ISSN No:-2456-2165

IJReal-Time Object Detection by using Deep Learning: A Survey

[1]Manish Yewange,[2]Kanishk Gaikwad,[3]Ramkishan Kamble,[4]Sagar Maske,[5]Prof.D.T.Salunke
[1,2,3,4]UGStudent, [5]Assistant Professor
ITdepartment, JSPM's Rajarshi Shahu College of Engineering,
Tathawade, Pune, India

Abstract:- Object detection has received a significant attention from researchers in recent years because of its close relationship with video analysis and image understanding. Traditional object detection approaches are built on the basis of classifiers and shallow trainable architectures. Their performance can easily plateau by developing sophisticated ensembles that incorporate various low-level visual features with high-level context from object detectors and scene classifiers. More powerful tools that can learn semantic, high-level, and deeper features are being offered with the rapid growth of deep learning to address the difficulties that afflict traditional systems. These models react differently in terms of network architecture, training approach, and optimization function. We present a review of deep learning-based object detection frameworks in this paper. Object recognition tasks, on the other hand, are more difficult to assess, consume more energy, and necessitate more computing power than image categorization. To address these issues, a novel technique for real-time object detection application is being developed in order to improve the detection process' accuracy and energy efficiency.

Keywords:- *CNN, F-RCNN, Dataset, Images, Videos.*

I. INTRODUCTION

Deep Learning is a type of algorithm learning purely based on synthetic neural networks, in which more than one layer of processing is utilised to extract regularly higher-level functions from data. The use of CNN for item detection requires human-computer interaction, which is a difficult and interesting problem. It could be used to build smarter, more accurate robots with a capped potential for increased competence in the goods. This paper contains a variety of real-lifestyles packages, such as highway surveillance cameras to prevent speeding, interactive recreation improvement, and driverless autos.Nearly a thousand regions are offered to incorporate an item within the picture, which can be referred to as the proposed areas, in object detection, which includes proximity detection and detecting the sorts of several items identified in one single picture. For item detection responsibilities, a number of device learning and function extraction algorithms have been developed. SVM (Support Vector Machines), SGD (Stochastic Gradient Descent), Convolutional Neural Support Vector Machines (CNSVMs), and the integration of several functions have all been presented as hand-crafted function extraction methodologies for item detection. Convolutional Neural Network (CNN) has been used in item detection tasks as a result of its success in picture classification tasks.In contrast to traditional Computer Vision (CV) architectures and distinct device learning responsibilities, where each function must be described in advance manually, CNN learns to extract functions from a predetermined database of functions on a regular basis. The CNN is combined with neural community classifiers that may be fed ahead of time, allowing the CNN community to be trained across the entire dataset. CNN requires a large quantity of well-informed data to establish an appropriate position. The ability of CNN to work on huge datasets despite limited computer power will increase the cost of CNN and lead to its development.

II. LITERATURE SURVEY

A. Real-Time Objects Recognition Approach for Assisting Blind People.

In computer vision, such as navigation and path finding, blind assistance is posing a significant challenge. To give the essential information about the environment, this research employs two cameras installed on a blind person's glasses, a GPS-free service, and an ultrasonic sensor. A dataset of objects acquired from everyday scenarios is built in order to apply the requisite recognition. Faces, bicycles, seats, doors, and tables are ubiquitous in blind situations, and object identification is a technique for distinguishing objects in the real world from a digital image. The disparity map is created using two cameras, the GPS service is used to group objects depending on their locations, and the sensor is utilised to identify any barrier at a medium to long range. The descriptor of the Speeded-Up Robust Features approach is optimised to conduct the recognition. The proposed technique for the blind intends to provide greater possibilities for those with vision loss to attain their full potential. The proposed work operates well in a real-time environment, according to the experimental results. for the blind is to provide greater possibilities for persons who have lost their vision to attain their full potential. The outcomes of the experiments show how well the suggested work functions in real-world scenarios. [1]

B. Object Detection Combining Recognition and Segmentation.

Liming Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen; Liming Wang1, Jianbo Shi2, Gang Song2, et al created a method for detection of objects that combines top-down recognition and also bottom-up image segmentation. The two essential steps in this strategy are generating hypotheses and verifying them. In the top-down hypothesis generation process, they increase the Shape Context feature, making it more resistant to object deformation and backdrop clutter. With the improved Shape Context, a collection of

ISSN No:-2456-2165

item placement hypotheses and figure ground masks with a high recall and low precision rate are generated. In the verification step, they also compute a set of viable segmentations that are consistent with top-down object assumptions, and suggest a False Positive. [2]

C. Human objects detection, behavior recognition

SeyedYahyaNikouei et al created Human object detection, behavior recognition, and prediction in smart surveillance fall into this category, where a substantial volume of video streaming data can take a long time to transition and impose a lot of strain on communication networks Video processing and object detection are widely accepted as computationally difficult and too costly for resource-constrained edge devices to handle. A lightweight Convolutional Neural Network (L-CNN) inspired by depth wise separable convolution and the Single Shot Multi-Box Detector is introduced in this paper (SSD). By restricting the classifier's searching space to focus on human objects in surveillance video frames, the proposed L-CNN technique detects pedestrians with a low compute workload on an edge device. [3]

D. The movement of SCARA robot

In this work, MohannadFarag et al. Deep learning-based object detection and edge detection-based position measurement guide SCARA's movement in grasp and place tasks, respectively. Using a transfer learning technique, a deep convolutional neural network (CNN) model called KSS net is constructed based on CNN Alex net for object detection. A SCARA training dataset comprising 4000 photos of two object categories associated with 20 distinct positions is constructed and tagged to train the KSS net model. The position of the observed item is included in the prediction result at the output classification layer. With 100% object detection precision, 100% robotic positioning accuracy, and 100% effective real-time robotic grasping in 0.38 seconds, this method achieved state-of-the-art results. [4]

E. FPGA and neural networks technologies to solve the real-time object recognition problem

Edward Rzaev et al stated Object detection is one of the most active research and application areas of neural networks, as discussed in this review. In this research, they employ FPGA and neural networks to address the real-time object recognition challenge. This article discusses the integration of the YOLOv3 neural network on the DE10-Nano FPGA. When compared to more expensive GPU-based solutions, differences in the cost and dimensions of the FPGA board utilised offset slightly inferior indications of the major metrics (Map, FPS, inference time) when operating a neural network on a De10- Nano board. Based on the findings of a study that investigated several methods for converting neural networks to FPGA. [5]

F. Region Proposal Network (RPN) that shares full-image convolutional features

RenShaoqing et al. introduce a Region Proposal Network (RPN),In this paper, the detection network and the full-image convolutional features are shared, allowing for almost cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object limits and objectless scores at each place. From start to finish, the RPN is trained to create high-quality region proposals, which Fast R-CNN employs for detection. They further combine RPN and Fast R-CNN into a single network by sharing their convolutional features, using the lately trendy concept of neural networks with'attention' processes. The RPN component directs the unified network's search. For the model VGG-16, which has a depth. [6]

G. Object Discovery and Grasp Detectionusing CNN

Di Guoet al In robotics, grabbing an object from a stack of objects in real time is still a challenge. This necessitates the robot's capacity to execute both quick item discovery and grasp detection: first, a target object should be selected from the stack, then the right grip configuration should be applied to grab the object. In this paper, we present a shared convolutional neural network (CNN) capable of performing these two tasks in real-time. On a GPU, the model's processing speed is around 100 frames per second, which is more than enough to meet the requirement. We also construct a labelled RGBD dataset with scenes of stacked objects for robotic grasping at the same time .[7]

H. RCNN AndF-RCNN

In this review, Chandan G, et al. Deep learning has had a significant impact on how the world has adapted to Artificial Intelligence in recent years. Some of the most prominent object identification algorithms include Region-based Convolutional Neural Networks (RCNN), Faster-RCNN, Single Shot Detector (SSD), and You Only Look Once (YOLO). Among these, faster-RCNN and SSD have superior accuracy, although YOLO outperforms when speed is favoured over accuracy. Deep learning combines SSD and Mobile Nets to efficiently implement detection and tracking. Without sacrificing performance, this method recognises objects rapidly and efficiently. [8]

I. Object detection, localization and tracking system

Chen, Zhihao, et al presented object detection, localization, and tracking system for smart mobility applications such as traffic, road, and railway environments. in this reviewTo begin, object detection and tracking were performed using two deep learning approaches: You Only Look Once (YOLO) V3 and Single Shot Detector (SSD). By contrasting the two methodologies, we can establish their applicability in the traffic context. The results were evaluated on both the road and the train. Second, based on the Mono depth technique, a distance estimate algorithm was developed. [9]

J. TensorFlow Object Detection API.

This paper presents a deep learning approach for traffic light detection by retraining the Inceptionv3 model and adapting a single shot detection (SSD) approach and image classification of two categories of objects. An open-source tool called TensorFlow Object Detection API is used to implement picture categorization and object detection.

It's been proven that more data implies better outcomes, therefore network depth must be properly selected. Recall values of up to 95% were achieved even for little objects. The frequency of false positives per image

ISSN No:-2456-2165

varies depending on whatever operating point you are using. The use of graphics processing units such as the NVIDIA Tesla K40 could help to improve the outcomes. Object proposal generation was also achieved via sharing a base network as CNNs became more popular. [10]

K. You Only Look Once (YOLO) method

The goal is to use the You Only Look Once (YOLO) method to distinguish between objects. This technique has only a few focal points as compared to other object detection techniques. Object detection is used in face detection, vehicle detection, web photos, and safety systems.

The dimension measure, box in charge of the item, is the only factor taken into consideration by the loss function, which quantifies mistakes in deduced boundary box positions. Loss functions like as classification, confidence, and localization are examples. The YOLO algorithm is used and presented for object recognition in this paper. This method can be applied to a wide range of real-world problems, including security, road safety, and, in any case, supporting outwardly disabled persons through the use of input. [11]

L. TensorFlow and MCNN

According to machine learning researchers, efficiently optimised feature extraction is the key to enhancing object detection accuracy utilising deep learning algorithms. This proposed method enhances the accuracy and speed of tracking objects in a video stream by employing MCNN instead of CNN for feature extraction.

MCNN has a success record of over 98 percent when it comes to detecting objects from videos. In terms of accuracy, MCNN exceeds all other object detection models. In Python code, Tensorflow and MCNN are used to build a native technique for monitoring the object detector's performance. [12]

M. SVM and NMS

On the PASCAL 2007 test dataset, their model's mean average precision (mAP) is 37.38 percent. Deep learning is the key to enhancing object detection accuracy. They also address how to increase performance using this approach in this publication.

The SVM's training method uses a similar number of object and background boxes. HOGs of sliding windows of three shapes and three scales are passed into the trained SVM to obtain scores on each label. Bounding boxes that overlap higher-score bounding boxes are removed using NMS.

The amount of proposals per image must be reduced heuristically. The VOC 2007 test dataset contains a vast number of items. To speed up our calculation, they can safely eliminate the. [13]

N. (RPN) and Fast R-CNN

Object detection has become one of the most important applications of deep learning. The Region Proposal Network (RPN) and Fast R-CNN are used together. Without the requirement for hand-crafted technology, deep models can be employed as a classifier and detecting device.

The trials were done out using the Tensor Flow API on the NVIDIA platform. An NVIDIA GeForce GTX 660 GPU with 2GB of RAM was used to speed up the learning process. Inception V2 and ResNet50 are the two models that deliver the best performance on the available RAM. 5000 training photos and 1000 validation images with a 500x500 pixel image size were used to create the ResNet50 model. The model had a prediction accuracy of 88.6%, with a 12.4 percent error rate of 27. [14]

O. Theoretical and experimental analysis of YOLO, R-CNN, CNN Algorithm

Neeraj Joshi, ShubhamMaurya, Sarika Jain et al conducted a theoretical and experimental analysis of the existing works in this paper. they discovered a research gap in existing works for visually impaired people. The method that could have been used to create a model that is much more feasible, i.e., a model that can run on low-power devices such as a smartphone.

Most of the time is spent in real-time. Object detection systems use YOLO object detection. This method is used to recognise and identify items due to its speed of detection and identification. It's not as accurate as Fast R-CNN or other Region-based CNNs, though. [15]

P. Possible future direction of image processing

In this paper they presented an overview of previous object detection research, outline the current main research directions, and discuss open problems and possible future directions in this manuscript. Learning incrementally, detecting new classes, and distinguishing between subclasses are all important problems. If this can be done unsupervised, we will be able to build new classifiers based on existing ones with little additional effort, dramatically lowering the effort necessary to learn new object classes. There has been some development in the application of new sensing modalities in recent years, especially depth and thermal cameras.In general, this type of problem has been solved by segmenting the image first and then labeling each segment e.g., Penget al. (2013).[16]

Q. YOLOv3 and Simple Online Real Time Tracking (SORT)

As a result of the impact of efficient data, performance benchmarks in terms of speed and accuracy have shifted. Objects are tracked across frames in traffic surveillance video using YOLOv3 and Simple Online Real Time Tracking (SORT). Effective detection and tracking may be seen on a dataset of urban automobiles.

In solving the tasks, Artificial Intelligence outperformed image processing techniques. The CNN model trained on an on-road vehicle dataset had a validation accuracy of 95.7 percent for auto, 95.5 percent for car, and 96 percent for heavy vehicles for single object detection. [17]

ISSN No:-2456-2165

R. RCNN, Faster-RCNN, SSD

Deep learning has had a significant impact on how the world is adapting to Artificial Intelligence. Region-based Convolutional Neural Networks (RCNN), Faster RCNN, Single Shot Detector (SSD), and You Only Look Once are some of the most popular object detection algorithms (YOLO).A python program was created and implemented for OpenCV. A total of 21 items were trained in this model. A dog, a motorcycle, a person, a potted plant, a bird, a car, a sofa, a sheep, a bottle, a chair, an airplane, a train, and a bicycle were among the 21 object classes that the model was trained to detect with 99 percent accuracy. [18]

S. Text to the speech synthesizer

Soniya1, B. Mounica1, A. Joshpin Shyamala1, Mr. D. Balakumaran2 et al stated

The main goal of this paper is to use deep learning to assist visually impaired people with object and text detection. There are two types of object detection: object localisation and image categorization. The recognised output is converted into speech using a text to speech synthesiser. Standard text forms in a range of fonts are recognised by the suggested algorithm. The best results are obtained when the item is closer to the camera than 70 cm and the lighting is good. The accuracy of the outcome is affected by the camera's specifications as well as the object's focus. A valuable feature is the ability to recognise several objects in a single frame. [19]

T. The ideal generic object detection algorithm

Li Liu1, Wanliet al.examined the most recent advances in this field, as a result of deep learning techniques.Deep learning approaches have made significant advances in the field of generic object detection. The optimum generic object detection method achieves a balance of high quality/accuracy and efficiency. For high-quality detection, objects in pictures or video frames must be precisely identified and recognised.

The entire detection operation must be accomplished in real time, with low memory and storage requirements, for high efficiency.[20]

ISSN No:-2456-2165

Sr.No.	Title	Year	Features
1	Real-Time Objects detection and Approach for Assisting Blind Humans	2017	Methods used-Ultrasonic sensor Accuracy-70%
2	Object Detection Combining Recognition and Segmentation.	2018	Methods used-Topdown object hypothesis Accuracy-76%
3	Human objects detection, behavior recognition	2018	Methods used-L-CNN, SSD Accuracy-80%
4	The movement of SCARA robot	2019	Methods used-CNN, SCARA Accuracy-95%
5	Neural and FPGA technologies used to solve the real-time object/image recognition problem	2021	Methods used-FPGA, YOLOv3 Accuracy-80%
6	Region Proposal Network (RPN) which shares entire image convolutional features	2016	Methods used-RPN, YOLOv3 Accuracy-92%
7	Object Discovery and Grasp Detection using CNN	2016	Methods used-CNN, Shared CNN Accuracy-78%
8	RCNN And F-RCNN	2019	Methods used-R-CNN, F-RCNN, SSD Accuracy-95%
9	Object detection, localization and tracking system	2019	Methods used-YOLOv3, SSD Accuracy-93%
10	TensorFlow Object Detection API.	2018	Methods used-Tensor flow, SSD Accuracy-95%
11	You Only Look Once (YOLO) method	2021	Methods used-YOLO Accuracy-72%
12	TensorFlow and MCNN	2019	Methods used-MCNN, Tensor flow Accuracy-98%
13	SVM and NMS	2018	Methods used-SVM, NMS Accuracy-86%
14	(RPN) and Fast R-CNN	2019	Methods used-RPN, F-RCNN Accuracy-88.6%
15	Theoretical and experimental analysis of YOLO, R-CNN, CNN Algorithm	2020	Methods used-YOLO, RCNN Accuracy- 77%
16	Possible future direction of image processing	2015	Methods used-MCNN, R-CNN Accuracy-
17	YOLOv3 and (SORT)	2019	Methods used-YOLOv3, SORT Accuracy-93%
18	RCNN, Faster-RCNN, SSD	2018	Methods used-RPCNN, FRCNN, SSD Accuracy-80%
19	Text to the speech synthesizer	2020	Methods used-NLP Accuracy- 92%
20	The ideal generic object detection algorithm	2019	Methods used-Generic Object detection algorithm Accuracy-95%

Table 1: Summary Of Literature Review

III. CONCLUSION

In this paper, we have reviewed deep learning algorithms for real time object detection. The mainobjective of CNN algorithmis to detect various objects in real time video sequence and track them in real time. This model works well on the object trained in terms of detection and tracking, and it can be used in certain situations to identify, track, and respond to specific targeted objects in video surveillance. This real-time ecosystem analysis may produce great results for any business by allowing security, order, and utility. The models' accuracy varied depending on the complexity of the learning and the quantity of the data set, with results ranging from early 70% to late 98 %. We conclude that using the correct data collection and CNN, we can get high-accuracy outcomes.

IV. CHALLENGES

- Viewpoint Variations
- Deformation
- Occlusion
- Illumination Conditions
- Cluttered or textured Background

ISSN No:-2456-2165

REFERENCES

[1.] Real-Time Objects Recognition Approach for Assisting Blind People. Jamal S. ZraqouWissam M. Alkhadour and Mohammad Z. Siam,201.

[2.] Object Detection Combining Recognition and Segmentation [Fudan University, Shanghai, Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen] 2018

[3.] Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN SeyedYahyaNikouei†, Yu Chen†, Sejun Song‡, RonghuaXu†, Baek-Young Choi‡, Timothy R. Faughnanξ018

[4.] Real-Time Robotic Grasping and Localization Using Deep Learning-Based Object Detection Technique Mohannad Farag,2nd Abdul NasirAbdGhafar Mohammed Hayyan ALSIBAI2019

[5.] Neural Network for Real-Time Object Detection on FPGA Edward Rzaev Moscow, Russian Federatio Anton Khanaev Moscow, Russian Federation AleksandrAmerikanov 2021

[6.] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks ShaoqingRen, Kaiming He, Ross Girshick, and Jian Sun2016

[7.] Object Discovery and Grasp Detection with a Shared Convolutional Neural Network Di Guo, Tao Kong, Fuchun Sun and Huaping Liu 2016 IEEE

[8.] Real Time Object Detection and Tracking Using Deep Learning and Open CV Real Time Object Detection and Tracking Using Deep Learning and Open CV

[9.] Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility Zhihao Chen 2nd RedouaneKhemmar 3rd Benoit Decoux 2019 IEEE

[10.] Object detection and classification through deep learning approaches DeepikaSolanki, Mr. Shrwan Ram

[11.] Object Detection using Deep Learning with OpenCV and Python Shreyas N Srivatsa1, Amruth2,SreevathsaG3, Vinay G4, Mr. Elaiyaraja P5 1-4Student, Dept. of Computer Science Engineering, Sir MVIT, Karnataka, India 5Professor, Dept. of Computer Science Engineering, Sir MVIT, Karnataka, India

[12.] Research on Object Detection in Video Streaming Using Deep Learning Shaba Irram, Sheikh Fahad Ahmad.

[13.] Object Detection based on Convolutional Neural Network Shijian Tang Department of Electrical Engineering Stanford University sjtang@stanford.edu Ye Yuan Department of Computer Science Stanford University.

[14.] Real Time Object Detection using Deep Learning: A Webcam Based Approach Sandipan Chowdhury1, Prabuddha Sinha2 1,2Department of Information Technology, Government College of Engineering and Ceramic Technology, West Bengal.

[15.] Real-Time Object Detection and Identification for Visually Challenged People Using Mobile Platform Neeraj Joshi, ShubhamMaurya, Sarika Jain.

[16.] : Current and Future Directions Article in Frontiers in Robotics and AI · December 2015

[17.] Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications Mohana1, HV Ravish Aradhya 2.

[18.] Real Time Object Detection and Tracking Using Deep Learning and OpenCV Conference Paper · July 2018

[19.] Text and Object Recognition using Deep Learning for Visually Impaired People R. Soniya1, B. Mounica1, A. Joshpin Shyamala1, Mr. D. Balakumaran2

[20.] Li Liu1,2 · Wanli Ouyang3 · Xiaogang Wang4 · Paul Fieguth5 · Jie Chen2 · Xinwang Liu1 · Matti Pietikäinen2 Deep Learning for Generic Object Detection: A Survey

[21.] DipmalaSalunke, Ram Joshi, Prasad Peddi, Deepak Mane, Deep Learning Techniques for Dental Image Diagnostics: A Survey, Springer “Lecture Notes in Social Network, 4th International Conference on Computer Networks, Big Data and IoT, ICCBI 2021.

[22.] Krishna Sheth, Prof. D. T. Salunke,Atul Raj, AsthaKubavat, Dental Disease Classification using Deep Learning: A Survey, GIS Science Journal, 2021.volume 8 issue 6,pp.1829-1841.

[23.] HarshalKapase, DipmalaSalunke, KalyaniGalande, TanmaySonna, DeepaliPawar,A review on: Sentiment polarity analysis on Twitter data from different Events, International Research Journal of Engineering and Technology (IRJET),Volume 5, Issue 3,2018,pp. 1479-1482.