IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 3, September 2023, pp. 1033~1043
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i3.pp1033-1043

A review on object detection for autonomous mobile robot

Syamimi Abdul-Khalil[1], Shuzlina Abdul-Rahman[1,2], Sofianita Mutalib[1,2], Saidatul Izyanie Kamarudin[1], Siti Sakira Kamaruddin[3]

[1]School of Computing Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA, Selangor, Malaysia
[2]Research Initiative Group of Intelligent Systems, College of Computing, Informatics and Media, Universiti Teknologi MARA, Selangor, Malaysia
[3]School of Computing, Universiti Utara Malaysia, Kedah Darul Aman, Malaysia

Article Info

Article history:

Received Jan 17, 2022
Revised Dec 29, 2022
Accepted Jan 10, 2023

Keywords:

Artificial intelligence
Autonomous mobile robot
Deep learning
Object detection

ABSTRACT

The advancement of autonomous mobile robots (AMR) is vastly being discovered and applied to several industries. AMR contributes to the development of artificial intelligence (AI), which focuses on the growth of human-interaction systems. However, it is safe to understand that mobile robots work closely in real-time and under changing surroundings. Similarly, some limitations may affect the efficiency of mobile robots. Thus, to improve the system's efficiency and accuracy, mobile robots should adopt the ability to detect incoming obstacles accurately. This paper presents the findings of a brief technology review aimed at identifying the current state of the art and future needs for AMR in object detection. This review paper is presented in the form of a narrative-literature review. Review articles were collected from 2015 until 2022 from journals or conference papers from well-known sources like IEEE Xplore, Science Direct, Scopus, and Web of Science (WOS). The analysis of the articles was discussed in four main topics, AI, object detection, AMR, and deep learning.

Corresponding Author:

Shuzlina Abdul-Rahman
Research Initiative Group of Intelligent Systems, College of Computing, Informatics and Media
Universiti Teknologi MARA
Selangor, Malaysia
Email: shuzlina@uitm.edu.my

1. INTRODUCTION

An autonomous mobile robot (AMR) is a system that operates in an unpredictable and partially unknown environment. AMR should have unrestricted movements and avoid any incoming obstacles within a surrounding [1]. Recently, AMR has been the core of technological advancement in daily services such as humanitarian assistance. It has been used as an autonomous agent in automotive, agriculture, education, and healthcare [2]. Hence, to accomplish intelligent systems, mobile robots are acquired to work remotely with computer systems, such as moving machine parts within a factory for storage or driving automated vehicles [3].

One of the challenges for the mobile robot is visual perception alongside interaction in the real world [4]. Object detection is a crucial robot vision system that trains them to perform complex tasks and overcome prevailing complications. For instance, one of robotics' primary duties is grasp detection, which helps robots collect objects in front of them [5]. Besides that, the AMR should provide the capability to detect any dynamic obstacles in real-time [6]. To make the robot function properly, the observation and integration using the navigation system, localization systems, and detection systems (sensors), along with motion and kinematics and dynamics systems are essential [7]. This paper will focus on reviewing the statesof-the-arts that improve the accuracy of the detection system for AMR technology. Current studies of the sensor technology for AMR tackle the effectiveness of sensor fusion or multiple sensors for perception [1], [8]. Additionally, more accurate sensors may affect the quality of information perceived by AMR [9].

Computer vision is fundamental in various applications relative to automation and robotics. In robotics applications, it is a prerequisite to acquire explainability for any algorithms that help to narrow down the issues [10]. Technologies associated with object detection have scaled up to various types, such as face detection, pedestrian detection and obstacle detection [11]. These technologies call for unsupervised learning in artificial intelligence (AI) that works well with limited data and raw data typically using methods such as deep learning. There are two types of deep learning frameworks: i) single-stage and ii) two-stages detectors. Examples of techniques for single-stage detectors are you only look once (YOLO), single shot detector (SSD), and RetinaNet, and for two-stages are convolutional neural networks (CNN) and faster region-convolutional neural networks (Faster R-CNN).

The crucial questions remain whether both criteria, employing sensors in AMR systems and deep learning algorithms, help a working AMR in a real-time environment achieve greater accuracy or vice versa. Therefore, this paper will discuss the previous literature works to analyze the performances of sensors employed in AMR and deep learning techniques with detection accuracy for the AMR. The structure of this paper is as: section 1 introduces the concept of object detection for AMR. Section 2 describes the literature search strategy, while section 3 analyzes the results. Lastly, section 4 summarizes and concludes the paper.

2. LITERATURE SEARCH STRATEGY

From the points learned from the past literature, this review outlines the indicated topic related to object detection for AMR and deep learning. The method used in writing this review paper is a narrative-style literature review, as studied in [12]. The main aim of this paper is to analyze and conclude past research to avoid replications of ideas. Besides, it helps to discover and outline the unseen areas related to the research topic.

The papers are collected through an open-access site, MyKnowledge Management by Universiti Teknologi MARA, published from 2015 until 2022, to ensure up-to-date resources. The references mainly consist of research papers, journals, and conference papers from well-known repositories like IEEE Xplore, Science Direct, Scopus, and Web of Science (WOS). For practical storing, Mendeley is used as a reference management tool to help filter the collected papers according to selected criteria. In total, there were approximately 154 articles searched, and 52 were included as references. Specific inclusion criteria are full paper, open access, and English articles. The keywords light detection and ranging (LiDAR) (OR) and AMR (OR) object detection are used due to the preliminary study of the area discussed in the next section.

3. RESULTS AND DISCUSSION

This review paper explains the research domain covering AI, object detection, AMR, and deep learning. Furthermore, it will emphasize the predominant topic of the connection between multiple sensors, deep learning, and accuracy detection. All four components are correlated together and comprehensively describe the results obtained from the methodology.

3.1. Artificial intelligence

Studies about object detection for the AMR have escalated following the emergence of AI in the said field. Some papers focus on one aspect simultaneously, while others combined the techniques. Based on the reading, several writings explained generally the effectiveness of multiple sensors and deep learning in object detection for the AMR.

Different algorithms and techniques for image-based recognition in mobile robotic systems were studied in [13]. They studied the recent technology with three-dimensional (3D) scenes captured by Microsoft Kinect and time-of-flight (ToF) sensors, which have become the future technology of scientific research, engineering, and virtual reality. The main point of the paper is to explain different algorithms that contribute to autonomous vehicle (AV) navigation. The paper includes a brief introduction to each algorithm from several perspectives of signal processing, machine learning (ML), statistical learning, and neural networks. The study by [14] evaluates the capabilities and technical performance of sensors commonly used in autonomous vehicles like vision cameras, LiDAR, and radar sensors. Alzaabi *et al.* presented the least discussed primary categories of sensor calibration, a technique to notify the system of the sensors' position in an autonomous system. The paper included fusion algorithms accumulated from various established literature papers and challenges in the sensor fusion field. Their study further categorized the techniques and algorithms into classical and deep learning sensor fusion algorithms.

Another study by [15] conducted a systematic literature review on ML in object detection for security. The paper explained the impact of safety in daily life and how human beings can monitor it through the deployment of AI. The review also includes the common differences between three types of ML: i) supervised learning (SL), ii) unsupervised learning (UL), and iii) reinforcement learning (RL). Additionally, it gives the reader a run-through of what has been happening in the security field of object detection. They first identified the research questions and then searched for relevant documents related to the chosen topic. After that, the list will be sorted following the requirements to answer the questions. According to [7], robots can accomplish their tasks and goals in the real-world by focusing on the control unit or mechanical structure utilized in mobile robots [7] discusses a brief overview of the autonomous mobile robot system's localization, perception, cognition, and navigation. Each paper adequately performed extensive analysis from the preface of object detection to techniques applied using deep learning for the AMR. However, existing work did not focus on a specific review regarding the applications of many sensors to evaluate deep neural networks in attaining accuracies during obstacle detection.

3.2. Object detection

Image classification has been one of the critical tasks in achieving a better detection of objects in images. However, more is needed to perform object detection as it should combine with evaluating the concept and position of objects in an image [16]. According to [17], proper motion estimation and compensation techniques are required to track the object in large data surveillance accurately. The study proposed the hardware-level architecture involving motion detection, estimation and compensation in real-time implementation. Kogge-stone adder is utilized to improve the speed of operation of the architecture. Theoretically, the proposed method achieves a 4.21% of false detection rate, but the experiment managed to achieve 11.91% false detection rate. However, to achieve cost-effective, simple and effective solutions, an integrated robot system has been proposed [18] which uses cartesian and articulated configuration to detect objects in agricultural applications. Nevertheless, the proposed design must work collaboratively with humans as the accuracy level is low.

The study by [19] interprets the meaning of object detection as an activity that could reinstate the demographic location only if there are objects instances from the presumed categories. This task emphasizes marking over vast choices of natural objects rather than limiting them to specific categories like faces, trees, or cars. However, among the countless predefined objects, it is undeniable that most research was done on positioning exceptionally structured objects (e.g. faces, aeroplanes) and articulated objects like animals. Moreover, object detection performs different tasks for various applications like face recognition, autonomous driving, and analysis of human behaviours [20].

3.2.1. Static and dynamic objects

There are two different representations of objects or obstacles: i) static objects (SO) and ii) dynamic objects (DO). SO refer to entirely stationary that are fixed at a specific position [21]. Sometimes these objects can be seen as buildings, the surface of the roads, infrastructures, or indoor components like tables and chairs. Meanwhile, DO appear at different locations due to their moving natures. Another study defined it as a motion with differences in displacement value and time changes within the previous and current frames [17]. Detecting the DO is difficult compared to SO as it requires the camera's motion estimation [22].

There are different ways of performing object detection. Both background subtraction-mixture of Gaussian (BS-MOG) and two-frame and three-frame differencing are the enhancement of previous algorithms. A study [23] mentioned that the capabilities to execute detection contradict based on identified obstacles and the background's condition, either fixed or vice-versa. The applied technique from the same study could not detect the moving objects due to confusion about the setting's notion perceived as static or quasi-static. Like moving objects, the environment plays a significant role in determining the capacity for fixed object detection. Table 1 compiled the techniques and results for detection based on the object type to measure the issues of uncertain surrounding.

In addition, a study has been done on the non-uniform and dynamic environments, delivering an optimal path. The paper [24] proposed an algorithm that leverages the high capability of an embedded computer with a graphics processing unit (GPU) NVIDIA Jetson TX2 for computing optimal paths to objects of interest to assist blind people. The system's execution times depend highly on the environment's complexity, such as the grid size and the unknown obstacle.

3.2.2. Object detection for autonomous mobile robot

Object detection for the AMR should distinguish the surroundings precisely as a human does. Therefore, re-enact of human abilities by the AMR should cover how we process the information given and act with the solutions accordingly [25]. The standard object detection process consists of several primary stages, like identifying targeted features from the images before making predictions about the result [26]. If the prediction's result satisfies the searched objects, the output in different aspects will be produced. Figure 1 describes a series of events that occurred in object detection starting with receiving input data in images or vídeos and then pre-processing the input to remove any noises that could affect the detection's efficiency. Next, the crucial part of object detection will begin with predicting the location and scale of the selected objects from the data. Finally, the whole process will form the desired output.

Table 1. Comparison of static and dynamic object detection techniques

Type of objects	Detection techniques	Findings	First used in
Static	Background subtraction (BS)	Applying the static-background method gives less accurate results due to the illuminating changes.	Beynon *et al.* [27]
	BS–mixture of Gaussian (MOG)	Using the BS method entirely and reducing the noise effects from the static objects with MOG.	Butt and Servin [23]
Dynamic	Pixel wise system approach	Particular segmentation of moving objects by subtracting two different background models.	Porikli *et al.* [28]
	Two-frame and three-frame differencing	Using the simple OR operation to achieve the foreground objects.	Bayona *et al.* [29]

Source: adapted from [30]

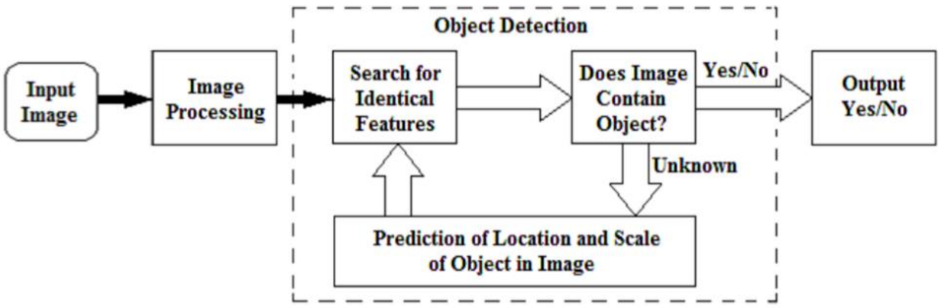


Figure 1. The standard process of object detection [25]

Nowadays, the AMR diverse application of object detection is seemingly increasing. Moreover, as agriculture, healthcare, and factories start investing in labour automation, it is vital to secure a safe projection for the AMR to improve their working efficiency. In research by [31], they produced a robotic vacuum called floor washing robot for professional users (FLOBOT) to identify obstacles like humans, house equipment, and dirt. In addition, there is a comparative analysis of tesseract optical character recognition (OCR) and Google Cloud Vision to improve object detection accuracy [32]. The proposed application is for the Thai vehicle registration certificate. The study found that Google Cloud Vision API works well for the Thai vehicle registration certificate with an accuracy of 84.43%, whereas the tesseract OCR showed an accuracy of 47.02% [32].

3.3. Autonomous mobile robot

Mobile robotics is scaling up rapidly in the fulfilment of scientific research. The intelligence of mobile robots can substitute the physical workforces in various fields because they can shift autonomously from one place to another [7]. Some mobile robot applications include rescue and research operations, surveillance, and research with education [33]. The performance of the AMR can be measured by its capacity to work within complex environments like indoors and outdoors. One of the basic principles of the AMR is understanding current environments and knowing future works that need to be accomplished [34].

3.3.1. Perception

The study by [35] described the current representation of perception as one of the AMR applications drawbacks. Perception is essential when studying mobile robots [7]. Perception is collecting information and extracting relevant knowledge from the environment. The use of sensors allows tasks to position and localize the autonomous robot. The mobile autonomous robot cannot accurately locate an object when it cannot efficiently observe the environment [7]. Previous research mentioned by [36] that a robotic system failed to recognize the visited pathway and sometimes moved differently from the planned trajectory. Another experiment by [37] saw the mobile robot becoming static as they were slow in collecting the requested object while trying to process the whole scenario. Both mishaps lead to a longer execution time and overall defeated the purpose of the AMR to work robustly in a safer environment. Regardless, ongoing studies have been conducted to counter mobile robot perception problems. The collective learning will ensure that every development of mobile robots is practical for deployment in the final phase.

3.3.2. Sensor's deployment in AMR

One of the significant tasks for mobile robots is sustaining precise knowledge of their current positioning and orientation. As a means to carry out that particular task, they are equipped with different sensor systems. The term sensor fusion methods refer to using multiple sensors that can perform various functions simultaneously. Utilizing several sensors with higher accuracy may affect the quality of information perceived by AMRs [9], [25] applied two kinect sensors to assist mobile robots with their coexistence task. This implementation helps them measure the actual bottleneck of computing distance from different directions.

Another perspective of multiple sensors was remarked by [6], which focuses on obstacle avoidance by the AMR. The study implies dynamic targets as a recurring issue that brings difficulties to object detection but managed to overcome it using multiple sensors. They applied the kinect sensor to recognize any suitable dynamic object and LiDAR to further notify incoming movements from the target. Following the techniques implemented to address the issue of perception for the AMR, it can be seen that multiple sensors work attentively to encounter them. Furthermore, the combination of gathered information enabled them to look attentively at the unnoticed space. Therefore, besides feasibly helping mobile robots to perform localization, they are deemed to execute precise navigation and recognition.

Instead of using multiple sensors, the study by [38] proposed an omnidirectional vision camera as a visual sensor for a robot to recognize the object's information. Furthermore, the proposed system utilizes PeleeNet as a deep learning model for object detection. The experiment has been compared between PeleeNet, MobileNetSSD, and SSD. The study found that the proposed system using PeleeNet has a balance between speed recognition, memory and accuracy. In addition, [39] also proposed a robot environment using an omnidirectional visual sensor equipped with a LiDAR sensor for 2D mapping in a room. The hector simultaneous localization and mapping (SLAM) algorithm is used to discover the robots position based on scan matching of the LiDAR data. Finally, the results show that the robot accurately and automatically constructs maps of the actual room with an accuracy of 95.41%. Thus it can be concluded that both types of multiple and omnidirectional sensors give the best performance for the AMR system.

3.4. Deep learning

Deep Learning is categorized as a type of ML method that processes the most prominent features from any data. Both deep learning and ML form the basis for intelligence studies, however, there is a limitation to where ML can perform in computer vision. Nevertheless, the performance of deep learning will help achieve the in-depth ML model with different algorithm advancements [40]. In robotics, implementing a deep learning algorithm helps with aspects of object detection. As it has become the most relevant domain within computer vision, real cases like autonomous vehicles, pedestrian detection, face recognition, or video surveillance depend on comprehensive algorithms research [41].

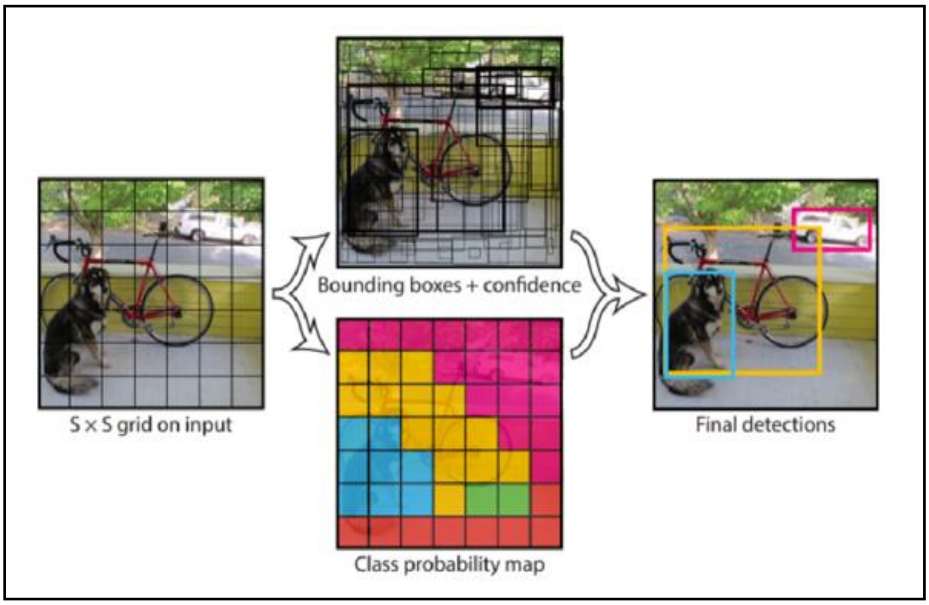
Object detection favours deep learning as it practices low-level feature development before critical enhancement. As computer vision involves extracting features from an image, known as image classification, the variety of prints allows it to adapt conveniently as inputs to deep learning [38], [42]. Nevertheless, the deep neural network architectures have varied based on performance detection due to the presence of different detectors, single-stage and two-stage detectors.

3.4.1. Single stage and two-stage detectors

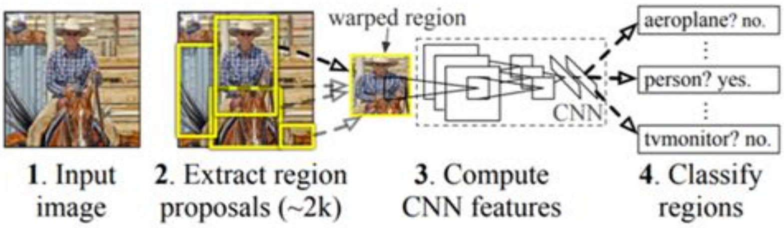
Both detectors fall under the same task category, each having compatibility to solve any surfaced problems. As the detection tasks become more challenging, modifying the traditional method into more robust and modern algorithms is crucial [43]. The main points that separate the practicality of both detectors are the processes to generate the detection. The difference relies on the architecture of the single-stage comprising one network, while the two-stage acquires integration with the single-stage and another network [44].

Figure 2(a) is an example of a single-stage detector called the YOLO model. The structure of YOLO begins during the image classification performance, where it takes inputs from the image and uses the regression method to understand its class values and bounding-boxes coordinates. The final detection will be made once the evaluation of the information is completed. Thus, the structure of single detectors can be seen as superficial, and their fast recognition of the objects confirms this [45]. Meanwhile, Figure 2(b) depicts the architecture of a two-stage detector called R-CNN. After the network has extracted the input images, it will generate a sparse region of interest (RoI) or region proposal network (RPN) to perform a specific detection.

Then, the next stage will see the selected region sent for classification and regression of the bounding boxes. As a result, the two-stage method generates better accuracy than the single-stage due to region extractions at the beginning of this network. Hence, applying a two-stage algorithm when performing object detection of dynamic objects is preferable.



(a)



(b)

Figure 2. The difference relies on the architecture of the single-stage and two-stage (a) single-stage architecture using YOLO model and (b) two-stage architecture using R-CNN model [38]

Table 2 summarises the advantages, disadvantages, and examples of single-stage and two-stage detectors. The table shows that both detectors have their respective disadvantages; however, the two-stage sensor still dominates in terms of accuracy. A single-stage sensor is usually faster for detection as it does not comprise multiple stages, unlike a two-stage detector. Furthermore, the two-stage detector, gives better accuracy due to the RPN or RoI.

The study in [46] discussed work-oriented assistive robotics, where a scenario is established for a robot to successfully reach a tool in the hand of a user when they have verbally requested it by the object's name. In addition, Useche *et al.* discussed the development of an algorithm in charge of detecting, classifying and grabbing occluded objects using AI techniques. The tools used in the study are fast region-based convolutional neural network (Fast R-CNN) and Haar classifier. It has been found that Fast R-CNN exceeded the Haar classifiers by 20% accuracy experimentally. In addition, according to [47], CNN has shown high performance in recognition of objects.

Meanwhile, Table 3 distinguishes the functionalities and deliverability of each algorithm. According to the table, it can be seen that the Faster R-CNN is the best choice to perform detection for real-time data as it can achieve higher accuracy. Furthermore, the Faster R-CNN is reported as the algorithm to achieve the highest processing speed compared to other two-stage networks [48].

Table 2. Comparison between the single-stage and two-stage detectors

Type	Functions		Examples
Single-stage detector	How it works: The single-layer feed-forward network will perform image or object classification and regression to the bounding boxes.		a. YOLO b. YOLOv3 c. SSD d. RetinaNet
	Advantages	Disadvantages	
	It is a simpler and faster algorithm for detection.	It may have less computational performance in terms of accuracy.	
Two-stage detector	How it works: This detector has two different networks. The first one will generate a sparse RoI. After that, it will further classify the images and do regression.		a. CNN b. R-CNN c. Faster R-CNN d. Cascade R-CNN
	Advantages	Disadvantages	
	This algorithm provides better accuracy in detection due to the RoI pooling.	Computational time will increase as it has different stages to go through.	

Source: adapted from [46]

Table 3. Deliverability of algorithms in both detectors

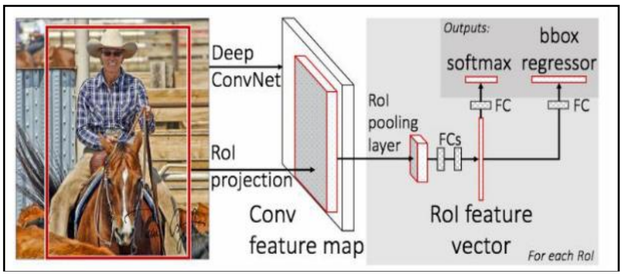
Type	Examples	Deliverability
Single-stage detector	YOLO	YOLO is anchors dependent alternative that segments the image into several regions. After that, it will give predictions and probabilities of bounding boxes.
	Single Shot MultiBox Detector (SSD)	It uses a feed-forward CNN and forms predefined numbers for bounding boxes. For each detected object, there will be confidence scores in the image.
	YOLO Version 3 (YOLOv3)	It is an enhanced algorithm of YOLO by multi-scale predictions. Moreover, it comprises a more incredible backbone network, DarkNet-53, compared to the previous one, DarkNet-19.
Two-stage detector	(R-CNN	R-CNN will perform an extra selective search that will be used to construct proposals before further detection. After that, it will imply the same method as CNN for classification and regression.
	Fast R-CNN	Fast R-CNN will send the inputs alongside multiple RoIs to a fully convolutional network. Then, each RoI will be pooled into a feature vector. The final step will perform regression of bounding boxes.
	Faster R-CNN	Faster R-CNN is the improvised version of Fast R-CNN, with shorter time processing and higher accuracy. The first stage of Faster R-CNN will undergo region proposal network before detecting with Fast R-CNN.

Source: adapted from [38], [46], [48]

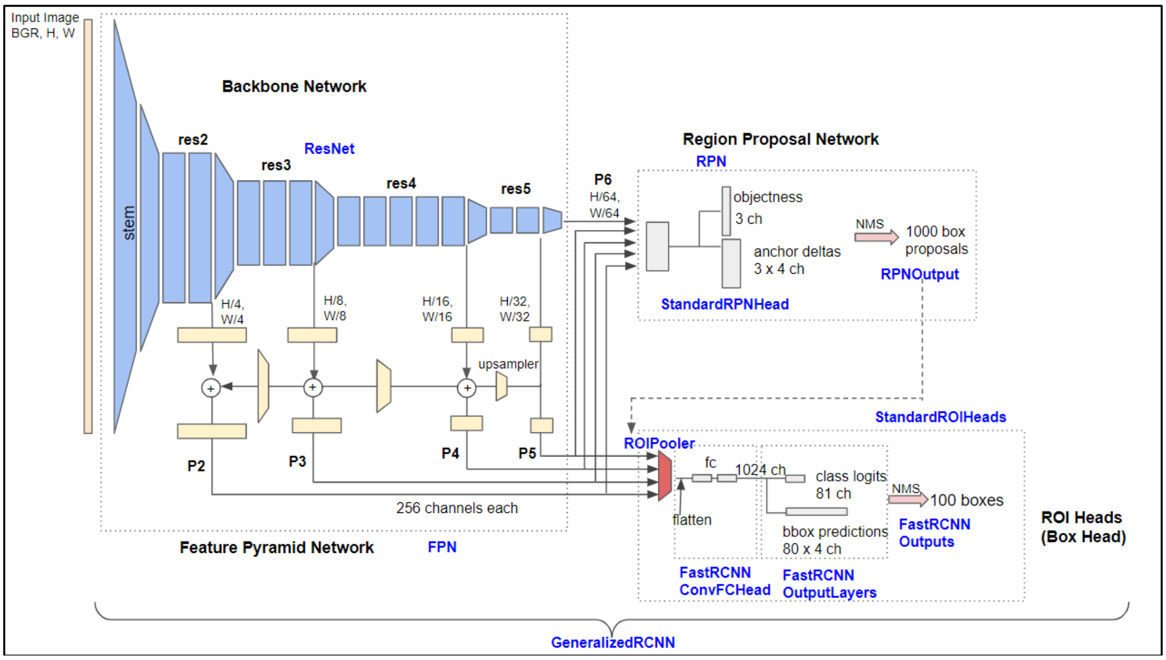
3.4.2. Faster R-CNN

Faster R-CNN is a deep learning algorithm proposed in a study by [49] to perform object detection. It is also an enhanced algorithm from the previous model, Fast R-CNN, which solves a speed bottleneck in that generation. But due to another network stage known as RPN, this algorithm has been deemed to consume detection time, although it reaches higher accuracy for detection [50]. Figure 3(a) describes how the architecture of Fast R-CNN was created with no network to perform RPN. After selecting a region, a selective search (SS) will generate a detection frame and execute max-pooling from the gathered feature maps.

Unlike Fast R-CNN, the upgraded version of it is a single and combined network for object detection. As seen in Figure 3(b), Faster R-CNN still adapted the detection stage with RoI pooling like the previous algorithm; however, there is an additional meticulous stage, RPN. RPN helps to generate the detection box synchronously, which is beneficial in increasing the processing speed. Hence, making Faster R-CNN the fastest and most accurate algorithm for two-stage detectors [51]. This state-of-the-art detector connects to the proposed project, which comprises real-time data. Therefore, in mobile robot applications that promptly perceive data, Faster R-CNN could be another stepping stone that provides precise results.



(a)



(b)

Figure 3. Architecture of (a) Fast R-CNN and (b) Faster R-CNN [48]

4. CONCLUSION

This paper presented a brief technology review to identify the current state of the art and future needs for AMR in object detection. We look into the underlying principle behind object detection; the technique of static and dynamic, the single-stage and two-stage detectors and the functionalities and deliverability of the algorithms that fall under those categories. As we examine the methods and past studies, many have reported that two-stage effectively detects any obstacles for the AMR. Besides providing a more meticulous stage to ensure the region detected is correct, it also enhances the detection probabilities. The AMR works to serve human labourers who are already precise when completing any tasks. Therefore, accuracy plays a more prominent role in mobile robot applications. As for the techniques, the two-stage detectors have various algorithms to perform modelling specifically for object detection. Past studies applied the Faster R-CNN and achieved minimal error rates. Hence, it gives an extensive overview of why the Faster R-CNN is the most suitable algorithm for AMR's static and dynamic object detection. For the employment of sensors in the autonomous system, multiple and omnidirectional sensors give the best accuracy performance for the AMR. Nonetheless, thorough research for future works is needed, particularly on other parameters that may help achieve the best accuracy.

ACKNOWLEDGEMENTS

The authors would like to express the gratitude to the Malaysian Ministry of Higher Education, grant (FRGS/1/2019/ICT02/UITM/02/10), and School of Computing Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia for the research support.

REFERENCES

[1] M. B. Alatise and G. P. Hancke, "A review on challenges of autonomous mobile robot and sensor fusion methods," *IEEE Access*, vol. 8, pp. 39830–39846, 2020, doi: 10.1109/ACCESS.2020.2975643.

[2] J. O. Pinzon-Arenas and R. Jimenez-Moreno, "Obstacle detection using faster R-CNN oriented to an autonomous feeding assistance system," in *Proceedings - 3rd International Conference on Information and Computer Technologies, ICICT 2020*, Mar. 2020, pp. 137–142, doi: 10.1109/ICICT50521.2020.00029.

[3] A. G. C. Gonzalez, M. V. S. Alves, G. S. Viana, L. K. Carvalho, and J. C. Basilio, "Supervisory control-based navigation architecture: a new framework for autonomous robots in industry 4.0 environments," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1732–1743, Apr. 2018, doi: 10.1109/TII.2017.2788079.

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

[5] V. L. Popov, S. A. Ahmed, N. G. Shakev, and A. V. Topalov, "Detection and following of moving targets by an indoor mobile robot using Microsoft Kinect and 2D LiDAR data," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Nov. 2018, pp. 280–285, doi: 10.1109/ICARCV.2018.8581231.

[6] M. S. Bahraini, A. B. Rad, and M. Bozorg, "SLAM in dynamic environments: A deep learning approach for moving object tracking using ML-RANSAC algorithm," *Sensors (Switzerland)*, vol. 19, no. 17, p. 3699, Aug. 2019, doi: 10.3390/s19173699.

[7] N. A. K. Zghair and A. S. Al-Araji, "A one decade survey of autonomous mobile robot systems," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 4891–4906, Dec. 2021, doi: 10.11591/ijece.v11i6.pp4891-4906.

[8] L. A. Nguyen, P. T. Dung, T. D. Ngo, and X. T. Truong, "Improving the accuracy of the autonomous mobile robot localization systems based on the multiple sensor fusion methods," in *3rd International Conference on Recent Advances in Signal Processing, Telecommunications and Computing (SigTelCom)*, Mar. 2019, pp. 33–37, doi: 10.1109/SIGTELCOM.2019.8696103.

[9] K. Okarma, "Applications of computer vision in automation and robotics," *Applied Sciences (Switzerland)*, vol. 10, no. 19, p. 6783, Sep. 2020, doi: 10.3390/app10196783.

[10] J. Wang, T. Zhang, Y. Cheng, and N. Al-Nabhan, "Deep learning for object detection: a survey," *Computer Systems Science and Engineering*, vol. 38, no. 2, pp. 165–182, 2021, doi: 10.32604/CSSE.2021.017016.

[11] R. Ferrari, "Writing narrative style literature reviews," *Medical Writing*, vol. 24, no. 4, pp. 230–235, Dec. 2015, doi: 10.1179/2047480615z.000000000329.

[12] T. A. Q. Tawiah, "A review of algorithms and techniques for image-based recognition and inference in mobile robotic systems," *International Journal of Advanced Robotic Systems*, vol. 17, no. 6, p. 172988142097227, Nov. 2020, doi: 10.1177/1729881420972278.

[13] D. J. Yeong, G. Velasco-hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: a review," *Sensors*, vol. 21, no. 6, pp. 1–37, Mar. 2021, doi: 10.3390/s21062140.

[14] A. Alzaabi, M. A. Talib, A. B. Nassif, A. Sajwani, and O. Einea, "A systematic literature review on machine learning in object detection security," in *IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Oct. 2020, pp. 136–139, doi: 10.1109/ICCCA49541.2020.9250836.

[15] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[16] L. Liu *et al.*, "Deep learning for generic object detection: a survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020, doi: 10.1007/s11263-019-01247-4.

[17] N. Sridevi and M. Meenakshi, "Efficient reconfigurable architecture for moving object detection with motion compensation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 2, pp. 802–810, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp802-810.

[18] V. Raja, B. Bhaskaran, K. K. G. Nagaraj, J. G. Sampathkumar, and S. R. Senthilkumar, "Agricultural harvesting using integrated robot system," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 25, no. 1, pp. 152–158, Jan. 2022, doi: 10.11591/ijeecs.v25.i1.pp152-158.

[19] 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.

[20] 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*, vol. 11133 LNCS, Springer International Publishing, 2018, pp. 1–16.

[21] M. N. Chapel and T. Bouwmans, "Moving objects detection with a moving camera: a comprehensive review," *Computer Science Review*, vol. 38, p. 100310, Nov. 2020, doi: 10.1016/j.cosrev.2020.100310.

[22] Y. Li, G. Liu, and S. Chen, "Detection of moving object in dynamic background using Gaussian Max-pooling and segmentation constrained RPCA," *arXiv preprint arXiv:1709.00657*, pp. 1–9, 2017, doi: 10.48550/arXiv.1709.00657.

[23] W. ur R. Butt and M. Servin, "Static object detection and segmentation in videos based on dual foregrounds difference with noise filtering," 2020, doi: 10.48550/arXiv.2012.10708.

[24] A. Díaz-Toro, P. Mosquera-Ortega, G. Herrera-Silva, and S. Campaña-Bastidas, "Path planning for assisting blind people in purposeful navigation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 26, no. 1, pp. 450–461, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp450-461.

[25] K. U. Sharma and N. V. Thakur, "A review and an approach for object detection in images," *International Journal of Computational Vision and Robotics*, vol. 7, no. 1–2, pp. 196–237, 2017, doi: 10.1504/IJCVR.2017.081234.

[26] Z. Yan, S. Schreiberhuber, G. Halmetschlager, T. Duckett, M. Vincze, and N. Bellotto, "Robot perception of static and dynamic objects with an autonomous floor scrubber," *Intelligent Service Robotics*, vol. 13, no. 3, pp. 403–417, Jun. 2020, doi: 10.1007/s11370-020-00324-9.

[27] M. D. Beynon, D. J. Van Hook, M. Seibert, A. Peacock, and D. Dudgeon, "Detecting abandoned packages in a multi-camera video surveillance system," in *Proceedings - IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2003, pp. 221–228, doi: 10.1109/AVSS.2003.1217925.

[28] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 1–11, Dec. 2007, doi: 10.1155/2008/197875.

[29] Á. Bayona, J. C. SanMiguel, and J. M. Martínez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques," in *6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Sep. 2009, pp. 25–30, doi: 10.1109/AVSS.2009.35.

[30] F. Fabrizio and A. De Luca, "Real-time computation of distance to dynamic obstacles with multiple depth sensors," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 56–63, Jan. 2017, doi: 10.1109/LRA.2016.2535859.

[31] F. Rubio, F. Valero, and C. Llopis-Albert, “A review of mobile robots: concepts, methods, theoretical framework, and applications,” *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, pp. 1–22, Mar. 2019, doi: 10.1177/1729881419839596.

[32] K. Thammarak, P. Kongkla, Y. Sirisathitkul, and S. Intakosum, “Comparative analysis of Tesseract and Google Cloud Vision for Thai vehicle registration certificate,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 1849–1858, Apr. 2022, doi: 10.11591/ijece.v12i2.pp1849-1858.

[33] G. Shan, T. Wang, X. Li, Y. Fang, and Y. Zhang, “A deep learning-based visual perception approach for mobile robots,” in *Proceedings 2018 Chinese Automation Congress (CAC)*, Nov. 2019, pp. 825–829, doi: 10.1109/CAC.2018.8623665.

[34] M. Lakrouf, S. Larnier, M. Devy, and N. Achour, “Moving obstacles detection and camera pointing for mobile robot applications,” in *Proceedings of the 3rd International Conference on Mechatronics and Robotics Engineering*, Feb. 2017, vol. Part F1280, pp. 57–62, doi: 10.1145/3068796.3068816.

[35] C. Y. Lee, H. Lee, I. Hwang, and B. T. Zhang, “Visual perception framework for an intelligent mobile robot,” in *17th International Conference on Ubiquitous Robots (UR)*, Jun. 2020, pp. 612–616, doi: 10.1109/UR49135.2020.9144932.

[36] M. Lee *et al.*, “Fast perception, planning, and execution for a robotic butler: wheeled humanoid m-hubo,” in *IEEE International Conference on Intelligent Robots and Systems*, Nov. 2019, pp. 5444–5451, doi: 10.1109/IROS40897.2019.8968064.

[37] Y. N. Dong and G. S. Liang, “Research and discussion on image recognition and classification algorithm based on deep learning,” in *Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Nov. 2019, pp. 274–278, doi: 10.1109/MLBDBI48998.2019.00061.

[38] Winarno, A. S. Agoes, E. I. Agustin, and D. Arifianto, “Object detection for KRSBI robot soccer using PeleeNet on omnidirectional camera,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4, pp. 1942–1953, Aug. 2020, doi: 10.12928/TELKOMNIKA.V18I4.15009.

[39] M. Rivai, D. Hutabarat, and Z. M. Jauhar Nafis, “2D mapping using omni-directional mobile robot equipped with LiDAR,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1467–1474, Jun. 2020, doi: 10.12928/TELKOMNIKA.v18i3.14872.

[40] M. Abagiu, D. Popescu, F. L. Manta, and L. C. Popescu, “Use of a deep neural network for object detection in a mobile robot application,” in *Proceedings of the 2020 11th International Conference and Exposition on Electrical And Power Engineering (EPE)*, Oct. 2020, pp. 221–225, doi: 10.1109/EPE50722.2020.9305648.

[41] R. Kalliomäki, “Real-time object detection for autonomous vehicles using deep learning.” p. 92, 2019, doi: 10.1007/978-981-33-6862-0_28.

[42] W. Hongtao and Y. Xi, “Object detection method based on improved one-stage detector,” in *5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Jun. 2020, pp. 209–212, doi: 10.1109/ICSGEA51094.2020.00051.

[43] X. Long, Z. Zheng, Y. Chi, and R. Liu, “A mixed two-stage object detector for image processing of power system applications,” in *International Conference on Communication Technology Proceedings (ICCT)*, Oct. 2020, vol. 2020-Octob, pp. 1352–1355, doi: 10.1109/ICCT50939.2020.9295843.

[44] P. Soviany and R. T. Ionescu, “Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction,” in *20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sep. 2018, pp. 209–214, doi: 10.1109/SYNASC.2018.00041.

[45] 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, pp. 1–23, Dec. 2021, doi: 10.3390/rs13010089.

[46] P. Useche, R. Jimenez-Moreno, and J. M. Baquero, “Algorithm of detection, classification and gripping of occluded objects by CNN techniques and Haar classifiers,” *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 4712–4720, Oct. 2020, doi: 10.11591/ijece.v10i5.pp4712-4720.

[47] R. Jiménez-Moreno, J. O. Pinzón-Arenas, and C. G. Pachón-Suescún, “Assistant robot through deep learning,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 1053–1062, Feb. 2020, doi: 10.11591/ijece.v10i1.pp1053-1062.

[48] C. Wang and Z. Peng, “Design and implementation of an object detection system using faster R-CNN,” in *International Conference on Robots and Intelligent System (ICRIS)*, Jun. 2019, pp. 204–206, doi: 10.1109/ICRIS.2019.00060.

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

[50] B. Liu, W. Zhao, and Q. Sun, “Study of object detection based on faster R-CNN,” in *Proceedings - 2017 Chinese Automation Congress (CAC)*, Oct. 2017, vol. 2017-Janua, pp. 6233–6236, doi: 10.1109/CAC.2017.8243900.

[51] M. Karthikeyan and T. S. Subashini, “Automated object detection of mechanical fasteners using faster region based convolutional neural networks,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5430–5437, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5430-5437.

BIOGRAPHIES OF AUTHORS

Syamimi Abdul Khalil is an undergraduate student from College of Computing, Informatics and Media, in Universiti Teknologi MARA (UiTM), where she pursued a Bachelor's Degree in Information Systems (Hons.) Intelligent Systems Engineering. She was the recipient of an Endowment Scholarship from UiTM, which has allowed her to engage in students' development while expanding her connection with the university. In her studies, Syamimi focuses on studying artificial intelligence in aspects of machine learning, deep learning, optimization, data analytics, data visualization, and data mining. Her background study also includes learning Computer Science in the Diploma program. She can be contacted at email: syamimikhalil99@gmail.com.

Shuzlina Abdul Rahman is an Associate Professor of Information Systems at the College of Computing, Informatics and Media, in Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia. She received the Ph.D. degree in Science and Systems Management, specialized in data mining and optimization from the Universiti Kebangsaan Malaysia (UKM) in 2012. Currently, She is the head of Intelligent Systems Research Group, UiTM. Her publication fingerprints include data mining, feature selection, mobile robots, sentiment analysis, and object detection. Her teaching and primary research interests involve the computational intelligence, data mining and optimization and intelligent data analytics. She can be contacted at email: shuzlina@uitm.edu.my.

Sofianita Mutalib is currently a senior lecturer in the College of Computing, Informatics and Media, in Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia. She received the Ph.D. degree, specialized in Data Mining in 2019 from UiTM. She teaches courses related to intelligent systems such as intelligent system development, decision support systems and data mining. Her primary research interest involves data analytics for unstructured data and also data science. She can be contacted at email: sofianita@uitm.edu.my.

Saidatul Izyanie Kamarudin obtained her Bachelor of Engineering in Communication Engineering (Honors) from International Islamic University Malaysia, Malaysia in 2011. She received her Master of Science in Communication System Engineering from the University of Portsmouth, Portsmouth, United Kingdom in 2013. Then, she got her PhD in Network Communication Engineering from Universiti Putra Malaysia in 2021. Currently, she joined the College of Computing, Informatics and Media, in Universiti Teknologi MARA, UiTM as a lecturer. Her teaching includes artificial intelligence, data analytics and machine learning. Her research interest includes intelligent systems, data science, the internet of things, wireless power transfer, RFID, antenna and microwave systems. She can be contacted at email: saidatulizyanie@uitm.edu.my.

Siti Sakira Kamaruddin received her Diploma in Computer Science from Universiti Putra Malaysia (UPM), Serdang, Selangor, Malaysia in 1990, the bachelor's and master's degrees in Computer Science from Universiti Teknologi Malaysia (UTM), Skudai, Johor, Malaysia in 1995 and 1998, respectively, and the Ph.D. degree in Science and System Management from the Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia in 2011. She is currently an Associate Professor at the School of Computing, Universiti Utara Malaysia (UUM), Sintok, Kedah, Malaysia. She has published various papers in scientific journals and international conferences in the area of computational intelligence, text mining, natural language processing, social media analytics, and data mining. She can be contacted at email: sakira@uum.edu.my.