Journal of Basrah Researches (Sciences) **50**(1), 46 (2024)
DOI: https://doi.org/10.56714/bjrs.50.1.5

A survey study in Object Detection: A Comprehensive Analysis of Traditional and State-of-the-Art Approaches

Marwa A. Hameed[1], Zainab A. Khalaf[2*]

[1]Department of Computer Science, College of Computer Science & Information Technology, University of Basrah, Basrah, Iraq.
[2]Department of Computer Science, College of education for Pure Sciences, University of Basrah, Basrah, Iraq.

ARTICLE INFO

Received 11 December 2023
Accepted 04 February 2024
Published 30 June 2024

Keywords:

Object Detection, Deep Learning, Traditional Detectors, Image Object Detection.

Citation: M.A. Hameed, Z.A. Khalaf, J. Basrah Res. (Sci.)**50**(1), 46 (2024). DOI:https://doi.org/10.56714/bjrs.50.1.5

ABSTRACT

Object detection is an essential field within computer vision, focusing on identifying objects' presence and category within image or video data. The significance of this issue is paramount in numerous domains that directly impact people's lives, including autonomous driving, healthcare systems, and security monitoring. In contrast to traditional methodologies employed for object detection, deep learning-based algorithms have demonstrated substantial progress in computational efficiency and precision in recent years. This study aims to provide a comprehensive review of object detection by methodically employing deep learning to facilitate a comprehensive and in-depth comprehension of the fundamental principles in this field. The discussion has encompassed various subjects, such as the obstacles and complexities associated with object detection and the traditional and deep learning detectors. The detection of objects within images and videos, the real-time detection of objects, detection of 3D objects, commonly used datasets, and the metrics employed for evaluating object detection performance. This study will likely yield scientific benefits for academics working in the field of object detection and deep learning.

1. Introduction

Object detection stands as a critical pillar in the domain of computer vision, encompassing the pivotal goal of not only recognizing but also precisely locating instances belonging to a specific class of visual objects within the intricate landscape of digital images. As a central challenge in computer vision, object detection doesn't merely fulfill a singular role but acts as the foundational bedrock for addressing a spectrum of advanced visual tasks. These tasks span a wide array of complexities, including but not limited to object tracking, segmentation, image captioning, scene understanding, activity recognition, and event detection [1].The exploration of object detection is not confined to the academic realm; it extends its reach into the pragmatic domains of real-world applications. These applications are as diverse as the challenges they seek to address, ranging from the precision demands of autonomous driving and the nuanced intricacies of robotic vision to the vigilant oversight in security

*Corresponding author email : zainab.khalaf@uobasrah.edu.iq

ISSN: 1817-2695 (Print); 2411-524X (Online)
Online at: https://jou.jobrs.edu.iq

monitoring, the insightful analysis in drone scene examination, and the comprehensive surveillance required in transportation systems [2]. The existing literature on object detection could be categorized into three distinct groups, as elucidated subsequently. Figure 1. illustrates the fundamental differentiation between them:

1. **Object Detection (OD):** At the forefront of these categorizations is Object Detection, an endeavour focused on the identification of objects with an emphasis on their location, devoid of consideration for their class category. Within the realm of OD, algorithms are strategically designed to generate a multitude of region proposals within an image. The subsequent selection process involves choosing optimal candidates based on predefined criteria [2]. Pioneering techniques like R-CNN [3] and its derivatives showcase the efficacy of OD in various visual recognition tasks.

2. **Salient Object Detection (SOD):** A unique facet within this classification is Salient Object Detection (SOD). Drawing inspiration from the intricacies of human attention mechanisms, SOD methods aspire to recognize and localize objects that stand out within an image or video. The emphasis here is not on the identity of objects but on their visual prominence. SOD plays a pivotal role in applications where highlighting visually significant regions is paramount, such as computer graphics, image understanding, robot navigation, and video comprehension [4]. Cutting-edge models, including Deep Saliency and SALICON [5], showcase the prowess of SOD in capturing and accentuating salient elements.

3. **Category-specific Object Detection (COD)**: Distinguishing itself within this tripartite categorization is Category-specific Object Detection (COD). COD goes beyond the realm of mere identification by not only detecting multiple objects but also specifying their class categories within the image or video. This dual focus on object presence and categorization is pivotal for applications where a nuanced understanding of both elements is crucial. Techniques like You Only Look Once (YOLO) [6] and SSD (Single Shot multi-box Detector) [7] exemplify the efficiency and accuracy achievable through COD methodologies [2].

This three-part classification doesn't just help categorize various object detection methods, but it also gives researchers and practitioners a comprehensive framework to handle the complexities involved in visual recognition tasks. The careful arrangement of these categories sets the stage for progress in computer vision, offering the potential for a better grasp and use of visual data in many applications.

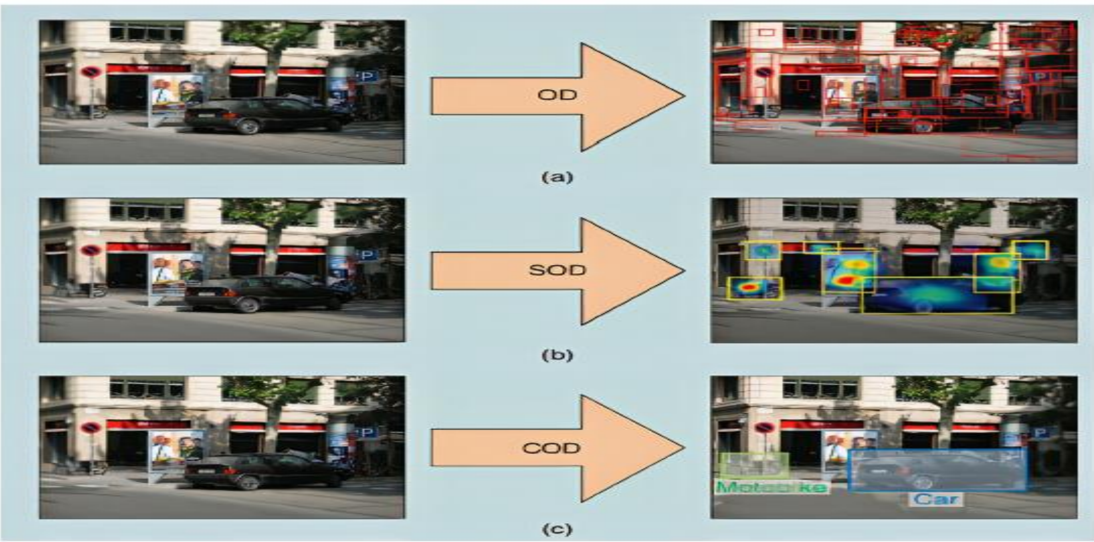

Fig. 1. Visual representation of the three possible object detection directions [2].
(a) OD: Object Detection, (b) SOD: Salient Object Detection, (c) COD: Category-specific Object Detection.

This study thoroughly explores object detection methods that leverage deep learning. The article delivers a detailed summary of the recent developments in object detection and computer vision, intending to provide researchers with substantial information given the swift advancements in this domain. The key contributions of this paper can be summarized as follows:

- This article summarized relevant literature on object detection utilizing deep learning approaches.
- A compilation of the data sets that are publicly available for object detection have been summarized.
- More over this paper differs from previous object detection surveys by presenting the different types of images used to detect objects based on deep learning.

The subsequent sections of the research are structured as follows. Section 2 pertains to relevant surveys on object detection, while Section 3 outlines the obstacles and complexities encountered in object detection. Section 4 elucidates conventional techniques and juxtaposes them with deep learning methodologies. Section 5 and Section 6 of the study concentrate on the identification of objects utilizing Remote Sensing Images (RSI), Synthetic Aperture Radar (SAR) images, and images obtained from drones and Unmanned Aerial Vehicles (UAV). Sections 7 and 8 elucidate the concept of object detection utilizing video data rather than static images and the implementation of real-time detection. Section 9 elucidates the identification of three-dimensional objects and their respective classifications, while Section 10 expounds upon prevalent datasets and assessment criteria.

2. Related Object Detection surveys

Over the past few years, there has been a notable increase in the number of published reviews concerning object detection, as can be shown in Table 1. Most of these studies focus on applying deep learning for broad object detection. Recent years have seen a proliferation of assessments concentrate not on a particular object or application but on object detection in general. However, the reviews summarized in this paper date back to the previous four years.

Table 1. Overview of corresponding surveys on object detection in the last four years.

References	Year	Content
[2]	2019	A survey of describing and analyzing object detection-based deep learning and reviewing the typical detection models.
[8]	2019	This survey reviews 400+ papers(from 1999 to 2019) of object detection.
[9]	2019	Comprehensive and accessible presentation of recent advancements in object detection utilizing deep learning.
[10]	2019	A review of convolutional neural networks models and applications to object detection.
[11]	2019	A detailed assessment of current improvements in visual object detection-based deep learning.
[4]	2020	Introduction of traditional and deep convolutional neural networks object detection algorithms.
[12]	2020	This survey encompasses approximately 300 recent publications on visual object detection.
[13]	2021	This paper reviews recent advances in deep learning-based object detectors.
[14]	2021	This survey examines the most recent advancements in object detection and briefly introduces the literature reviews.
[15]	2022	This study covers the current advancements in remote sensing object detection technologies, including classical and deep learning approaches.
[16]	2022	This paper presents an extensive analysis of existing literature that explicitly examines object detection.

3. Object Detection Difficulties and Challenges

The field of object detection for practical applications encounters numerous problems and complexities that have a direct impact on the accuracy and efficiency of the detection procedure. Among the fundamental challenges that exist are:

- **Dense Occlusion** The issue of dense occlusion is a common occurrence in various practical application settings such as pedestrian detection and autonomous driving. The phenomenon can be categorized into two scenarios: occlusion occurring between objects belonging to the same category, and occlusion occurring between objects belonging to distinct categories. The occurrence of occlusion can lead to the loss of object information, resulting in the potential for missed or inaccurate detection. Researchers can employ the new item information in typical object detection methods. To tackle the dense occlusion problem, use grey information, boundary information, and local characteristics [4].
- **Intra-Class Variation** The intra-class variance among variants of the same object is, by definition, somewhat constant. This variation may be attributable to a number of factors, including illumination, position, occlusion, and viewpoint, among others. These unrestricted exteriors can significantly alter the aspect of the object. The objects are anticipated to have non-rigid deformation or to be rotated, scaled, or fuzzy. Some objects may be surrounded by unobtrusive surroundings, making extraction challenging [12, 16].
- **Class Imbalance** Class imbalance contributes to the diminished accuracy observed in one-stage object detection compared to two-stage object detection. Utilizing the two-stage object detection zone concept has effectively mitigated the class imbalance issue. The class consists of challenging positive examples, challenging negative examples, straightforward positive examples, and straightforward negative instances. The quantity of difficult instances is smaller than that of straightforward ones. OHEM, Focal Loss, CC-Net, and RON are representative solutions for overcoming this problem [4].
- **Efficiency Challenge** The efficiency problems arise due to the requirement of localizing and recognizing objects, whereby the computing complexity escalates with the possibility of a substantial number of item categories and an extensive range of locations and scales inside a singular image. In order to provide accurate detection results, contemporary models require a substantial allocation of computer resources. Given the widespread use of mobile and peripheral devices, efficient object detectors are crucial in advancing computer vision [12, 16].
- **Multi-scale object detection** poses significant challenges in the domain of object detection. A typical hierarchical structure is the convolutional neural network. Each layer outlines the image's feature map, and the semantic information conveyed by the maps of the feature differs. Based on DCNNs, this property by itself specifies the method for detecting multi-scale objects. The RCNN and YOLO, for instance, only serve the classification of objects as well as the bounding-box regression on the final layer of feature maps. This results in a significant loss of representation information of object features, which is plainly detrimental to the recognition of multi-scale objects. YOLO, for instance, lacks robustness in detecting small objects because it fails to capture the properties of small objects in the last convolutional layer [4].
- **Small object detection** Due to indistinguishable characteristics, low resolution, complex backgrounds, limited context information, and so forth, small objects are challenging to detect [18]. In recent years, the notable achievements of deep learning methodologies have revitalized the field of small object detection, moving it to the vanguard of research attention. Small object detection has been extensively employed in academic research and practical scenarios, including domains such as robot vision, autonomous driving, intelligent transportation, drone scene analysis, military reconnaissance, and surveillance [19]. Several conventional detection techniques, such as Faster RCNN, YOLO, and SSD, have limitations in effectively recognizing small objects [4].

4. Traditional detectors vs. deep learning detectors

Deep learning (DL) is a state-of-the-art methodology utilized for extracting highly accurate features from images and analyzing data, resulting in remarkable outcomes and capabilities. The phrase "deep learning" denotes a category of machine learning techniques. This addition enhances the complexity of the model [20]. Recently, significant advancements have been witnessed in deep learning across many domains, encompassing speech recognition, stock market prediction, video surveillance, weather forecasting, and object detection [17- 20]. Modern object detectors typically use deep learning networks as their detection network and backbone to extract features from input images (or videos), classify objects, and locate them [2], as shown in Fig. 2.

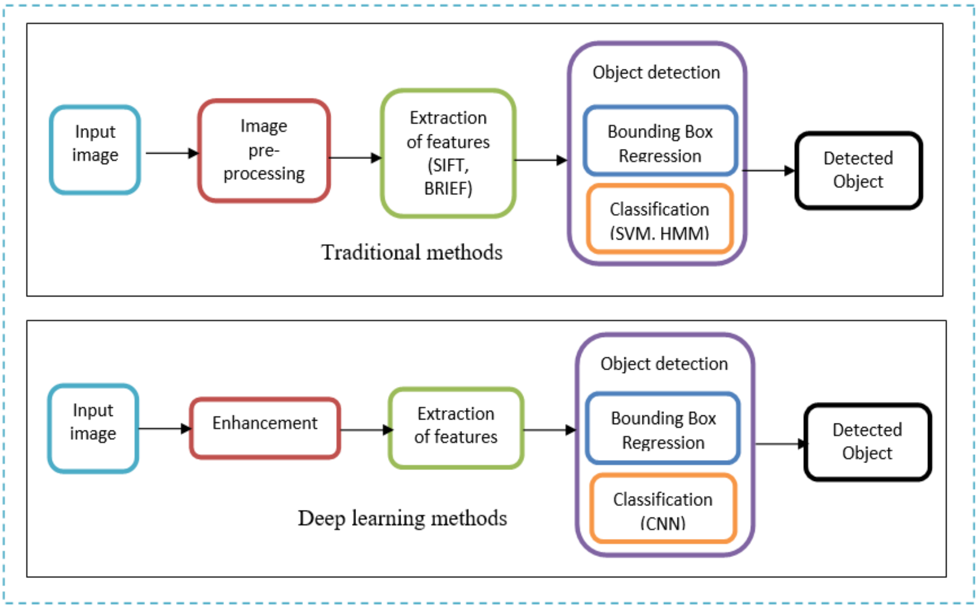

Fig. 2. Comparison of object detection based classic machine learning methods and deep learning approaches.

In the traditional machine learning approach, object detection is regarded as a classification problem. The model initially scans an image to identify potential regions containing objects and then extracts various features such as histogram of gradients (HOG), bag of words (BoW), texture, contextual, and other relevant information from these regions. Subsequently, it employs a separate classifier to differentiate between several item types in order to ascertain the presence of objects within the sub-region. Due to limitations such as the inability to train the feature extractor and classifier seamlessly, significant computing burden, and an imprecise positioning function, the machine learning approach has been progressively substituted by deep learning (DL), and its utilization is gradually discontinued [15].The evolution to state-of-the-art approaches, particularly those using deep learning, marks a significant shift. Unlike traditional detectors, modern ones leveraging deep learning networks can automatically learn intricate features from data, leading to more accurate and flexible detection. The deep learning-based detectors excel in handling complex patterns and variations in images, providing a notable improvement in object detection performance compared to their traditional counterparts. This shift towards deep learning signifies a paradigmatic advancement in the field of object detection [15]. Deep learning-based object detection approaches can be categorized into two types: two-stage detectors and one-stage detectors [23]. The two-stage detectors employ a two-step process to identify items within an image. These detectors frequently yield cutting-edge outcomes or exceptional precision on existing datasets. However, these detectors demonstrate a slower inference speed than one-stage detectors [24]. Region-based Convolutional Neural Network (R-CNN) [3] and Fast Region-based Convolutional Neural Network (Fast R-CNN) [25] are the most famous two-stage algorithms. The one-stage detector is primarily used in real-time object identification and offers significantly faster results compared to two-stage detectors [16]. YOLO [6] and SSD [7] are the most famous one-stage algorithms. Figure 2 shows a comparison between classic machine learning and deep learning detectors. For more details, Table 2 states the key differences between traditional and state-of-the-art object detectors.

Table 2. Comparison of traditional and deep learning detectors

Aspect	Traditional Approach Detector	State-of-the-Art Approach Detector (Deep Learning)
Feature Extraction	Manual selection of features	Automatic feature learning from data
Model Complexity	Simple models with limited layers	Complex models with deep neural networks
Adaptability	Limited adaptability to varied tasks	High adaptability to diverse challenges
Performance	May plateau on complex datasets	Superior accuracy and handling complexity
Data Requirements	Relies on predefined features	Can learn from less data but benefits from large datasets
Training Process	Manual tuning and feature engineering	End-to-end training with automated feature learning
Flexibility	Limited flexibility in handling varied tasks	Highly flexible, excelling in diverse applications

5. Image object detection

The image serves as the foundational component in the object detection process across images that differ in size and resolution. Image object detection has consistently posed a complex issue in the domain of computer vision, garnering significant research interest. Due to the advancements in deep convolutional neural networks and the increased speed of GPU processing, image object detection technology has experienced fast growth [14]. Obtaining images for the detection process can be accomplished by accessing numerous datasets that are accessible online. Some of these datasets have been previously addressed in this research. Object detection primarily involves the analysis of two types of images:

5.1 Remote Sensing Images (RSIs)

RSIs images are generated by capturing information about a target object using sensors located on distant platforms. This data is obtained via electromagnetic waves that are reflected, transmitted, or scattered [27]. Remote sensing images have been increasingly crucial since a decade ago, with object detection being a critical task that has direct relevance to a wide range of applications, such as mapping natural resources, analyzing crop harvests, managing disasters, planning traffic, and aiding navigation. Remote sensing images encompass a wide range of items of interest, varying in size from small to enormous. Hence, the problem of detecting objects in remote sensing images is complex since it involves many scales. However, recent advancements in deep learning have resulted in significant progress in the recognition and localization of objects [26].

5.2 Synthetic Aperture Radar (SAR) images

SAR images are exact images of a vast expanse of the Earth's surface, obtained using a specialized form of radar known as SAR [27]. SAR image-based object detection is crucial for land surveillance, military information gathering, and marine management [28].

Object detection is an essential issue in the realm of SAR image interpretation. Its objective is to locate and identify targets of interest from SAR imaging precisely. Extensive research has been conducted on this topic for many years. With the swift advancement of SAR imaging techniques, accessing high-resolution SAR images has become more accessible, thus providing more excellent research opportunities for complex SAR target detection tasks [29].

6. Drone and UAVs (Unmanned Aerial Vehicles) Captured Images

Drones have recently experienced significant growth and are projected to be integrated into numerous industries in the future, offering substantial benefits. Affordable drone photography technology, specifically, can assist local businesses and researchers in investigating cultural heritage areas along the coastline [30]. Object detection methods are extensively utilized in practical applications such as plant conservation, wildlife safeguarding, and urban surveillance, using scenes filmed by drones. Directly applying past models to the task of detecting items in drone-captured scenes has three notable disadvantages. The object scale varies significantly as a result of fluctuations in drone flight height. Furthermore, things with high density in drone-captured photographs can cause occlusion between other objects. Furthermore, as a result of the vast expanse covered, drone-captured images consistently exhibit intricate geographical features [31]. Due to the growing demand for autonomous cars in various terrains, computer vision is making significant advancements to enable these vehicles to identify the elements present in their surroundings accurately. New specialized integrated hardware and software are regularly being produced, which enable faster processing and seamless integration with UAV systems [32]. See Fig. 3.

Fig. 3. Drone-captured images of experimental situations [30].

7. Video object detection

Video object detection includes the detection of objects in video data and is of great importance in video analysis. Compared to still picture object detection, video object detection is more difficult due to motion blur, variable viewpoints/poses, and occlusion, as shown in Fig. 4. Existing approaches used temporal information during video detection and outperformed static-image detectors [33]. Although there are proven object detection methods that work well with static images, they have two disadvantages when applied to frame-by-frame video data: (i) Insufficient computational efficiency caused by redundant information across image frames or the lack of chronological and spatial relationship of information between consecutive image frames, and (ii) Insufficient ability to handle real-world scenarios such as motion blur and occlusion [34]. Recently, there has been an increase in interest in the field of object detection in videos, which is significant in a wide variety of real-world applications like video surveillance, automated driving, intelligent robotics, and so on [29, 31]. Video detection commonly utilizes image detection methods due to the similarities between the two. Nevertheless, when utilized on a dataset of video data, an object detection algorithm necessitates more stringent criteria due to the multitude of phenomena present, including motion blur, morphological diversity, occlusion and variations in illumination inside the video [36]. As illustrated in Fig. 4.

Fig. 4. An illustration of the typical difficulties connected with video object detection [33].

8. Real time object detection

The primary aim of real-time object detection is to precisely determine the location of an object within a given image and assign it to the corresponding category [37]. In recent times, there has been a rise in the development of algorithms such as DCNN, RCNN, YOLO, and SSD. These algorithms have demonstrated exceptional performance in the domain of real-time object detection. However, the video surveillance sector necessitates using potent gear to employ them effectively [35]. Real-time object detection systems must adhere to the real-time limitations, which may vary based on the specific execution contexts in which they operate. Given the widespread adoption of these systems in real-time situations, it is imperative to ensure that time limitations are not overlooked, as failure to do so could potentially result in catastrophic events such as collisions. Therefore, these systems must be capable of promptly detecting objects within a particular period. The time limit could differ depending on the actual execution conditions [38].

9. 3D object detection

It is customary to depict 3D objects as 3D boxes in point-cloud representations. The present depiction emulates the extensively researched technique of image-based 2D bounding-box detection, although with the inclusion of supplementary complexities [39]. Detecting three-dimensional objects in point clouds is critical in various practical domains, including autonomous driving and augmented reality [35, 37], as shown in Fig. 5. In contrast to the extensively researched 2D detection problem, the task of 3D detection on point-clouds presents a range of intriguing obstacles:

- The point-clouds exhibit sparsity, with many 3D objects lacking measurements.
- The output obtained is a three-dimensional box frequently needing proper alignment with a global coordinate frame.
- 3D objects exhibit diverse dimensions, forms, and proportions. For instance, in traffic, bicycles tend to have a nearly planar shape, whereas buses and limousines possess an elongated structure, and pedestrians typically have a taller stature [39].

The approaches for processing standard features in 3D detection can be classified into three categories depending on the type of point cloud representation utilized:

- Voxel-based approaches: Voxel-based methodologies include partitioning irregular point clouds into regular voxels, subjected to sparse 3D convolutions to acquire high-dimensional features.

While voxel-based techniques have proven beneficial, they challenge balancing efficiency and accuracy. In particular, employing smaller voxels enables enhanced precision at the expense of increased processing burden. On the contrary, utilizing bigger voxels eliminates potential localized information inside the densely populated voxels [35, 36].

- Point-based approaches: Point-based methodologies directly utilize raw points to acquire 3D representations. This approach effectively addresses the limitation of turning point clouds into regular structures. By utilizing learning techniques for point sets, point-based systems effectively circumvent information caused by voxelization. These approaches also capitalize on the sparsity inherent in point clouds, as they solely perform computations on valid data points [35, 36].

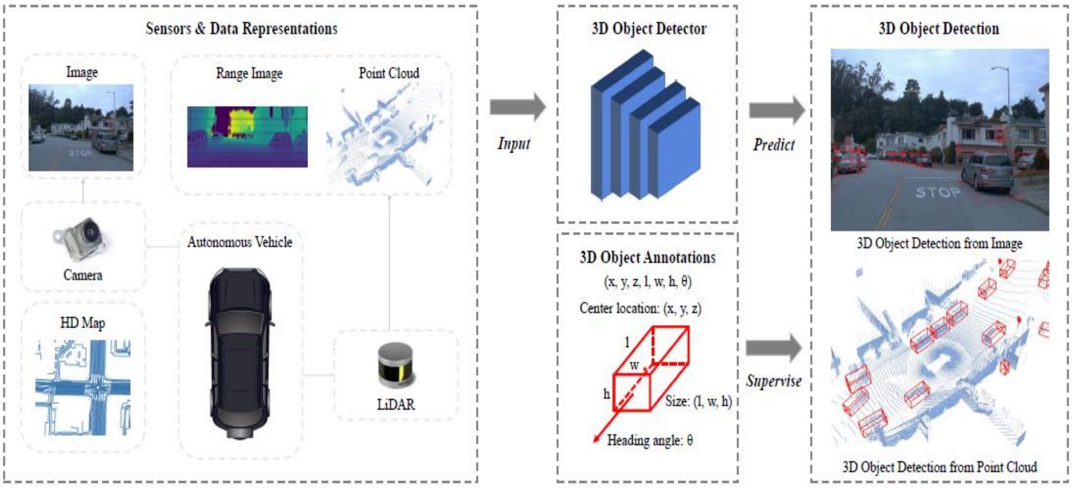

Fig. 5. Visual representation of 3D object detection in autonomous driving techniques [41].

- Point-voxel-based approaches (hybrid approaches): Point-voxel-based techniques employ an integrated architecture that utilizes both points and voxels to facilitate the detection of 3D object [36 ,38].

10. Datasets and Evaluation Criteria

10.1.Datasets

Datasets have been essential in the history of object detection research, serving as a means to measure and compare the performance of different algorithms. Additionally, they have driven the field towards more intricate and demanding tasks. Deep learning approaches, specifically, have recently achieved significant success in many image recognition challenges, with the presence of extensive annotated data being crucial to their achievements. Access to large numbers of images on the Internet enables the creation of comprehensive datasets that are able to capture immense variety and abundance of objects, thereby enabling object detection to achieve unprecedented levels of performance [17]. Over the past two decades, some well-known datasets have been made available for object detection, such as Caltech (Caltech-101, Caltech-256) and PASCAL VOC (e.g., VOC2007, VOC2012), ImageNet Large Scale Visual Recognition (e.g., ILSVRC2014), MS-COCO, and Open ImagesV5, as shown in Fig. 6.

- **Caltech dataset**

Caltech-101 dataset was proposed by Fei et al. [42]. The dataset comprises around 9144 images that are divided into 101 classes of objects, along with an extra class for background clutter. The number of photographs in classes might range from 40 to 800 images [39,40]. The Caltech-101 dataset was created by a methodical procedure that included choosing particular object categories, obtaining appropriate image samples from Google Images, and then manually reviewing and removing any images that did not match the intended category [45].

Caltech-256 dataset [45] represents an advancement over the CalTech-101 dataset in several respects. Notably, the total number of images has been expanded from 9,144 to 30,607, while the lowest number of images per class has risen from 31 to 80. Furthermore, the CalTech-256 dataset encompasses more than twice the number of classes found in the CalTech-101 dataset, among other enhancements [44].

- **PASCAL VOC**

The Pascal Visual Object Classes (VOC) dataset [46] serves as the primary standard benchmark for object detection and has been extensively utilized in this field. The PASCAL VOC dataset consists of 20 distinct types of things [9]. This dataset exists in two different versions:

PASCAL VOC 2007 dataset comprises a total of 9,963 images. 5,001 images are allocated for training and validation purposes, while the remaining 4,952 images are designated for testing. Each image has been tagged with the corresponding class label and accurate bounding boxes [8,10].

PASCAL VOC 2012 dataset is an expanded iteration of the VOC 2007 dataset, encompassing a comprehensive collection of 22,531 images. The train-validation dataset consists of 11,540 images, while the test dataset comprises 10,991 images. It is important to note that the test dataset does not include any publicly available ground-truth box boundaries [8,10].

- **ImageNet**

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [47] was a yearly competition from 2010 to 2017. It emerged as a widely recognized standard for assessing the effectiveness of algorithms. The size of the dataset was increased to about one million images, encompassing 1000 classes for object categorization. A total of 200 classes were carefully selected for the purpose of object detection. These classes encompass a collection of over 500,000 images [2,12]. The ImageNet is utilized for pre-training the backbone and object detection models. It is also usedto train detectors such as R-CNN, OverFeat, and SPPnet [9].

- **MS COCO**

The Microsoft Common Objects in Context (MS-COCO) dataset [48] is widely recognized as highly demanding and complex. There are 91 often encountered things in their natural environments that a typical 4-year-old human may readily identify. The product was introduced in 2014, and its popularity has grown. The dataset contains almost two million occurrences, averaging 3.5 categories per image. In addition, it is worth noting that the dataset in question exhibits an average of 7.7 instances per image, which surpasses the quantity observed in other widely used datasets. The MS COCO dataset includes photos captured from several perspectives [8, 10].

- **Open Images**

The Open images dataset [49] provided by Google comprises a collection of 9.2 million photographs. These images have been annotated with various types of information, including image-level labels, segmentation masks and object-bounding boxes. The product was introduced to the market in the year 2017 and has subsequently undergone six upgrades. Open Images contains a vast collection of 16 million bounding boxes, encompassing 600 categories, throughout a dataset of 1.9 million photos. This extensive dataset establishes Open Images as the most extensive object-local resource. The designers of the dataset showed meticulous attention in selecting captivating, intricate, and varied photos, resulting in an average of 8.3 item categories per image [2, 12].

Table 5. Popular datasets for object detection [12,16,41,40]

Dataset	Classes	Total images	Image size	Size
Caltech-101	101	9,144	300 × 200	131MB
Caltech-256	256	30,607	300 × 200	1.2GB
PASCAL VOC07	20	9963	375 × 500	2GB
PASCAL VOC12	20	11,540	470 × 380	2GB
ImageNet	1000	14 million+	500 × 400	150GB
MS COCO	91	328,000	640 × 480	40GB
Open Images	6000	9 million+	Varied	1.5GB

Fig. 6. Several examples of images from:
(a) PASCAL VOC, (b) ILSVRC,(c) MS COCO, (d) Open Images and with object annotations [17].

10.2. Evaluation Criteria

Evaluation metrics are an excellent technique to determine how good the object detection algorithm is, the most popular metrics are: Precision, Recall, and F1 score. In recent years, Average Precision (AP) metric has gained significant popularity due to its derivation from precision and recall [17]. Before delving into the forms of metrics, let's go over some concepts they all have in common. The following are the most fundamental:

- True Positive (TP): A correct detection of an existing object.
- False Positive (FP): An incorrect detection of a non-existent object or an inaccurate detection of an actual thing;
- True Negative (TN): A correct detection of non-existent object.
- False Negative (FN): An incorrect detection of an existing object.

In this paper, we will briefly describe some of evaluation metrics:

Precision

Precision [1] is the predicted region's percentage corresponding to the true region. The precision formula is given below:

$$\frac{Predicted\ area\ in\ region\ truth}{Total\ area\ of\ predicted\ region} = \frac{TP}{TP+FP} \quad (1)$$

Recall (or Sensitivity)

Recall (Sensitivity) [1, 46] Is the proportion of the ground-truth region that is present in the anticipated region. This is the formula for the recall:

$$\frac{Ground\ truth\ area\ in\ predicted\ region}{Total\ area\ of\ ground\ truth\ region} = \frac{TP}{TP+FN} \quad (2)$$

F-Measure (F1-score)

The F-measure [1] is calculated by averaging the recall and precision scores.
Below is a mathematical explanation:

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Average Precision (AP)

Average Precision [17] Is the most popular metric which resulting from precision and recall, AP is typically evaluated in a category-specific manner, that is, separately for each object category. The following formula is used to determine average precision:

$$\text{AP} = \sum_n (R_n - R_{n-1}) P_n \quad (4)$$

Mean Average Precision (mAP)

Mean Average Precision [17] averaged across all object categories is used as the final performance measure when comparing performance across all object categories.

$$\text{mAp} = \frac{1}{N} \sum_i^N AP_i \quad (5)$$

Accuracy

One of the critical and standard measures used to evaluate performance. It is defined as the ratio between correct samples to the number of total samples [51].

$$\text{AC} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Specificity

$$\text{Specificity} = \frac{TP}{FP+TN} \quad (7)$$

11. Conclusion

Object detection is one of the essential branches within computer vision which has garnered the interest of numerous researchers, particularly since the advent of deep learning tools, which have significantly contributed to the field's advancement. This article demonstrates traditional and deep learning techniques for detecting objects and distinguishing between them. In addition, it provided a summary of all previous reviews and surveys conducted over the past four years, object detection in images, drone and UAVs (Unmanned Aerial Vehicles) captured images, videos, 3D object detection, evaluation metrics, and standard datasets.

12. References

[1] M. Ahmed, K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, "Survey and performance analysis of deep learning based object detection in challenging environments," Sensors, vol. 21, no. 15. 2021. Doi: https://doi.org/10.3390/s21155116.

[2] L. Jiao et al., "A Survey of Deep Learning-Based Object Detection," IEEE Access, vol. 7, pp. 128837–128868, 2019, Doi: https://doi.org/10.1109/ACCESS.2019.2939201.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 580–587, 2014,Doi:https://doi.org/10.1109/CVPR.2014.81.

[4] Y. Xiao et al., “A review of object detection based on deep learning,” Multimedia Tools and Applications, vol. 79, no. 33–34, pp. 23729–23791, 2020, Doi:https://doi.org/10.1007/s11042-020-08976-6.
[5] M. Jiang, S. Huang, J. Duan, and Q. Zhao, “Salicon: Saliency in context,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1072–1080.
[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
[7] W. Liu et al., “SSD: Single shot multibox detector,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016, Doi: https://doi.org/10.1007/978-3-319-46448-0_2.
[8] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, “Object Detection in 20 Years: A Survey,” Proceedings of the IEEE, 2023, Doi: https://doi.org/10.1109/JPROC.2023.3238524.
[9] H. Zhang and X. Hong, “Recent progresses on object detection: a brief review,” Multimedia Tools and Applications, vol. 78, no. 19. pp. 27809–27847, 2019. Doi:https://doi.org/10.1007/s11042-019-07898-2.
[10] A. Dhillon and G. K. Verma, “Convolutional neural network: a review of models, methodologies and applications to object detection,” Progress in Artificial Intelligence, vol. 9, no. 2, pp. 85–112, 2020.
[11] X. Wu, D. Sahoo, and S. C. H. Hoi, “Recent advances in deep learning for object detection,” Neurocomputing, vol. 396, pp. 39–64, 2020, Doi:https://doi.org/10.1016/j.neucom.2020.01.085.
[12] 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:https://doi.org/10.1109/ACCESS.2020.3021508.
[13] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, “A survey of modern deep learning based object detection models,” Digital Signal Processing: A Review Journal, vol. 126. 2022. Doi: https://doi.org/10.1016/j.dsp.2022.103514.
[14] 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:https://doi.org/10.32604/CSSE.2021.017016.
[15] Z. Li et al., “Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey,” Remote Sensing, vol. 14, no. 10, pp. 1–41, 2022, Doi:https://doi.org/10.3390/rs14102385.
[16] J. Kaur and W. Singh, “Tools, techniques, datasets and application areas for object detection in an image: a review,” Multimedia Tools and Applications, vol. 81, no. 27. pp. 38297–38351, 2022. Doi:https://doi.org/10.1007/s11042-022-13153-y.
[17] 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:https://doi.org/10.1007/s11263-019-012474.
[18] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, “A survey and performance evaluation of deep learning methods for small object detection,” Expert Systems with Applications, vol. 172, no. April 2020, p. 114602, 2021, Doi:https://doi.org/10.1016/j.eswa.2021.114602.
[19] K. Tong, Y. Wu, and F. Zhou, “Recent advances in small object detection based on deep learning: A review,” Image and Vision Computing, vol. 97, p. 103910, 2020, Doi:https://doi.org/10.1016/j.imavis.2020.103910.
[20] M. Abdulla and A. Marhoon, “Agriculture based on Internet of Things and Deep Learning,” Iraqi Journal for Electrical and Electronic Engineering, vol. 18, no. 2, pp. 1–8, 2022, Doi: https://doi.org/10.37917/ijeee.18.2.1.
[21] R. S. Khudeyer and N. M. Al-moosawi, “Fake Image Detection Using Deep Learning EfficientNet-V2 network,” vol. 47, pp. 115–120, 2023.
[22] N. Odey and A. Marhoon, “Feature Deep Learning Extraction Approach for Object Detection in Self-Driving Cars,” 2023.
[23] G. Jin, R. I. Taniguchi, and F. Qu, “Auxiliary Detection Head for One-Stage Object Detection,” IEEE Access, vol. 8. pp. 85740–85749, 2020.Doi:https://doi.org/10.1109/ACCESS.2020.2992532.

[24] K. Zhao and X. Ren, "Small Aircraft Detection in Remote Sensing Images Based on YOLOv3," IOP Conference Series: Materials Science and Engineering, vol. 533, no. 1, 2019, Doi: https://doi.org/10.1088/1757-899X/533/1/012056.
[25] R. Girshick, "Fast R-CNN," Proceedings of the IEEE International Conference on Computer Vision, vol. 2015 Inter, pp. 1440–1448, 2015,Doi:https://doi.org/10.1109/ICCV.2015.169.
[26] Y. Wang et al., "Remote sensing image super-resolution and object detection: Benchmark and state of the art," Expert Systems with Applications, vol. 197. 2022.Doi:https://doi.org/10.1016/j.eswa.2022.116793.
[27] P. Singh, M. Diwakar, A. Shankar, R. Shree, and M. Kumar, "A Review on SAR Image and its Despeckling," Archives of Computational Methods in Engineering, vol. 28, no. 7, pp. 4633–4653, 2021, Doi: https://doi.org/10.1007/s11831-021-09548-z.
[28] L. Tang, W. Tang, X. Qu, Y. Han, W. Wang, and B. Zhao, "A Scale-Aware Pyramid Network for Multi-Scale Object Detection in SAR Images," Remote Sensing, vol. 14, no. 4, pp. 1–24, 2022, Doi:https://doi.org/10.3390/rs14040973.
[29] Y. Zhao, L. Zhao, Z. Liu, D. Hu, G. Kuang, and L. Liu, "Attentional Feature Refinement and Alignment Network for Aircraft Detection in SAR Imagery," Jan. 2022,Doi: https://doi.org/10.1109/TGRS.2021.3139994.
[30] H. K. Jung and G. S. Choi, "Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions," Applied Sciences (Switzerland), vol. 12, no. 14, 2022, Doi:https://doi.org/10.3390/app12147255.
[31] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," Proceedings of the IEEE International Conference on Computer Vision, vol. 2021-Octob, pp. 2778–2788, 2021, Doi: https://doi.org/10.1109/ICCVW54120.2021.00312
[32] R. Walambe, A. Marathe, and K. Kotecha, "Multiscale object detection from drone imagery using ensemble transfer learning," Drones, vol. 5, no. 3, p. 66, 2021,Doi:https://doi.org/10.3390/drones5030066.
[33] R. Zhang, Z. Miao, Q. Zhang, S. Hao, and S. Wang, "Video Object Detection by Aggregating Features across Adjacent Frames," Journal of Physics: Conference Series, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012039.
[34] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," Applied Sciences (Switzerland), vol. 10, no. 21. pp. 1–24, 2020. Doi:https://doi.org/10.3390/app10217834.
[35] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and trackingsystem for video surveillance system," Multimedia Tools and Applications, vol. 80, no. 3, pp. 3981–3996, 2021, Doi:https://doi.org/10.1007/s11042-020-09749-x.
[36] L. Jiao et al., "New generation deep learning for video object detection: A survey," IEEE Transactions on Neural Networks and Learning Systems, 2021.
[37] V. V, C. R. K, and R. A. C., "Real Time Object Detection System with YOLO and CNN Models: A Review," vol. XIV, no. 7, pp. 144–151, 2022.
[38] S. Heo, S. Cho, Y. Kim, and H. Kim, "Real-Time Object Detection System with Multi-Path Neural Networks," Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, RTAS, vol. 2020-April, pp. 174–187, 2020,Doi:https://doi.org/10.1109/RTAS48715.2020.000-8.
[39] T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3D Object Detection and Tracking," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, no. Figure 1, pp. 11779–11788, 2021,Doi:https://doi.org/10.1109/CVPR46437.2021.01161.
[40] X. Pan, Z. Xia, S. Song, L. E. Li, and G. Huang, "3D Object Detection with Pointformer," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7459–7468, 2021,Doi:https://doi.org/10.1109/CVPR46437.2021.00738.
[41] J. Mao, S. Shi, X. Wang, and H. Li, "3D Object Detection for Autonomous Driving: A Comprehensive Survey," International Journal of Computer Vision, no. February, 2023, Doi:https://doi.org/10.1007/s11263-023-01790-1.

[42]L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories,” Computer Vision and Image Understanding, vol. 106, no. 1, pp. 59–70, 2007, Doi:https://doi.org/10.1016/j.cviu.2005.09.012.
[43] A. S. Rao and K. Mahantesh, “Learning Semantic Features for Classifying Very Large Image Datasets Using Convolution Neural Network,” SN Computer Science, vol. 2, no. 3, pp. 1–9, 2021, Doi:https://doi.org/10.1007/s42979-021-00589-6.
[44] S. H. S. Basha, S. K. Vinakota, V. Pulabaigari, S. Mukherjee, and S. R. Dubey, “AutoTune: Automatically Tuning Convolutional Neural Networks for Improved Transfer Learning,” Neural Networks, vol. 133, pp. 112–122, 2021,Doi:https://doi.org/10.1016/j.neunet.2020.10.009.
[45] G. Griffin, a Holub, and P. Perona, “Caltech-256 object category dataset,” Caltech mimeo, vol. 11, no. 1, p. 20, 2007.
[46] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge,” International Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010,Doi:https://doi.org/10.1007/s11263-009-0275-4.
[47] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015,Doi:https://doi.org/10.1007/s11263-015-0816-y.
[48] T. Y. Lin et al., “Microsoft COCO: Common objects in context,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014,Doi:https://doi.org/10.1007/978-3-319-10602-1_48.
[49] A. Kuznetsova et al., “The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale,” International Journal of Computer Vision, vol. 128, no. 7, pp. 1956–1981, 2020, Doi:https://doi.org/10.1007/s11263-020-01316-z.
[50] Z. A. Khalaf, S. S. Hammadi, A. K. Mousa, H. M. Ali, H. R. Alnajar, and R. H. Mohsin, “Coronavirus disease 2019 detection using deep features learning.,” International Journal of Electrical & Computer Engineering (2088-8708), vol. 12, no. 4, 2022.
[51] G. S. Ohannesian and E. J. Harfash, “Epileptic Seizures Detection from EEG Recordings Based on a Hybrid System of Gaussian Mixture Model and Random Forest Classifier,” Informatica (Slovenia), vol. 46, no. 6, pp. 501–505, 2022,Doi:https://doi.org/10.31449/inf.v46i6.4203.

Journal of Basrah Researches (Sciences) **50**(1), 46 (2024)
DOI: https://doi.org/10.56714/bjrs.50.1.5

A survey study in Object Detection: A Comprehensive Analysis of Traditional and State-of-the-Art Approaches

مروة عبد المجيد حميد[1]، زينب علي خلف*[2]

[1]قسم علوم الحاسوب، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة البصرة، البصرة، العراق.
[2]قسم علم الحاسوب، كلية التربية للعلوم الصرفة، جامعة البصرة، البصرة، العراق.

معلومات البحث

الاستلام	11 كانون الأول 2023
القبول	4 شباط 2024
النشر	30 حزيران 2024

الكلمات المفتاحية

كشف الكائنات ، التعلم العميق ، الكاشفات التقليدية ، كشف الكائنات في الصور .

Citation: M.A. Hameed, Z.A. Khalaf, J. Basrah Res. (Sci.)**50**(1), 46 (2024). DOI:https://doi.org/10.56714/bjrs.50.1.5

الملخص

يعد اكتشاف الكائنات مجالًا أساسيًا في رؤية الكمبيوتر، مع التركيز على تحديد وجود الكائنات وفئتها ضمن بيانات الصورة أو الفيديو. وتكتسب أهمية هذا الموضوع أهمية قصوى في العديد من المجالات التي تؤثر بشكل مباشر على حياة الناس، بما في ذلك القيادة الذاتية، وأنظمة الرعاية الصحية، والمراقبة الأمنية. وعلى النقيض من المنهجيات التقليدية المستخدمة للكشف عن الكائنات، أظهرت الخوارزميات القائمة على التعلم العميق تقدمًا كبيرًا في الكفاءة والدقة الحسابية في السنوات الأخيرة. تهدف هذه الدراسة إلى تقديم مراجعة شاملة لاكتشاف الكائنات من خلال الاستخدام المنهجي للتعلم العميق لتسهيل الفهم الشامل والمتعمق للمبادئ الأساسية في هذا المجال. وقد تناولت المناقشة مواضيع مختلفة، مثل العوائق والتعقيدات المرتبطة باكتشاف الكائنات وأجهزة الكشف التقليدية والتعلم العميق. اكتشاف الكائنات داخل الصور ومقاطع الفيديو، والكشف عن الكائنات في الوقت الفعلي، والكشف عن الكائنات ثلاثية الأبعاد، ومجموعات البيانات شائعة الاستخدام، والمقاييس المستخدمة لتقييم أداء اكتشاف الكائنات. من المرجح أن تسفر هذه الدراسة عن فوائد علمية للأكاديميين العاملين في مجال اكتشاف الأشياء والتعلم العميق.

*Corresponding author email : zainab.khalaf@uobasrah.edu.iq

ISSN: 1817-2695 (Print); 2411-524X (Online)
Online at: https://jou.jobrs.edu.iq