (e.g., combining camera data with LIDAR or depth sensors) to improve detection reliability [46].

3.2. Performance of object detection models

Object detection models are commonly evaluated based on precision, recall, and other metrics, with Faster R-CNN and SSD among the most accurate for indoor applications. However, there are inherent trade-offs between accuracy and inference time, which are especially critical for real-time AMRs. Models like YOLO, although slightly less accurate, often strike the best balance for real-time applications, making them well-suited for indoor environments where quick decisions are necessary [47].

In addition to precision and recall, other important performance metrics include mean average precision (mAP), which assesses accuracy across various classes, and intersection over union (IoU), which measures the overlap between predicted and ground truth bounding boxes. Furthermore, the F1 score, which combines precision and recall, serves as a balanced indicator of the system's overall performance. These metrics provide a comprehensive assessment of the accuracy, robustness, and reliability of object detection systems in AMRs [48].

Moreover, scalability and adaptability are important considerations. While SSD and YOLO efficiently detect objects in diverse environments, ensuring high performance across different lighting conditions, varying object orientations, and potential occlusions remains a challenge. Advanced techniques such as contextual reasoning and multi-scale feature extraction can improve detection accuracy, particularly in complex indoor settings [38].

Lastly, the impact of hardware limitations on performance must be acknowledged. Real-time detection systems must balance between lightweight models for deployment on embedded devices and heavier, more accurate models for server-based processing. This balance is especially important in resource-constrained environments where inference speed is crucial for decision-making.

3.3. Future directions

The future of object detection in indoor AMRs may involve [49]:

- Fusion of sensors: combining data from cameras, LiDAR, and depth sensors can provide richer contextual information, improving detection robustness.
- Self-supervised learning: approaching the issue of limited labeled datasets through self-supervised or semi-supervised learning techniques could enhance model training.
- Interpretability: as robots operate close to humans, enhancing the interpretability of machine learning models becomes necessary for trust in decision-making processes.

4. CONCLUSION

This paper has provided a comprehensive review of deep learning-based object detection techniques for indoor mobile robot navigation. It examined the transition from traditional computer vision methods to state-of-the-art deep learning models, with a particular emphasis on the YOLO family, transformer-based architectures, and multi-sensor fusion strategies. While these approaches have led to notable advancements in detection accuracy, real-time performance, and deployment feasibility, several limitations remain, particularly in handling dynamic indoor environments, limited annotated datasets, and computational constraints on embedded platforms. In our future work, we aim to prioritize the development of lightweight yet high-performance object detection models suitable for resource-constrained indoor environments. We will also explore self-supervised learning techniques to reduce dependence on annotated datasets and investigate advanced multimodal sensor fusion to improve perceptual robustness. Furthermore, we plan to design simplified and generalizable frameworks that enable reliable deployment in dynamic real-world indoor settings, thereby bridging the gap between theoretical advancements and practical applications.

FUNDING INFORMATION

The authors 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
Hind Messbah	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Mohamed Emharraf	✓	✓		✓			✓	✓		✓		✓	✓	
Mohammed Saber	✓	✓		✓						✓			✓	

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

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

Data availability does not apply to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] A. B. Rashid and M. A. K. Kausik, "AI revolutionizing industries worldwide: a comprehensive overview of its diverse applications," *Hybrid Advances*, vol. 7, Dec. 2024, doi: 10.1016/j.hybadv.2024.100277.

[2] A. Loganathan and N. S. Ahmad, "A systematic review on recent advances in autonomous mobile robot navigation," *Engineering Science and Technology, an International Journal*, vol. 40, Apr. 2023, doi: 10.1016/j.jestch.2023.101343.

[3] J. Fu, L. Zong, Y. Li, K. Li, B. Yang, and X. Liu, "Model adaption object detection system for robot," in *2020 39th Chinese Control Conference (CCC)*, Jul. 2020, pp. 3659–3664, doi: 10.23919/CCC50068.2020.9189674.

[4] S. Abdul-Khalil, S. Abdul-Rahman, S. Mutalib, S. I. Kamarudin, and S. S. Kamaruddin, "A review on object detection for autonomous mobile robot," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 3, pp. 1033–1043, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1033-1043.

[5] Y. Liu, S. Wang, Y. Xie, T. Xiong, and M. Wu, "A review of sensing technologies for indoor autonomous mobile robots," *Sensors*, vol. 24, no. 4, Feb. 2024, doi: 10.3390/s24041222.

[6] K. Okarma, "Applications of computer vision in automation and robotics," *Applied Sciences*, vol. 10, no. 19, Sep. 2020, doi: 10.3390/app10196783.

[7] J. Kang, S. Tariq, H. Oh, and S. S. Woo, "A survey of deep learning-based object detection methods and datasets for overhead imagery," *IEEE Access*, vol. 10, pp. 20118–20134, 2022, doi: 10.1109/ACCESS.2022.3149052.

[8] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sensing*, vol. 13, no. 1, Dec. 2020, doi: 10.3390/rs13010089.

[9] H. Wang, C. Liu, Y. Cai, L. Chen, and Y. Li, "YOLOv8-QSD: an improved small object detection algorithm for autonomous vehicles based on YOLOv8," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–16, 2024, doi: 10.1109/TIM.2024.3379090.

[10] S. Behera, B. Anand, and P. Rajalakshmi, "YoloV8 based novel approach for object detection on LiDAR point cloud," in *IEEE Vehicular Technology Conference*, Jun. 2024, pp. 1–5, doi: 10.1109/VTC2024-Spring62846.2024.10683316.

[11] S. Basha and G. Ram, "Real-time object detection in low-light environments using YOLOv8: a case study with a custom dataset," *International Journal of Engineering Research and*, vol. 13, 2024, doi: 10.17577/IJERTV13IS100050.

[12] A. Alotaibi, H. Alatawi, A. Binnouh, L. Duwayriat, T. Alhmiedat, and O. M. Alia, "Deep learning-based vision systems for robot semantic navigation: an experimental study," *Technologies*, vol. 12, no. 9, 2024, doi: 10.3390/technologies12090157.

[13] E. Heikel and L. Espinosa-Leal, "Indoor scene recognition via object detection and TF-IDF," *Journal of Imaging*, vol. 8, no. 8, 2022, doi: 10.3390/jimaging8080209.

[14] M. Afif, R. Ayachi, Y. Said, and M. Atri, "An evaluation of EfficientDet for object detection used for indoor robots assistance navigation," *Journal of Real-Time Image Processing*, vol. 19, no. 3, pp. 651–661, 2022, doi: 10.1007/s11554-022-01212-4.

[15] C. Tang, B. Abbatematteo, J. Hu, R. Chandra, R. Martín-Martín, and P. Stone, "Deep reinforcement learning for robotics: a survey of real-world successes," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 8, no. 1, pp. 153–188, 2025, doi: 10.1146/annurev-control-030323-022510.

[16] Z. Xiuling, W. Huijuan, S. Yu, C. Gang, Z. Suhua, and Y. Quanbo, "Starting from the structure: a review of small object detection based on deep learning," *Image and Vision Computing*, vol. 146, 2024, doi: 10.1016/j.imavis.2024.105054.

[17] Y. Sun, Z. Sun, and W. Chen, "The evolution of object detection methods," *Engineering Applications of Artificial Intelligence*, vol. 133, Jul. 2024, doi: 10.1016/j.engappai.2024.108458.

[18] L. Wang, L. Z. Chen, B. Peng, and Y. T. Lin, "Improved YOLOv5 algorithm for real-time prediction of fish yield in all cage schools," *Journal of Marine Science and Engineering*, vol. 12, no. 2, 2024, doi: 10.3390/jmse12020195.

[19] S. Gui, S. Song, R. Qin, and Y. Tang, "Remote sensing object detection in the deep learning era—a review," *Remote Sensing*, vol. 16, no. 2, 2024, doi: 10.3390/rs16020327.

[20] M. A. Hameed and Z. A. Khalaf, "A survey study in object detection: a comprehensive analysis of traditional and state-of-the-art approaches," *Basrah Researches Sciences*, vol. 50, no. 1, 2024, doi: 10.56714/bjrs.50.1.5.

[21] R. Sun *et al.*, "Survey of image edge detection," *Frontiers in Signal Processing*, vol. 2, 2022, doi: 10.3389/frsip.2022.826967.

[22] D. T. Nguyen, W. Li, and P. Ogunbona, "An improved template matching method for object detection," in *Asian Conference on Computer Vision*, 2009, pp. 193–202, doi: 10.1007/978-3-642-12297-2_19.

[23] N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision," *Technologies*, vol. 12, no. 2, 2024, doi: 10.3390/technologies12020015.

[24] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: a survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023, doi: 10.1109/JPROC.2023.3238524.

[25] P. K. Goswami and G. Goswami, "A comprehensive review on real time object detection using deep learning model," in *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, 2022, pp. 1499–1502, doi: 10.1109/SMART55829.2022.10046972.

[26] N. Alsharabi, "Real-time object detection overview: advancements, challenges, and applications," *Journal of Amran University*, vol. 3, pp. 267-278, 2023, doi: 10.59145/jaust.v3i6.73.

[27] S. M. Aamir, H. Ma, M. A. A. Khan, and M. Aaqib, "Real time object detection in occluded environment with background cluttering effects using deep learning," *The 7th International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII2021)*, 2021, pp. 1-6.

[28] G. Lavanya and S. D. Pande, "Enhancing real-time object detection with YOLO algorithm," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024, doi: 10.4108/eetiot.4541.

[29] W. Liu *et al.*, "SSD: single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[30] L. Du, R. Zhang, and X. Wang, "Overview of two-stage object detection algorithms," *Journal of Physics: Conference Series*, vol. 1544, no. 1, May 2020, doi: 10.1088/1742-6596/1544/1/012033.

[31] K. Mahajan and P. Mane, "Object detection using convolutional neural networks (CNNs): a study, implementation and challenges," *International Journal of Architecture, Engineering, and Construction*, vol. 12, no. 3, pp. 308–317, 2023.

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016, doi: 10.1109/TPAMI.2015.2437384.

[33] J. Yang, D. Chen, and H. Shi, "A object detection method based on attention mechanism and reinforcement learning," in *2022 18th International Conference on Computational Intelligence and Security, CIS 2022*, 2022, pp. 229–233, doi: 10.1109/CIS58238.2022.00055.

[34] W. Li, K. Liu, L. Zhang, and F. Cheng, "Object detection based on an adaptive attention mechanism," *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-67529-x.

[35] A. Devi *et al.*, "Transfer learning for object detection in remote sensing images with YOLO," *Journal of Electrical Systems*, vol. 20, no. 3, pp. 980–989, Apr. 2024, doi: 10.52783/jes.1412.

[36] N. O'Mahony *et al.*, "Deep learning vs. traditional computer vision," *Advances in Intelligent Systems and Computing*, vol. 943, pp. 128–144, 2020, doi: 10.1007/978-3-030-17795-9_10.

[37] OpenCV University, "Why traditional computer vision thrives alongside deep learning: a counterpoint to deep learning dominance," *LinkedIn,* 2024. Accessed: Sep. 10, 2024. [Online]. Available: https://www.linkedin.com/pulse/why-traditional-computer-vision-thrives-alongside-deep-wv0if

[38] D. S. Bacea and F. Oniga, "Single stage architecture for improved accuracy real-time object detection on mobile devices," *Image and Vision Computing*, vol. 130, Feb. 2023, doi: 10.1016/j.imavis.2022.104613.

[39] L. Aziz, M. S. B. H. Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: a comprehensive review," *IEEE Access*, vol. 8, pp. 170461–170495, 2020, doi: 10.1109/ACCESS.2020.3021508.

[40] B. R. Kiran *et al.*, "Real-time dynamic object detection for autonomous driving using prior 3D-maps," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2019, pp. 567–582, doi: 10.1007/978-3-030-11021-5_35.

[41] Z. Chen, R. Zhao, X. Guo, J. Xie, and X. Han, "Moving object detection in freely moving camera via global motion compensation and local spatial information fusion," *Sensors*, vol. 24, no. 9, Apr. 2024, doi: 10.3390/s24092859.

[42] Z. Zheng, Y. Hu, X. Li, and Y. Huang, "Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 215, no. 1, Dec. 2023, doi: 10.1016/j.compag.2023.108469.

[43] L. Shine and C. V. Jiji, "Comparative analysis of two stage and single stage detectors for anomaly detection," in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, 2021, pp. 1–6, doi: 10.1109/ICCCNT51525.2021.9580079.

[44] R. Raj and A. Kos, "Study of human–robot interactions for assistive robots using machine learning and sensor fusion technologies," *Electronics*, vol. 13, no. 16, Aug. 2024, doi: 10.3390/electronics13163285.

[45] M. Afif, R. Ayachi, and M. Atri, "Indoor objects detection system implementation using multi-graphic processing units," *Cluster Computing*, vol. 25, no. 1, pp. 469–483, 2022, doi: 10.1007/s10586-021-03419-9.

[46] H. Liu, C. Wu, and H. Wang, "Real time object detection using LiDAR and camera fusion for autonomous driving," *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-35170-z.

[47] T. Mahendrakar *et al.*, “Performance study of YOLOv5 and faster R-CNN for autonomous navigation around non-cooperative targets,” in *IEEE Aerospace Conference Proceedings*, 2022, pp. 1–12, doi: 10.1109/AERO53065.2022.9843537.
[48] J. Rezaie, “Mastering object detection metrics: from IoU to mAP,” *Medium,* 2024. Accessed: Sep. 12, 2024. [Online]. Available: https://medium.com/@rezaie.j/mastering-object-detection-metrics-from-iou-to-map-12b42ef78d4b
[49] Y. Dai, D. Kim, and K. Lee, “An advanced approach to object detection and tracking in robotics and autonomous vehicles using YOLOov8 and LiDAR data fusion,” *Electronics*, vol. 13, no. 12, Jun. 2024, doi: 10.3390/electronics13122250.

BIOGRAPHIES OF AUTHORS

Hind Messbah received her master’s degree in Big Data and her bachelor’s degree in Computer Engineering from the International University of Rabat, Morocco, in the year 2018. Presently, she holds the position of technical leader within the consulting industry. With a career spanning over five years, she has garnered extensive expertise as a data engineer, boasting a track record of successful project implementations across diverse sectors, encompassing telecommunications, insurance, retail, and banking. Her research interests include big data, artificial intelligence, robotics, and the internet of things (IoT). She can be contacted at email: h.mesbahi@ump.ac.ma.

Mohamed Emharraf is a Professor of Robotics at National School of Applied Sciences, Mohamed First University, Oujda, Morocco. He received his Ph.D. in 2017 from CEDOC-EMPO. His research interests include indoor robot control, smart agricultural, computer engineering, human-computer-interaction, and artificial intelligence. He has published 29 papers in peer-reviewed journals and conference proceedings. He has also served as a reviewer for several scientific journals and as a program committee member. He can be contacted at email: m.emharraf@ump.ac.ma.

Mohammed Saber is currently a Full Professor (PES) in the Department of Electronics, Computer Science, and Telecommunications at the National School of Applied Sciences at Mohammed First University, Oujda, Morocco (2013). He received a Ph.D. in Computer Science at the Faculty of Sciences, Oujda, Morocco, in July 2012, an engineer degree in Network and Telecommunication at the National School of Applied Sciences, in July 2004, and a Licence degree in Electronics at the Faculty of Sciences, in July 2002, all from Mohammed First University, Oujda. He is currently the Director of the Smart Information, Communication and Technologies Laboratory (SmartICT Lab). His interests include network security (intrusion detection systems, evaluation of security components, and security IoT), AI, robotics, and embedded systems. He can be contacted at email: m.saber@ump.ac.ma.