e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

A REVIEW OF OBJECT DETECTION ALGORITHMS BASED ON DEEP LEARNING

Sakshi Yadav*1, Shayesta Sana*2, Saurabh Singh*3, Dr. Anil K. Pandey*4

*1,2,3Student Computer Science And Engineering, Babubanarasi Das Institute Of Technology And Management, Lucknow, India.

*4Assistant Professor Computer Science And Engineering, Babubanarasi Das Institute Of Technology And Management, Lucknow, India.

DOI : https://www.doi.org/10.56726/IRJMETS39161

ABSTRACT

With the rise of facial detection, autonomous vehicles, smart video surveillance, and various people-counting applications, the demand for fast and accurate object detection systems is also rising day by day. Object detection or target detection has become an important research hotspot in past years and is being widely used in various applications. The main aim is to quickly identify and locate a large number of objects in an image or a video. This article presents a review of object detection algorithms. Firstly, computer vision and object detection are summarized and introduced. Then, some commonly used deep learning methods, R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, SSD and YOLO are selected for analysis and introduction. At the end of the article, various applications and challenges of object detection algorithms are also discussed.

I. INTRODUCTION

Computer vision is the field of computer science that focuses on making computers capable of understanding, interpreting and processing digital images, with the aim of creating systems that can recognize objects, identify shapes, and detect motion. It involves the use of computer algorithms to analyze digital images and extract high-level information from them. Computer vision systems are used in a wide range of applications, including machine vision, medical imaging, robotics, and security.

Gesture recognition is the ability to detect and interpret the movements of a person's body and hands. This technology is often used in augmented reality to allow users to interact with the virtual environment. Gesture recognition can be used for navigation, object manipulation, and control of virtual objects.

Object detection is a computer vision technique used to identify objects within an image or video. It is used to detect objects such as cars, people, animals, buildings, and other items. Object detection algorithms typically use bounding boxes to identify objects in an image or video by drawing a box around the object. Fig. 1 below shows Object Detection. Object detection can be used for a variety of applications such as autonomous vehicles, security systems, medical diagnostics, and robotics.

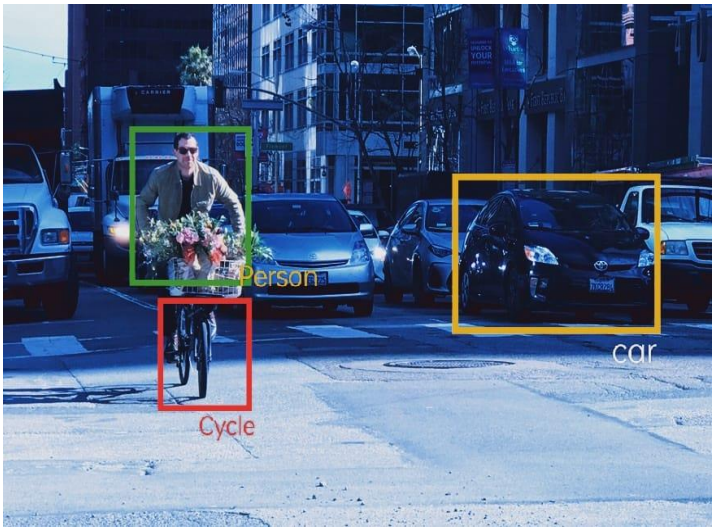

Fig. 1 Object Detection

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

Object detection is the computer vision technique for detecting and locating objects in an image or video. This can involve recognizing and locating a particular object, such as a face, or finding all instances of objects of a certain type, such as cars. Object detection can be used for a variety of applications, from security and surveillance to medical imaging to self-driving cars. It is an important part of many computer vision applications, from recognizing faces in photos to tracking objects in videos.

Object detection is a common task in computer vision, which has become increasingly popular with the advancement of deep learning. Deep learning-based object detection algorithms have achieved significant performance improvements on tasks such as image detection, object tracking, and video surveillance. In this review, we discuss various deep learning-based object detection algorithms, such as R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, SSD and YOLO. We also discuss the advantages and disadvantages of each algorithm. Finally, we examine the challenges and future directions of deep learning-based object detection.

II. SOME OBJECT DETECTION ALGORITHMS

1. R-CNN

Region-based Convolutional Neural Networks (R-CNN) are a type of deep learning algorithm used for object detection tasks. It works by first extracting regions or proposals from an image, then running a Convolutional Neural Network (CNN) on each of those regions to determine whether a certain object is present or not. The output from the CNN is then used to classify the region as containing an object or not. This makes R-CNNs a powerful tool for object detection, as it is able to detect objects in images that would otherwise be difficult for traditional image processing algorithms to detect.

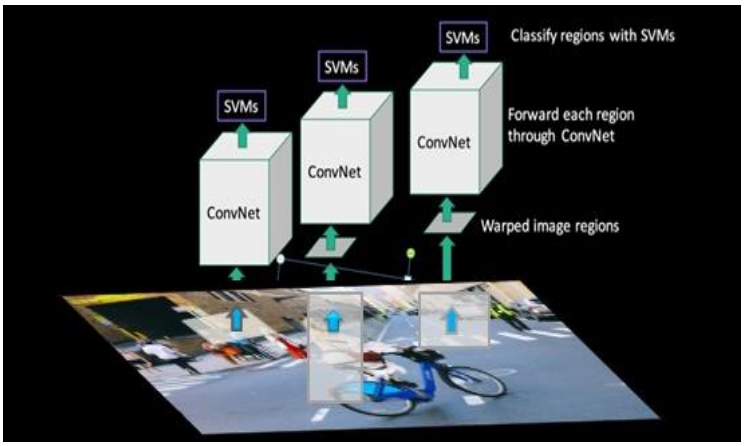

Fig. 2 R-CNN.

Region-based Convolutional Neural Networks (R-CNNs) are a type of convolutional neural network (CNN) that uses a region proposal algorithm to identify and classify regions of interest within an image. R-CNNs have been shown to be effective in object detection and recognition tasks. They are commonly used in applications such as autonomous driving, facial recognition, and medical imaging. R-CNNs can be trained using a variety of techniques, including supervised learning and reinforcement learning. The main difference between a standard CNN and an R-CNN is that the R-CNN includes an additional step of generating region proposals before the convolutional layers. These region proposals are then used to generate features that are used to classify the image. Fig. 2 above shows the working of R-CNN.

Region-based convolutional neural networks (R-CNNs) are a type of deep learning architecture used for object detection and semantic segmentation tasks. R-CNNs employ a region proposal algorithm, such as Selective Search or EdgeBoxes, to generate a set of candidate regions or bounding boxes. These regions are then used as input to a convolutional neural network (CNN) to classify the contents of each region, and to regress the bounding box parameters. R-CNNs are a popular choice for object detection and semantic segmentation tasks because they are able to accurately identify objects in complex scenes, while also providing rich semantic information about the objects in the scene. Additionally, R-CNNs are generally faster than their counterparts, such as the single-shot detection model (SSD) and the You Only Look Once (YOLO) detector. Despite their speed and accuracy, R-CNNs are not without their drawbacks. For example, they require a lot of manual effort to

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

generate the region proposals, and they are computationally expensive due to their two-step approach. Nevertheless, R-CNNs remain a popular choice for object detection and semantic segmentation tasks due to their accuracy and flexibility.

Advantages of Region-based Convolution Neural Networks:

(1) Region-based Convolutional Neural Networks (R-CNNs) are significantly more accurate than traditional object detectors. This is because R-CNNs use region proposals to detect objects in an image, which provides a more robust approach than sliding windows.
(2) R-CNNs are able to detect objects in different sizes and shapes. By using region proposals, R-CNNs are able to detect objects of different sizes that may not be captured by traditional detectors.
(3) R-CNNs are faster than the traditional sliding window approach. This is because the region proposals are used to identify potential objects in the image, reducing the number of windows that need to be evaluated.
(4) R-CNNs are able to classify objects in an image. By using region proposals and a deep neural network, R-CNNs are able to classify objects in an image and provide a confidence score for each identified object.

Disadvantages of Region-based Convolution Neural Networks:

(1) R-CNNs are computationally expensive. This is because the region proposals require a significant amount of computation to generate, and the deep neural network requires additional computing power to classify the objects.
(2) R-CNNs are limited to the region proposals. While the region proposals are effective at detecting objects, they may miss some objects that are not included in the region proposals.
(3) R-CNNs are not as efficient as single-shot detectors. Single-shot detectors are able to detect objects in an image in a single pass, while R-CNNs require multiple passes to detect objects. This makes R-CNNs less efficient than single-shot detectors.

2. FAST R-CNN

Fast R-CNN is a deep learning algorithm for object detection developed by Ross Girshick in 2015. It is an improvement over the original R-CNN algorithm which used Selective Search for generating region proposals.

Fast R-CNN uses a single convolutional neural network to perform both region proposal and object detection. The network is first trained to generate region proposals and then to predict the class label and bounding box coordinates for each proposal. This two-stage process is much faster than the previous R-CNN algorithm which required two separate networks to handle the two tasks.

The Fast R-CNN algorithm uses a region of interest pooling layer to efficiently handle multiple regions proposals. The region of interest pooling layer divides the input image into regions and then applies a max pooling operation to each region. This allows the network to extract features from each region more quickly.

The Fast R-CNN algorithm has been used successfully in many applications such as object detection, semantic segmentation, and image captioning.Fig.3 (a) and (b) below shows the working of Fast R-CNN.

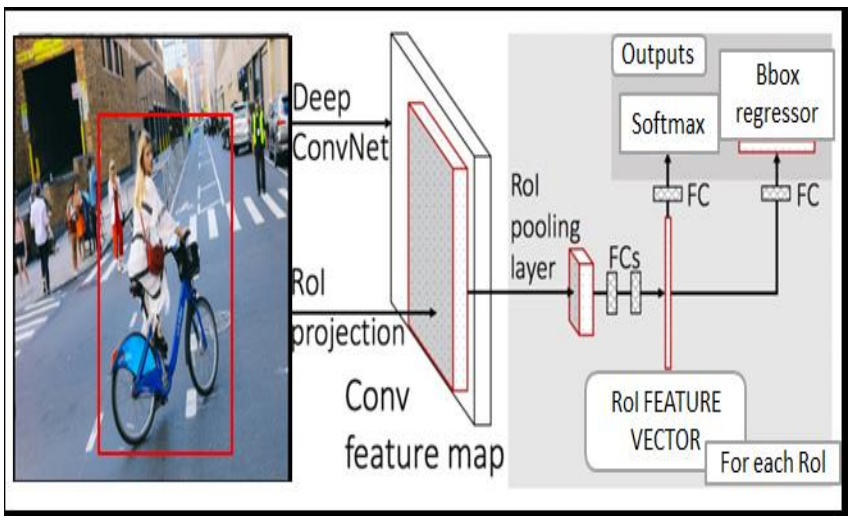

Fig. 3 (a) Process of Fast R-CNN

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

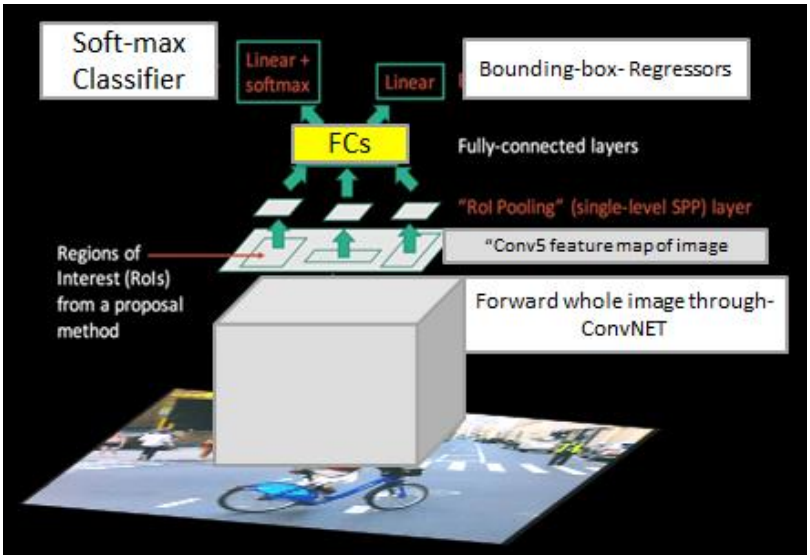

Fig. 3 (b) Working of Fast R-CNN

Advantages of Fast Region-based Convolutional Neural Networks:

(1) Fast region based convolutional neural networks are able to detect and classify objects in images very quickly and accurately.
(2) They are able to recognize faces and objects in real-time, making them ideal for applications that require fast processing.
(3) They are able to detect objects of different sizes and shapes in an image, making them suitable for a wide variety of applications.
(4) They have the capability to learn from data, which means they can be used for applications that require the learning of new patterns or rules.

Disadvantages of Fast Region-based Convolutional Neural Networks:

(1) Fast region based convolutional neural networks are limited in their ability to recognize complex patterns.
(2) They are also limited in their ability to learn from large datasets.
(3) They require a lot of computational resources, which can be a limitation in certain applications.
(4) They are also prone to overfitting, which can lead to inaccurate results.

3. FASTER R-CNN

Faster R-CNN is an object detection algorithm that uses a convolutional neural network (CNN). It combines region proposal, classification and bounding box regression and runs in an end-to-end manner. Faster R-CNN is an improved version of the original R-CNN which has been used for object detection for a few years.

Faster R-CNN uses a region proposal network (RPN) which is a fully convolutional network to generate region proposals. The region proposals are then used to classify the object and to get the bounding box coordinates of the object. This is done by a classifier and a regression layer.

The RPN is trained along with the classifier and the regression layer. The whole process is done in an end-to-end manner. Faster R-CNN is faster and more accurate than the original R-CNN and has been used in various applications such as face detection, pedestrian detection, etc.

Faster R-CNN is a deep learning algorithm used for object detection in images. It uses a region proposal network (RPN) to generate region proposals, which are then classified and localized by a fast convolutional neural network (CNN). The main goal of Faster R-CNN is to improve the speed and accuracy of object detection. The algorithm is composed of two main stages: a region proposal network (RPN) and a Fast R-CNN. The RPN is used to generate region proposals, which are then sent to the Fast R-CNN for classification and localization. The Fast R-CNN then produces the final output by combining the region proposals with the class scores and bounding box locations

Advantages of Faster Region-based Convolutional Neural Networks:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

(1) Faster Region Based Convolutional Neural Networks (Faster R-CNNs) are able to detect objects in images and videos more efficiently and with higher accuracy.
(2) Faster R-CNNs are able to detect multiple objects in a single image, making them ideal for object detection tasks.
(3) Faster R-CNNs are also able to learn from their mistakes, allowing them to become more accurate over time.
(4) Faster R-CNNs are able to process images and videos faster than traditional convolutional neural networks, making them ideal for applications where speed is important.

Disadvantages of Faster Region-based Convolutional Neural Networks:

(1) Faster R-CNNs require a lot of computing power, making them difficult to implement on small computers or devices.
(2) Faster R-CNNs can be difficult to train, as they require large amounts of data and a lot of computational resources.
(3) Faster R-CNNs are prone to overfitting, meaning that they can become too specialized and fail to generalize to new images or videos.
(4) Faster R-CNNs are also more prone to false positives, meaning they can mistakenly detect objects that are not actually there.

4. R-FCN

Region based Fully Convolutional Network (R-FCN) is a convolutional neural network (CNN) algorithm used for object detection. It is an advanced version of the popular Faster R-CNN model and is often used for applications requiring rapid detection of multiple objects in an image. R-FCN works by using a region proposal network (RPN) to generate region proposals, which are then classified and refined using a fully convolutional network. This model is also designed to be much faster than other object detection models due to its efficient use of convolutional layers. Furthermore, R-FCN is designed to be more accurate than other object detectors by using a multi-scale approach to handle objects of different sizes.

Advantages of Region based Fully Convolutional Network:

(1) Region based Fully Convolutional Networks (R-FCN) can process images of any size, due to their fully convolutional nature.
(2) R-FCN has fewer parameters than other deep learning architectures, making it more efficient and faster to train.
(3) R-FCN can detect objects at multiple scales, due to its ability to share features across multiple regions.
(4) It is also capable of detecting objects with varying aspect ratios.
(5) R-FCN is a more efficient network, as it eliminates the need for a region proposal network (RPN) and a fully connected layer.

Disadvantages of Region based Fully Convolutional Network:

(1) Training for a region based fully convolutional network algorithm can be computationally expensive.
(2) The algorithm can suffer from overfitting if the training data is insufficient.
(3) The accuracy of the algorithm is heavily dependent on the quality of the training data.
(4) The algorithm can be difficult to interpret and explain due to its complex nature.
(5) The algorithm can be sensitive to small changes in the input data.

5. SSD

Single Shot Multi-Box Detector (SSD) is an object detection algorithm based on deep learning. It is a single shot detector, meaning that it is able to detect objects in a single forward pass of the network. It uses a set of default boxes to detect a variety of object sizes and aspect ratios in an image. It is faster than other methods such as R-CNN and Fast R-CNN, and more accurate than YOLO. SSD is suitable for real-time object detection tasks and can be used in applications that require fast detection.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

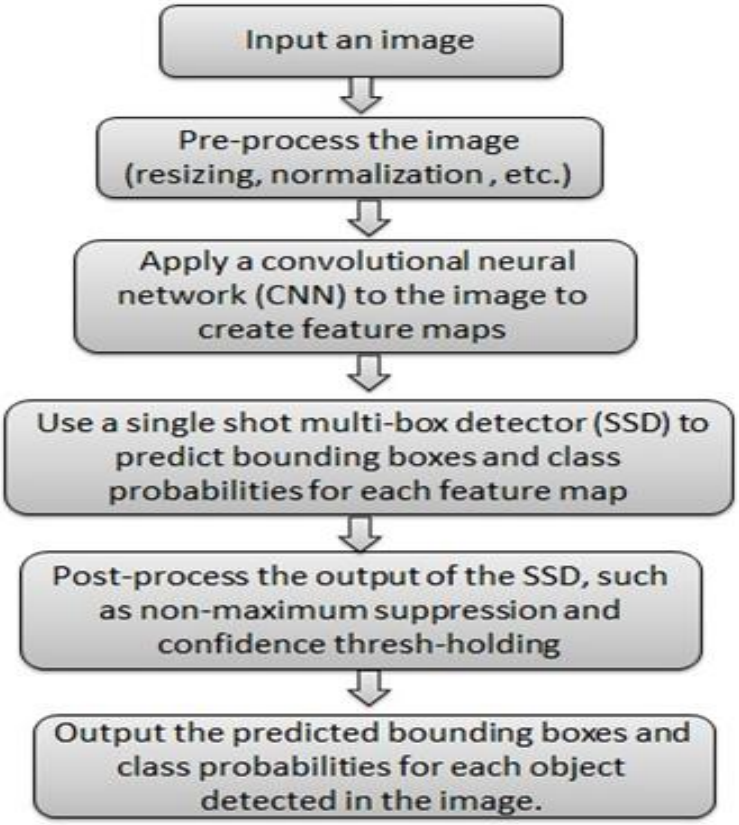

Fig.4. Flowchart of SSD Algorithm.

Advantages of single shot multi box detector algorithm

(1) High accuracy: Single shot multi box detectors are known for their high accuracy when it comes to object detection. This is because the model is able to recognize more than one object in an image and also detect objects in different sizes and shapes.

(2) Faster detection time: Single shot multi box detectors are able to process an image in less time than other algorithms. This is because the model is designed to be more efficient and can detect the object faster.

(3) Robustness: Single shot multi box detectors are able to handle different types of input images and can detect objects in different scales, making the model more robust.

(4) Less prone to overfitting: Single shot multi box detectors are less prone to overfitting since the model is designed to recognize multiple objects in an image. This means that the model is able to generalize better and is less likely to over fit on a particular dataset.

Disadvantages of single shot multi box detector algorithm

(1) It is computationally expensive, as it requires deep neural networks to run multiple times.

(2) It has limited accuracy, as it is based on the assumption that objects in an image have a fixed size and shape.

(3) It is unable to detect objects of varying size and shape.

(4) It is unable to detect objects that are partially occluded.

(5) It is unable to detect objects in an image that have multiple categories.

6. YOLO

You Only Look Once (YOLO) is an object detection system developed by Joseph Redmon and Ali Farhadi. It is a single shot multi-box detector, meaning it detects objects without needing to scan multiple times and can detect multiple objects in one shot. YOLO divides an image into a grid of cells and then applies a convolutional neural network (CNN) to each cell to make predictions. Fig. 8 below shows the flowchart of YOLO.The algorithm uses multiple bounding boxes per cell and predicts the probability of each bounding box containing an object as well as the class of the object. It is a fast and accurate algorithm, able to detect objects in real-time with a high accuracy rate.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

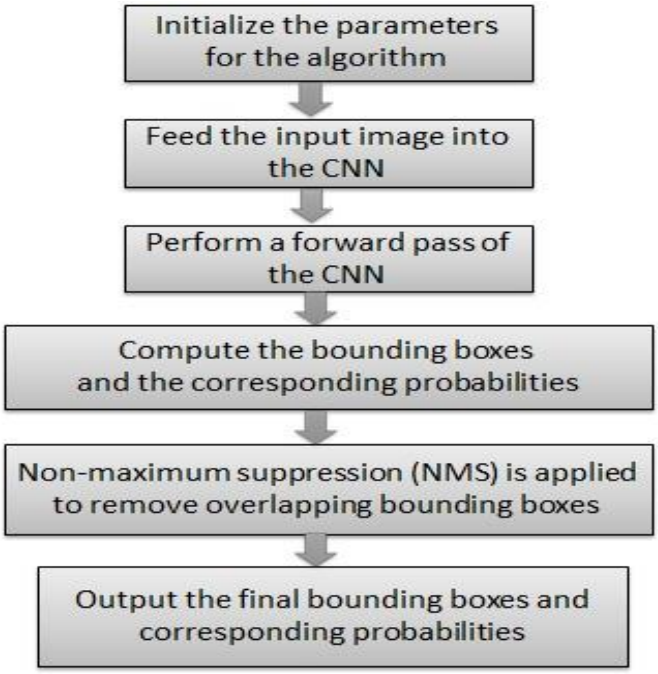

Fig. 5 Flowchart of YOLO Algorithm

Advantages of You Only Look Once:

(1) You Only Look Once (YOLO) is an extremely fast real-time object detection algorithm. It can detect objects in an image or video frame with a single pass, making it a very efficient algorithm.
(2) YOLO is an excellent choice for applications that require real-time object detection, such as robotics, autonomous vehicles, and security systems.
(3) YOLO is also very accurate, as it is able to detect objects in an image with high accuracy.
(4) YOLO is also very flexible and can be trained on different datasets to detect different types of objects.

Disadvantages of You Only Look Once:

(1) YOLO is not very precise and can only detect objects within a certain size range. It is not able to detect very small objects and may miss out on some details, such as facial features.
(2) YOLO is not the best choice for applications that require very precise object detection, such as facial recognition or medical imaging.
(3) YOLO can be computationally expensive and may not be suitable for applications that require low latency.
(4) YOLO is not well suited for detecting objects that are far away or in low-light conditions.

III. APPLICATIONS

Autonomous vehicles: Object detection algorithms are used in autonomous vehicles to identify and track objects in real-time. This helps the vehicles take decisions on the road and avoid collisions.

(1) Facial recognition: Object detection algorithms are used in facial recognition systems to identify and track faces in images and videos. This helps the systems to accurately recognize a person from a database of known individuals.
(2) Robotics: Object detection algorithms are used in robotics to recognize and track objects in the environment. This helps the robots make decisions based on their surroundings and perform tasks accordingly.
(3) Security systems: Object detection algorithms are used in security systems to identify and track objects in real-time. This helps the systems to detect intruders or suspicious activities and alert the authorities in time.

IV. CHALLENGES

(1) Occlusion: Objects can be partially or completely blocked by other objects, making it difficult for the algorithm to detect them.
(2) Scale Variations: Objects can appear in images of different sizes, making it difficult for the algorithm to detect them.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

(3) Lighting Variations: Different lighting conditions can affect the appearance of objects, making it difficult for the algorithm to detect them.
(4) Clutter: Objects can be surrounded by other objects in a cluttered scene, making it difficult for the algorithm to distinguish them.
(5) Background: The background of an image can be very complex, making it difficult for the algorithm to distinguish the object from the background.

V. CONCLUSION

In conclusion, deep learning-based object detection algorithms have shown promising results in a variety of applications. They have the potential to enable real-time detection of multiple objects with high accuracy. However, further research is needed to improve their accuracy and make them better suited for real-world scenarios. Additionally, the development of robust datasets and the implementation of new techniques such as attention mechanisms and Generative Adversarial Networks should be explored in order to further improve the performance of object detection algorithms.

Overall, deep learning-based object detection algorithms have demonstrated their ability to accurately detect objects in an image. These algorithms have shown their superiority over traditional detection models and are becoming increasingly popular in the field of computer vision. With the development of more advanced algorithms, deep learning-based object detection will continue to improve and provide better accuracy and faster detection. This review has provided an overview of the current state of deep learning-based object detection and highlighted the potential of this technology.

Object detection algorithms are a powerful tool for computer vision, allowing us to detect and classify objects in images and videos. These algorithms are used extensively in many areas, such as autonomous vehicles, security systems, facial recognition, and robotics. Their accuracy, speed, and robustness have improved significantly over the years, making them increasingly useful and reliable. With the rise of deep learning and machine learning, object detection algorithms have become an increasingly powerful tool for a wide range of applications. From autonomous vehicles to medical imaging, object detection algorithms are powering the future of computer vision. As technology advances, object detection algorithms will continue to improve and become more powerful. With the help of object detection algorithms, we can detect and classify objects more quickly, accurately, and efficiently. As their accuracy and speed continue to improve, they will become even more useful in a variety of applications.

Table 1: Comparison Table of SSD with other algorithms

S. No.	Algorithms	Comparison With SSD
1	R-CNN	Complex to implement and slower than SSD
2	Fast R-CNN	High resource usage and slow speed
3	Faster R-CNN	Memory usage is high and detection time is slow
4	YOLO	Less precise in detecting objects and uses multiple passes of the image
5	RFCN	Not suitable for small objects and Inference is slower

VI. REFERENCES

[1] Sik-Ho Tsang, "R-FCN- Positive - Sensitive Score Maps (Object Detection)", Towards data science, October 2018.
[2] Anil Pandey Solving Component Identification Problem by using Evolutionary Method, International Journal of Engineering and Advanced Technology , Vol 9 ISBN 3768-3772
[3] Manish Chablani, "YOLO – you only look once,real time object detection", Towards data science, August 2017.
[4] Jonathan Hui, "Understanding region based fully convolutional network (RFCN) for object detection",Medium, March 2018.

[5] YugeshVerma , "R-CNN vs Fast-RCNN vs Faster – RCNN –A comparative guide",Developers corner, September 2021.

[6] Ashwani Kumar, Zuopeng Justin Zhang &HongboLyu, "Object detection in real time based on improved single shot multi-box detector algorithm",Springer Open, October 2020.

[7] Wei Liu, DragomirAnguelov, DumitruErhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu & Alexander C. Berg, "SSD: Single Shot Multi-Box Detector",Springer Link ,September 2016.

[8] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng,Shou-tao Xu, and Xindong Wu, "Object Detection with Deep Learning: A Review",IEEE Transaction on neural network and learning system, April 2019.

[9] Jun Deng, Xiaojing Xuan, Weifeng Wang, Zhao Li4, Hanwen Yao, ZhiqiangWang, "A review of research on object detection based on deep learning",Journal of Physics: Conference Series,2020.

[10] Liu, W., Anguelov, D., Erhan, D., et al. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision, 2016, pp. 21-37.

[11] Anil Kumar Pandey , Review for Different Approaches for identify a Software Component, Journal of Emerging Technologies and Innovative Research (JETIR), 2019.

[12] Sita M. Yadav, Sandeep M. Chaware, "Video Object Detection through Traditional and Deep Learning Methods", 2020.