I.J. Image, Graphics and Signal Processing, **2022, 2, 47-60**
Published Online April 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2022.02.05

Recent Object Detection Techniques: A Survey

Diwakar
Research scholar at Babasaheb Bhimrao Ambedkar University Lucknow (A Central University)
Email: diwakarmsccs0@gmail.com

Deepa Raj
Associate professor at Babasaheb Bhimrao Ambedkar University Lucknow (A Central University)
Email: deepa_raj200@yahoo.co.in

Received: 27 October 2021; Revised: 20 November 2021; Accepted: 02 December 2021; Published: 08 April 2022

Abstract: In the field of computer vision, object detection is the fundamental most widely used and challenging problem. Last several decades, great effort has been made by computer scientists or researchers to handle the object detection problem. Object detection is basically, used for detecting the object from image/video. At the beginning of the 21st century, a lot of work has been done in this field such as HOG, SIFT, SURF etc. are performing well but can't be efficiently used for Real-time detection with speed and accuracy. Furthermore, in the deep learning era Convolution Neural Network made a rapid change and leads to a new pathway and a lot of excellent work has been done till dated such as region-based convolution network YOLO, SSD, retina NET etc. In this survey paper, lots of research papers were reviewed based on popular traditional object detection methods and current trending deep learning-based methods and displayed challenges, limitations, methodologies used to detect the object and also directions for future research.

Index Terms: Object detection, Convolutional Neural Network, deep learning techniques

1. Introduction

In the last decade, object detection is the most growing and widely used and active area of research in the field of computer vision. Worldwide object detection supports a lot of applications such as Automated CCTV surveillance, Self-driving cars, medical imaging, architecture, industrial robotics, satellite image and many more in real life. The goal of object detection is somehow identifying (classification) the object present in the image/video and draw the bounding box (localization) around the object. These two tasks (localization and classification), in traditional object detection methods (such as HOG, SIFT, SURF, VIOLA-JONES DETECTOR, etc.) obtains by handcrafted feature-based methods, which need to extract the various features of an image by using feature descriptors (HOG, HAAR like, SHIFT etc.) to recognize groups of pixels that may be related to an object. And then these features are fed into a regression model (such as SVM or Adaboot) to identify object categories. In 1999, scale-invariant feature transform is a feature detection algorithm proposed by David Lowe[1]. Which extracts the different key points (highly distinctive locations) from a similar type of image dataset. It is a way to describe the local image content or signature to a local area like circle, corner blob etc. It can rigidly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor extracts the features that are invariant to image scale and rotation, orientation, illumination changes and robust against affine distortion. SIFT result shows high accuracy (recall rates) but is not a good choice for real-time object detection since it's computationally heavy and not effective for low powered devices. Thereafter in 2001, Viola-Jones object detection method was proposed by Viola and Jones[2]. Used for face detection in the image and the idea behind the algorithm is looking for more relevant features like forehead, nose, eyes, and lips like that features detected by haar-like features. It is quite popular and it was the first real-time object detector and is still in use such as in face filter applications (Instagram, snap chat). Later on, in 2005, Researchers Dalal and Triggs proposed a HOG descriptor[3] for pedestrian detection from static image and video as well. HoG algorithm is too precise and high success rate which computes the histograms of image gradients orientation and uses them as image features. HOG is being almost a decade old but still heavily used today with excellent results. If you have a powerful computer/machine then HoG is not a bad choice. Directly after in 2006, SURF was proposed by H. Bay et al.[4]. inspired by the scale-invariant feature transform (SIFT) descriptor. SURF is fast and more robust than SIFT, used for image recognition, 3D Reconstruction, classification and many others. It uses 64 dimensions feature vector as compare to SIFT (128 dimensions) which is fast enough to be used in real-time object detection applications such as medical image registration (a fast method to retrieve CT images of a patient) effectively and efficiently. Moreover, to speed up the matching process, the Sign of laplacian is added to the feature vector. Hence, the matching gives a slight increase in performance. These are just a few of the popular traditional object detection methods. In the last few years, deep

learning-based object detection methods (like R-CNN family, SSD, YOLO, SPP-net etc.) which is now become a key component in the field of computer vision and gained fame for their capability to process visual information. But as of now still, deep learning can't solve all the computer vision problems. Whereas for few problems traditional techniques are still performing well. deep-learning opens many doors to do something with traditional techniques so that it could possible to overcome the limitations or challenges of traditional approaches by the use of deep learning strategies with it. The objective of this paper is to review the significant efforts done in deep learning-based object detection methods with various applications and research gaps.

2. Neural Based Models for Object Detection

Undoubtedly, traditional methods for object detection still working effectively but in the field of computer vision deep learning added a huge boost in development and now it becoming part of our everyday life. neural-based approaches added new features, improved the accuracy, efficiency and leading the lots of ways to develop the new technologies using deep learning methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) etc.

Object detection aims to locate the object in the image (localization), draw the bounding box around the object and identify the class of object. This is exactly what trying to do in object detection. Let's understand how neural network works and what is the intuition behind neural network.

Neural Network or artificial neural networks (ANNs) is a set of algorithms designed for computers or machines in such a way as the human brain works. So that machines will be able to recognize the patterns, learn the thing, and make decisions like humans. A simple Neural Network consists of the input layer, which takes signals as input and passes them to the next layer called 'hidden layer' (could be one or more hidden layers) that does feature extractions, some mathematical calculations and finally produced the result as output layer (As shown in below fig-1). Layers are connected through nodes and these connections form a network of interconnected nodes named as neural network. In the case of object detection, for better understanding let's take an image that forms arrays of pixels and fed as input to identify the image. Each input neuron has some value (called activation) ranging from 0 to 1 as represented grayscale values concerning the pixels. Then passes the input to the hidden layer. Inter-connection of neurons (between input and hidden layer) assigned some random weights those multiplied to corresponding input pixel values and add a bias to them. Then the weighted sum of input is fed to the activation function (e.g., sigmoid function, threshold function, ReLU etc.) to determine which node should be activated or fire for feature extraction. And finally, the model predicts the outcome by applying the appropriate function to the output layer. Then after calculating the cost function and minimizing the error rate by back-propagating through the network and keep adjusting the weights until they fit to the training models. For maximum accuracy need to do several iterations and each iteration weights are adjusted based on error. And finally, compare the outcome with the actual result.

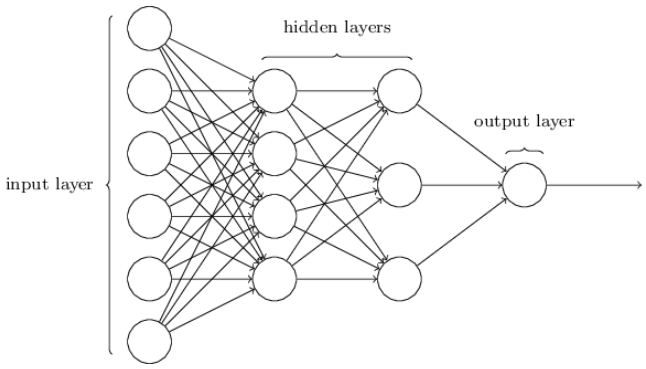

Fig.1. Neural Network

Here, in object detection, using CNN's because it is more powerful than ANN's. A Convolutional neural network (CNN or ConvNet) is a class of deep-learning that has one or more CNN layers, which is used for image classification, recognition and segmentation. Here, the term 'Convolution' refers to a mathematical operation on two functions that produce a third function that shows how the shape of one function is influenced by another function. CNN has many types of network layers (e.g., Convolutional layer, Activation layer, fully-connected layer) with different structures and mathematical operations. Each layer performs some pre-defined function on its input data and extract features and finally based on matching features CNN model predicts the belonging class.

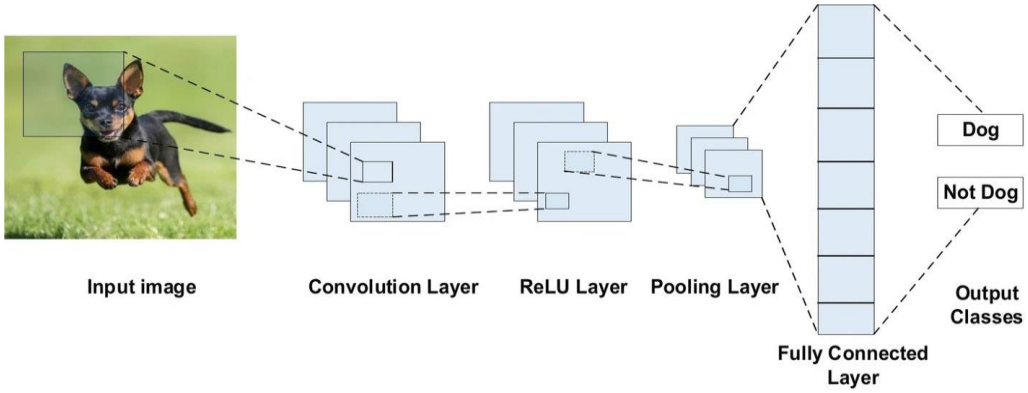

Fig.2. CNN architecture

As shown in the above fig-2, typical CNN consists of three types of layers, Convolutional layer, pooling layer, and fully-connected layer. The first layer is used to extract the various features from the input image by performing the convolutional mathematical operation between the input image and with a particular size of the filter by sliding over the image and computing the feature map (e.g., corners, edges etc). In the next layer (pooling), the size of the feature map reduces so that decrease the computational cost. There are several types of pooling methods (Max-pooling, average pooling etc.) available and can use any method depending upon work. In the next layer (fully-connected layer), every neuron in one layer is connected with every neuron of another layer in the same way as a traditional multi-layer perceptron neural network. A fully-connected layer takes input as flatten vector/matrix, classify the image and predict the output classes.

In the deep learning era, this object detection task is generally, categories into two stages. "Two-stage detectors" such as faster R-CNN, mask R-CNN are region-based detectors that obtain higher accuracy with a penalty of computational cost. Instead of a "One stage detector" such as YOLO (You look once only) or Single Shot Detector (SSD) is complete the detection task only in one network containing a single feed-forward fully convolutional network that directly localizes and classified the object. One stage detectors are performing well in terms of efficiency (speed) and providing excellent results but little bit lacking in accuracy. Such models reach lower accuracy rates, but they are much faster than two-stage object detectors.

2.1 Region-Based Convolutional Neural Networks (R-Cnn)

R-CNN author, R. Girshick et al.[5] proposed an outstanding region-based CNN model which is used to detect the object in the image. here, rather than discussing much about the history of CNN and R-CNN, directly jump to the intuition behind CNN and how it become popular? The ultimate goal of R-CNN is to take the image as input (in image objects could be car, pedestrian, motorcycle etc.) and as result (output) provide the bounding boxes and labels (class) to each object present in the image.

R-CNN detects the objects in three subsequent stages. The first stage is category-independent region proposals. lots of region proposal methods are available like slidingWindow, selective search, objectiveness, superpixels. here author adopts the selective search method which extracts approx. 2000 botton-up region proposals and each region proposal extract the 4096-dimensional feature vectors. The next step is (second module) extracts the fixed-length feature vector through a strong convolutional neural network (pre-trained AlexNet). And finally, in the third stage linear Support vector machine classifies to predict the object within each region and identify the class of object. On the VOC2010 dataset, R-CNN gives Mean Average Precision (mAPs) of 53.7% and the ILSVRC2013 object detection dataset gives an improved mAP of 31.4%.

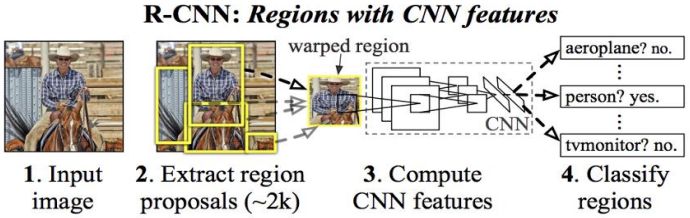

Fig.3. R-CNN (Regions With Cnn Architecture)

As shown in above fig-3, R-CNN first extracts region proposals using a selective search algorithm and classify ~2k region proposals per image and then resize all extracted regions so that regions can match with CNN input size and pass them through the network. This process takes a huge amount of time to train and test the image. So that can't use in

real-time image detection. Moreover, R-CNN uses selective search therefore no learning happens which is leads to bad region proposals. So overall can say 'training' is expensive in both space and time.

2.2 FAST R-CNN

A year later in 2015, author R. Girshick enhanced the R-CNN model and proposed updated architecture named as Fast R-CN[6]. Fast R-CNN approach is similar to R-CNN but the intuition behind fast R-CNN is the reduce the computation time by running the neural network once on the whole image. Or can say instead of feeding approx. on each of as many as two thousand regions of interest to the convolutional network every time to generate the feature map. It takes the whole image and region proposals as input in convolutional neural network architecture in one forward propagation. For faster detection author used 'Truncated Singular Value Decomposition' (compression method) which is speedup the forward pass computing the fully connected layers. Moreover, fast R-CNN removes the storage requirement or space problem (to store feature map) by combining the different parts such as ConvNet, RoI pooling, and classification layer in a single architecture. Experimental result shows that increase Mean Average Precision (mAPs) on VOC2007 to 66.9% from 66.0%(RCNN)

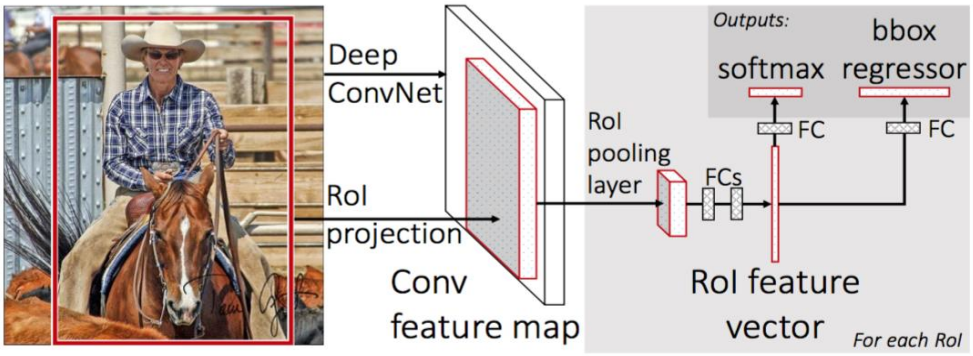

Fig.4. Fast R-Cnn Architecture

In just one year Author, introduced Fast R-CNN with training time reduced to 9.5 hrs. as compared to R-CNN with 84 hrs. Achieved high accuracy, but still uses selective search strategy which is a time-consuming process, slowing down the whole system. Moreover, the test time per image is approx. 2 seconds, so that not possible to use for real-time object detection.

2.3 FASTER R-CNN

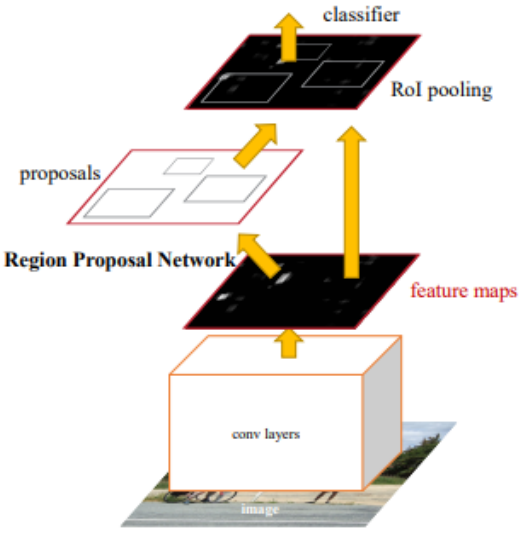

Fig.5. Faster Rcnn Architecture

Just a few months later in 2015, Ren et al. proposed Faster R-CNN[7]. So, for had seen region based-CNN (R-CNN and fast R-CNN) models are using selective search algorithm which is slowing down the detection, consuming more time and affecting the overall network. So that can't use appropriately for real-time object detection. Therefore,

the author mainly focuses on speed and replaced the selective search algorithm with "Region Proposal Network" (RPN) that provides almost cost-free region proposals. RPN is basically, a fully convolutional network is used for simultaneously predicting object bounds and objectness outcomes at each position. Region Proposal Network trained end-to-end specifically for the task of generating high-quality region-proposals. Faster R-CNN Achieved Mean Average Precision (mAPs) on VOC2007 to 69.9% from 66.9%(Fast RCNN).

Faster R-CNN fixes the problem of selective search algorithm by replacing it with Region Proposal Network and become first nearly real-time object detector with running time 10 times dropped than fast RCNN. This was really, excellent work and can see how deep learning growing exponentially. (See the table for more details). although complete system architecture is connected sequentially so that performance of the system depends on how efficiently previous systems performed.

2.4 MASK R-CNN

Region-based CNN (Faster RCNN) furthermore, extended by He et al. in march 2017[8] named as Mask R-CNN. This is another variant of a Deep Neural Network object detector that is Mainly focused on instance segmentation tasks. Mast RCNN is the extension of Faster R-CNN with an additional parallel mask branch for predicting segmentation masks on each Region of Interest(RoI). Another property of mask RCNN is ROIAlign which is just a simple improvement of the very popular ROI operation. however, there is no quantization in the ROIAlign operation. Moreover, Feature pyramid networks(FPN) are integrated as the effective backbone in mask R-CNN to generate ROI (region of interest) features. Mask R-CNN results on the Microsoft COCO test set based on ResNet-101, achieved a mask AP of 35.7 and running at 5 FPS.

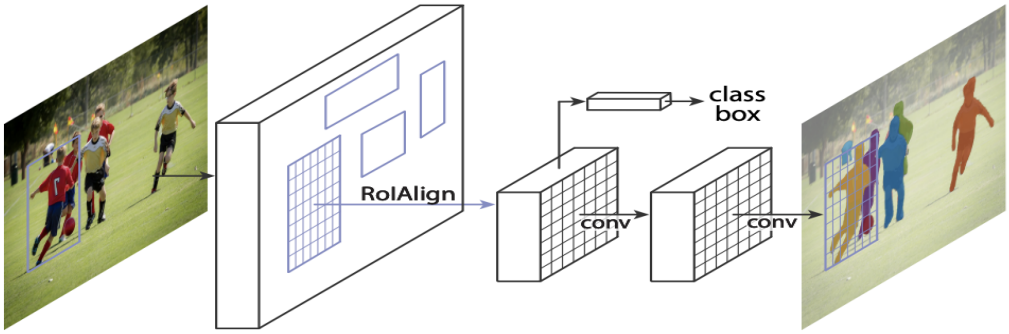

Fig.6. Mask R-Cnn Architecture

Mask- R-CNN, plays important role in real-world vision applications such as self-driving cars, road sign detection, video surveillance in security, tumour detection in the medical field are of them.

2.5 YOLO

Yolo stands for "You only look once". In the Deep learning era, Yolo is the real-time one stage object detector proposed by Redmon et al. in 2015 [9]. Over the world researcher trying to develop a kind of machine or algorithm that can able to detect the object with excellent speed and accuracy. YOLO is one of the success achievement till dated. it serving in our life in various applications such as traffic signals, person detection, parking meters, vehicle detection etc. From 2015 onwards, various improvements (different versions V1, V2, V3 etc.) were made by the author. YOLO is an extremely fast unified, Real-Time Object Detection model, simple to construct and can be trained directly on the full image. previous methods like the R-CNN family are detecting the object in several steps, but YOLO requires only a single neural network for the full image. The intuition behind YOLO is to divide the image into regions (cell) and split the regions into a square grid of dimensions S×S. the cell (grid cell) in which the centre of the object is present, the cell only responsible for detecting that object. Each cell in the grid will predict bounding boxes and a confidence score for each box. In real-time object detection, on the PASCAL VOC dataset, the YOLO model processes the object detection in 45 frames per second and achieved 63.4% mAP (approx. double the mAP of other real-time object detectors). Generally, one stage detector tends to be of lesser accuracy as compared to two-stage detectors (R-CNN family) but is significantly faster. Undoubtedly, YOLO doing excellent but still has few limitations such as struggles to detect small objects (like flocks of birds) that appear in groups, low recall and more localization error, can detect up to 49 objects etc. So later on, the author further improved the accuracy by keeping tremendous speed in the next YOLO versions.

2.6 YOLOV2

Author, Redmon and Farhadi in 2017, proposed second version of YOLO[10] mainly focused on improving recall and localization. in tend to improvement of YOLO, author used new ideas such as '**Batch Normalization'**- that gets

more than 2% improvement in mAP also helps regularise the model, **'High Resolution Classifier'**- earlier in first version of YOLO used classifier network (input size) at 224×224 now in second version increases the resolution to 448×448 that increases of almost 4% mAP, '**Convolutional With Anchor Boxes**'- author improve the convolutional layer by removing the fully connected layer from original YOLO (YOLOv1) and uses the anchor boxes for predict bounding boxes that leads to improve the recall (7% improved) from 81% to 88% but decrease mAP slightly from 69.5% to 69.2%, '**Dimension Clusters**'-instead of choosing anchor boxes(priors) by hand, author uses k-means clustering on the training set bounding boxes to automatically get good priors, '**Fine-Grained Features**'-In first YOLO author lagging to detect the small object then here in V2, basically improvement of localizing the smaller object. YOLO9000 concatenates the high-res features with the low-res features by stacking adjacent features into different channels which gives a modest 1% performance[10] improves, '**Multi-Scale Training**' for the network to be robust to running on object/images of different sizes author trained the model for different input sizes(changes the network in every few iterations). Every 10 batches of network randomly choose a new image dimension size. Therefore, the same network can able to predict detections at different resolutions.

YOLOV2 gives state-of-the-art detection accuracy on PASCAL VOC2007 increases in mAP 78.6% with 40 fps as compared to YOLO with 63.4% mAP and 45fps.

2.7 YOLOV3

After two years in 2018 YOLO authors, Redmon and Farhadi proposed YOLO-V3[11] which is an Incremental Improvement of the YOLO family. Towards object detection, a lot of algorithms are proposed by researchers or scientists and the competition is all about how accurately and efficiently (quickly) objects are detected. In this incremental improvement of the YOLO, the family author focused on accuracy with maintaining the speed. YOLOv3 uses logistic regression (Multi labels prediction) to predict the objectiveness (confidence) score for each bounding box. YOLOv3 can detect more smaller objects as compared to YOLOv2. Moreover, it used a robust feature extractor called the darknet-53 model (a convolutional neural network that is 53 layers deep). These improvements made YOLOv3 run in 22 MS at 28.2% mAP, as accurate as SSD but 3x faster.

After this version author stopped working on this algorithm due to concerns regarding the possible negative impact of his works as Joseph Redmon announced about it in his last paper of YOLOv3 [ref]. So next improve YOLO version proposed by Bochkovskiy et al. in April 2020.

2.8 YOLOV4

In YOLOv4[12], the Author made a significant upgrade in terms of accuracy (average precision) and Speed (FPS) both by 10% and 12%, respectively. The intuition of YOLOv4 is the optimization of neural networks detector for parallel computations by implemented new features such as Weighted-Residual-Connections (WRC), Cross mini-Batch Normalization (CmBN), CIoU loss, Cross-Stage-Partial-connections (CSP), Self-adversarial-training (SAT), DropBlock regularization, Mish activation, Mosaic data augmentation and other features which leads to achieved 43.5% AP (65.7% of AP50) on Microsoft COCO dataset at a real-time speed of approximately 65FPS on NVIDIA Tesla V100 (GPU). The main difference from YOLOv3 is the backbone (feature extractor) used by YOLOv3 is Darknet53 and YOLOv4 using CSPDarknet53 which reduces the training time.

The real-time object detection space remains hot and researchers and scientists are continuously working on it. Last year in May 2020, Glenn Jocher (Founder & CEO of Utralytics) released the next version of YOLO (**YOLOV5**) on GitHub [13]. But still, there is no paper available on it. Moreover, in real-time object detection still tricky and research is going on tends to improve accuracy and speed.

2.9 SSD

SSD stands for "Single Shot MultiBox Detector". it is another masterful one-stage detector (single deep neural network) proposed by Wei Liu et al. in 2015[14]. SSD architecture is shown in the below figure. SSD is significantly more accurate as region-based technique (R-CNN) as compared to previous single shot algorithms (YOLO) and use of multiple feature map which predicts different scale and aspects radios (core idea of SSD algorithm). This leads to easy training and high accuracy even though on low-res (smaller input image size) images. Experimented results on VOC2007 test, SSD achieves 59FPS with mAP 74.3% on Nvidia Titan X, as compare to Faster R-CNN 7 FPS with mAP 73.2% and YOLO 45 FPS with mAP 63.4%.

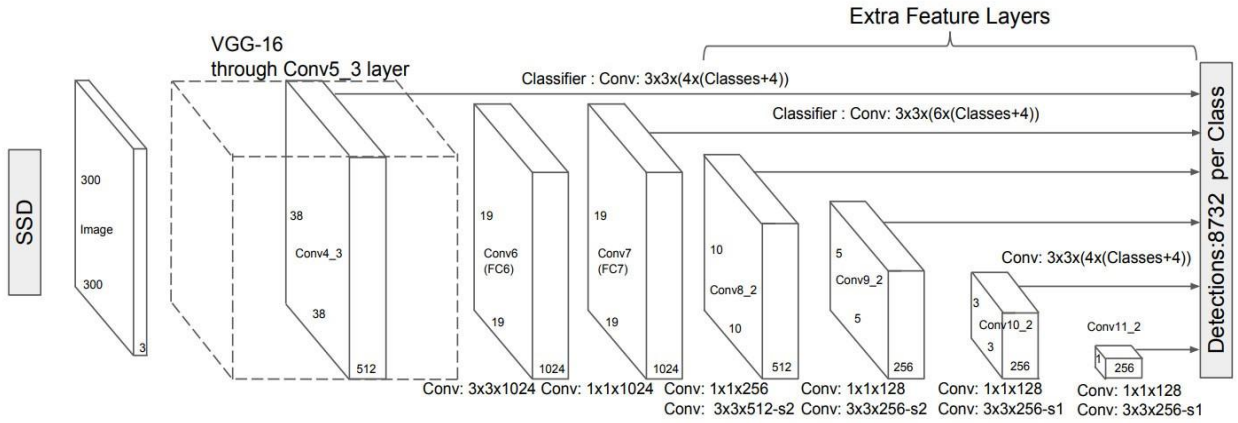

Fig.7. Architecture of Ssd

2.10 RETINANET

In the deep learning era, T.-Yi Lin et al. in 2018 proposed an algorithm[15] for object detection using RetinaNET. as discussed, algorithms for object detection and had seen detection is categorised in two stages, two-stage detectors such (RCNN family) and one stage detectors (YOLO). Two-stage detectors are good in accuracy but penalty as speed and one stage detectors are faster (speed) but lacking in accuracy. So, the author analyses, one-stage detectors suffer from an extreme foreground-background class imbalance problem which is encountered due to dense sampling of anchor boxes. So the author introduced a new loss function to address this class imbalance problem with one-stage object detection models is called- **Focal Loss** (an enhancement over Cross-Entropy Loss) which decreases the loss contribution from easy examples and increases the importance of correcting miss-classified examples. It is widely using in aerial and satellite imagery. Experimental result on Microsoft COCO dataset with ResNeXt-101-FPN achieved 40.8% AP (Average Precision) and with ResNet-101-FPN achieved 39.1% AP.

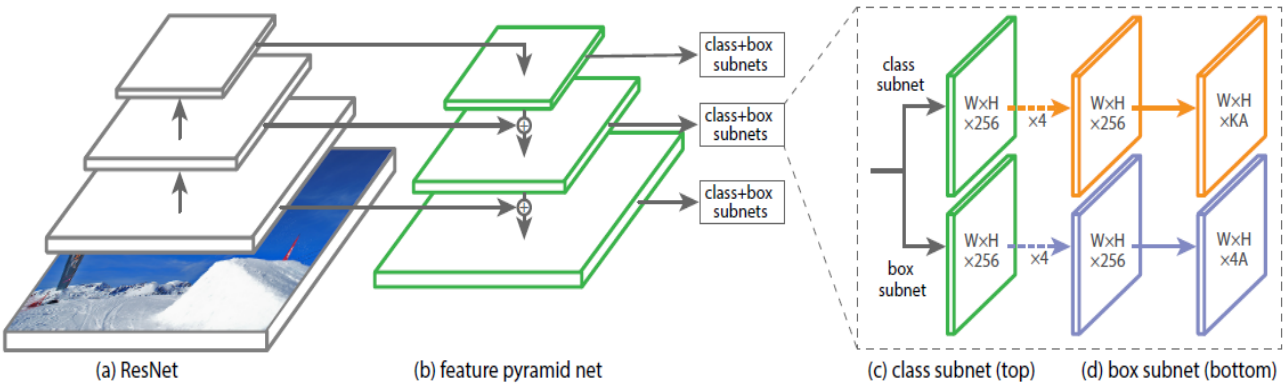

Fig.8. Architecture Retinanet

2.11 REFINEDET

In 2018, object detection has achieved significant advancement by the contribution of a model name as 'RefineDet'[16] proposed by Zhang et al. which achieves accuracy similar to two-stage detectors (like Faster R-CNN) with maintaining the efficiency (speed) as one-stage detectors (SSD). Author intuition here somehow inherits the highest accuracy from the two-stage detector (R-CNN) and high speed from the one-stage detector (SSD). As shown In below figure-7, RefineDet is constructed by two interconnected modules, one is the anchor refinement module and the second is the object detection module (ODM). ARM identifies and eliminate the negative anchors and narrow the search space for the object classifier. Moreover, in the feature map cell, fine-tune the locations and size of anchors to provide better initialization. While ODM refined the anchors for higher efficiency (speed) and finally transfer connection block and transfer the features in an ARM to accurately predict the size, locations and corresponding multiclass labels. The experimented result on VOC2007 achieved 81.8 mAP with 24.1 PFS (Titan X) and VOC2007 achieved 80.1 mAP.

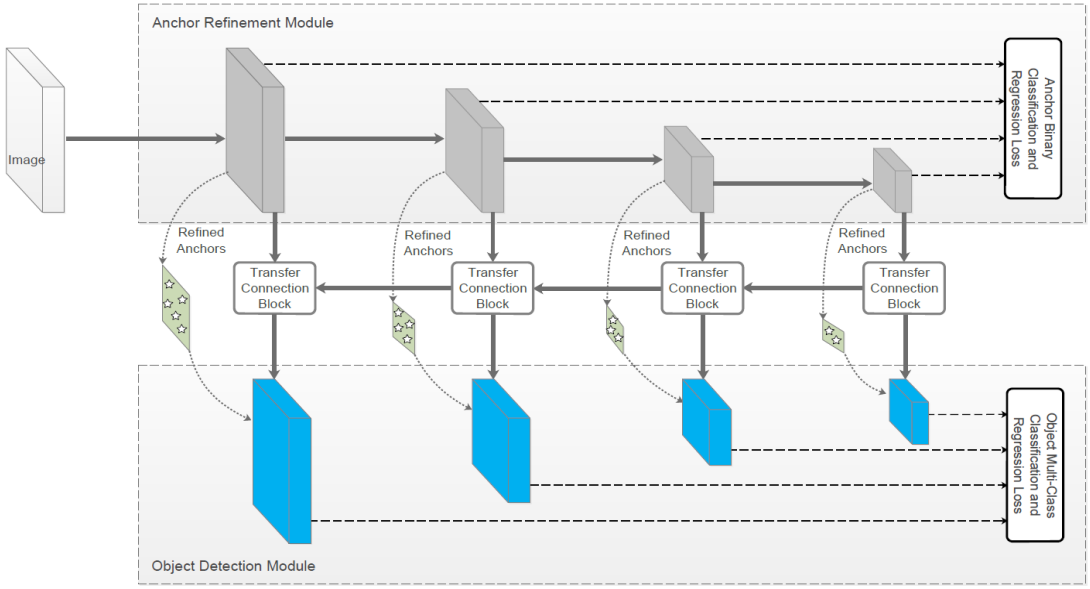

Fig.9. Refinedet Network Architecture

3. Survey Related to Object Detection

Lots of research are going on to detect the object in an image, in a video, underwater, in air, in the medical system and many more. Many research papers surveyed and display below.

Shen et .al. propose a novel method to detect a moving object

Felzenszwalb et al.[17] explain the Cascade object detection with deformable part models in his paper. Girshick et. al.[5] described region-based convolutional networks (R-CNN) for accurate object detection and segmentation. Girshick et.al.[6] explain Fast RCNN with detailed description and method using neural network.

Ren et.al.[7] described the Faster RCNN in his paper which is for real-time object detection with region proposal networks, Kaiming, et al.[8] explain the Mask RCNN.

Joseph, et al.[9] described YOLO: Unified, real-time object detection method which is also used for real-time object detection. Redmon et al.[10] explain the YOLO9000 which is better, faster, stronger.

Redmon et. al.[11] explain Yolov3 which is an incremental improvement of YOLO and YOLOv2.

Bochkovskiy et. al.[12] defined the new procedure of real-time object detection as YOLO4 Liu et al.[14] explained SSD: Single shot multibox detector.

Lin et. al.[15] explained focal loss for dense object detection. Kong et al.[18] described anchor-free framework for object detection with detailed method and description.[19]A new Moving Object Detection method based on Background Substation and Frame difference. [20]Deep Learning for real-Time capable of Object Localization and Detection on Mobile Platforms(2017). [21]Deep Lac: Deep Localization Alignment and classification for fine-grained Recognition.

Malisiewicz et al.[22] describe an exemplar-based approach in a discriminative framework for object detection based on training a separate linear SVM classifier for every exemplar in the training set (2011).

Yin et al.[23] presented a maximally stable extremal region (MSER) based Robust Text Detection method from natural scene images that work well in the low-quality image as well.

Zhang et al.[24] explained how ConvNets efficiently use for detection and localization tasks and proposed a multiscale, sliding window approach for localization, classification and detection using Convolutional Networks.

Erhan et al.[25] proposed an object localizing method based on a deep CNN as a base feature extraction and learning model which predicts multiple bounding boxes at a time.

He et al.[26] describes SPP-net which computes feature map from the entire image only once and shows excellent accuracy in classification or detection tasks.

Yoo et al.[27] proposed an object detection method using a deep convolutional neural network. Where cast an object detection problem into an iterative classification problem.

Gidaris et al.[28] in 2015, presented rich Convolutional Neural Network-based representation for object detection relies on diversification of the discriminative appearance factors and encoding of semantic segmentation-aware features.

Ghodrati et al.[29] proposed an algorithm for detection based on convolutional neural network activation features that used one kind of feature for both localization and detection.

Kong et al. presented HyperNet architecture which is a fully trainable deep architecture for handling region-proposal generation and object detection jointly[30].

Cai et al. proposed a unified deep neural network for fast multi-scale object detection[31].

Kang et al.[32] proposed a deep learning framework named T-CNN for object detection especially from videos by integrating the temporal and contextual information from tubelets obtained in videos.

Shen et al.[33] proposed a paper of another flavour of SSD as Deeply Supervised Object Detector framework for training the object from scratch. Also In 2017, Fu et al.[34] propose another variation of SDD to achieve greater performance by combining a classifier 'ResNet-101' with a fast detection framework namely 'SSD'

Kong et al.[35] in 2017, presented a framework for generic object detection focused on two major problems as negative sample mining and multi-scale object localization. In this paper, the author addressed these two problems by design the reverse connection and for the second problem describe the objectness prior to significantly narrow the searching space of objects.

In this paper[36], Dai et al. proposed two new models that enhance CNNs' capability of modelling geometric transformations named as 'deformable convolution' and 'deformable RoI pooling'.

Tychsen-Smith et al.[37], presented a region-of-interest detector and classification object detection framework for sparse estimation with CNNs which reduces manual engineering and improves performance with real-time and near real-time evaluation rates.

Zhou et al.[38] work on the Scale problem of object detection and develop a Scale-Transferrable Detection Network for detecting multi-scale objects in images by using it to make image super-resolution. Explained in detail in the paper.

He et al.[39] proposed an objects relation model for object detection in which processes a group of objects concurrently as objects interacted between their appearance feature and geometry, thus allowing modelling of their relations.

Law et al.[40] proposed a new approach for object detection. In which detecting an object bounding box as a pair of key points. This approach achieves a 42.1% AP on MS COCO and excellent performance over one stage detector.

Pang et al.[41] presented a model to tackle the imbalance problem during the training process. It is an effective model towards balanced learning for object detection.

Chen et al.[42] proposed a new cascade architecture for instance segmentation in which detection and segmentation tasks interweave for joint multi-stage processing and it adopts a fully convolutional branch to provide spatial context.

Li et al.[43], addressed the scale variation challenge in object detection focus to generate scale-specific feature maps with a uniform representational power. This model achieved state-of-the-art single-model results of 48.4 mAP with ResNet-101 backbone on the COCO dataset.

Wu et al.[44] describe an approach to handle the scale problem in which generates multiscale positive samples as object pyramids and refines the detectors at different scales.

Tang et al.[45], describes in object detection how can improve the efficiency using several key optimization methods such as weighted bi-directional feature pyramid network (BiFPN), compound scaling method etc.

Zhang et al.[46] propose a method to address the inconsistency problem of the training process of detectors and it is significantly improved by the use of Dynamic Label Assignment and Dynamic SmoothL1 Loss.

Sun et al.[47] presented a simple, unified network composed of a backbone network, Sparse method for object detection in image in which replace hundreds of thousands of candidates from RPN with a small set of proposal boxes.

Szegedy et al.[48] proposed a deep convolutional neural network architecture for classification and detection in the ImageNet and optimize the quality by increasing the depth and width of the network and also improve the utilization of the computing resources inside the network.

Pramanik et al.[49] proposed two new approaches for object detection and tracking from videos, named as granulated RCNN and multi-class deep SORT (MCD-SORT).

Table 1 shows the comparative study of all the methods used for object detection using different techniques, methods of localization, classification segmentation with important key points and application of method used.

Table 1. Comparative study of the object detection technique

METHODS	AUTHOR	YEAR	KEY POINTS	APPLICATIONS
SIFT	David Lowe	1999	• Describes the local image content or signature to a local area like circle, corner blob etc. • Invariant to scale, rotation, Illumination, viewpoints etc. • Not a good choice for real-time object detection	Object recognition, face recognition, navigation, image stitching, object recognition, 3D modelling, fingerprint recognition, video tracking and many more
Viola-Jones Detector	Viola and Jones	2001	• Mainly used for face detection • learn the structural features of the face such as nose, eyes, lips and detect the face in four stages -Haar Feature Selection, Creating an Integral Image, Adaboost Training and Cascading Classifiers.	face detection such as the password to unlock the device etc.

HOG	Dalal and Triggs	2005	• HoG descriptor counts the occurrences of gradient orientation (direction of intensity change) in localized portions of an image • Works on a grayscale image	Widely used in computer vision tasks such as biometrics, automatic target detection, and recognition etc.
SURF	H. Bay et al.	2006	• Fast and performant interest point detection description model • Contains three major stages- interest point detection (such as corners, blobs), Feature description (local neighbourhood description) and feature matching (between different images)	Object recognition, classification, image registration, 3D reconstruction, Medical X-Ray Images etc
DPM	P. Fezenszwalb	2008	• Extension of HOG model • Uses a star-structured part-based model developed for embedded vision applications	Self-driving cars, Biomedical imaging, autonomous vehicles etc.
Neural network-based model				
Overfeat	Zhang et al.	2014	• multiscale, sliding window approach for localization, classification and detection • similar to AlexNet • On the ILSVRC 2013 dataset, it achieved 4th rank in classification and 1st rank in localization and detection tasks	Object detection, object tracking, Image Recognition etc.
R-CNN	Ross Girshick and et al	2014	• The two-stage detector uses Selective Search to extract regions of interest • Extracts 2000 regions from the input image (Region proposals) • Produces a 4096-dimensional feature vector as output • Approx. 49 (including region proposal) seconds for each test image, that's why can't implement for real-time object detection	Surveillance systems, facial recognition, tracking objects, drone-mounted camera, object detection in Google Lens, CCTV surveillance in security etc
Fast R-CNN	Ross Girshick and et al.	2015	• Reduces the computation time by running the neural network once on the whole image • Test time 2.3 seconds (including region proposal) per image.	
Faster R-CNN	Shaoqing Ren and et al.	2015	• Uses almost cost-free region proposals namely RPN (Region Proposal Network) • Integrates the ROI generation into the neural network itself. • Test time 0.2 seconds (including region proposal) per image. • Can be used for real-time object detection	
Mask R-CNN	Kaiming He and et al.	2017	• Used for pixel-wise predictions • Extended Faster R-CNN to solve instance segmentation problems. • Uses a simple, quantization free layer, called RoIAlign	
AttentionNet	Yoo et al.	2015	• Uses top-down approach • Not scalable to multiple classes. However, has a potential for extension to generic object classes	Perform well in human detection tasks
Object detection via multi-region and semantic segmentation aware CNN model	Gidaris et al.	2015	• Rich CNN-based model, relies on a multi-region deep CNN and Segmentation aware[28] • On PASCAL VOC2007 achieve mAP of 78.2% and PASCAL VOC2012 achieved 73.9% mAP	Can use for Real-time object detection, object tracking, surveillance etc.
DSOD	Shen et al.	2015	• multi-scale Proposal-free detector • similar to SSD • Transition Without Pooling • Uses deep supervised concept to tackle the problem of vanishing gradient	Detection performed well in low-end devices as well.
SPPNet	He et al	2015	• Extract the feature maps from the entire image only once • Extracts window-wise features from regions of the feature maps • Feature extraction applicable in any windows from CNN feature maps	Vehicle Pedestrian Detection and many more

YOLOv1	Joseph Redmon et al.	2015	• One-stage detection model • A single neural network only in one evolution, predict the bounding boxes (anchor box) and class probabilities directly from full images • Realtime object detection speed achieved 45 fps but was not accurate enough • Limitations- can predict up to 49 objects, struggles to detect small objects, low recall and more localization error. • Input image divided into 7x7 grid (in the paper [10])	Traffic signals, Autonomous driving, Wildlife (identify animals in videos), person detection, parking meters, vehicle detection in smart cities etc
YOLOv2	Redmon and Farhadi	2017	• Improved recall and localization in the previous version • Can detect more than 9000 object categories • Uses DarkNet as feature extractors • YOLOv2 uses Batch Normalization, High-Resolution Classifier and uses anchor boxes to predict bounding boxes and more.	
YOLOv3	Redmon and Farhadi	2018	• Can detect more smaller objects as compare to YOLOv2 • Uses robust feature extractor called darknet-53 model	
YOLOv4	Bochkovskiy et al.	2020	• Improved speed and accuracy • Uses CSPDarknet53 which reduces the training time • On MSCOCO dataset at the real-time speed of approx. 65FPS on NVIDIA Tesla V100	
SSD	Wei Liu et al.	2015	• More accurate as region-based technique (R-CNN) as compare to the YOLO family • Uses multiple feature map which predicts different scale and aspects radios • On the VOC2007 test, SSD achieves 59FPS	
DSSD	Fu et al.	2017	• Combine the classifier ResNet-101 with SSD • Achieved 81.5% mAP on VOC2007 with 513 ×513 input and 80.0% mAP on • VOC2012	
MS-CNN	Cai et al.	2016	• MS-CNN consists of a detection sub-network and a proposal sub-network. • Achieved detection rates up to 15 fps.	
RetinaNET	Tsung-Yi Lin and et al.	2018	• One-stage object detection models which work well with dense and small-scale objects • Addresses the extreme foreground-background class imbalance problem with the Focal Loss model • Can have approx. 100k boxes with the resolve of class imbalance problem using focal loss • Achieves state-of-the-art accuracy and speed	
RefineDet	Zhang et al.	2018	• High accuracy and speed • Constructed by two inter-connected modules -ARM and ODM	
Hypernet	Kong et al.	2016	• Aimed to tackle region proposal generation and object detection jointly[30] • On PASCAL VOC 2007 test set, HyperNet (mAP of 76.3%,) leads 6.3 points higher than Fast R-CNN (70.0%) and 3.1 points higher than Faster R-CNN (73.2%).	Realtime object detection, Surveillance etc.
CornerNet	Law et al.	2018	• Detect an object bounding box (anchor box) as a pair of key points or pairs of corners using a single convolution neural network • Eliminate the need for designing a set of anchor boxes • achieved 42.1% AP on MS COCO	Object Detection in Surveillance, Medical, Daily Life and many more
EfficientDet	M. Tang et al.	2020	• Uses weighted bi-directional feature pyramid network (BiFPN) for fast multiscale feature fusion • Uses compound scaling method that uniformly scales the resolution, width, and depth for all backbone • Achieved the highest accuracy with fewest training epochs	High accuracy object detection method and can practically be useful for many real-world applications such as underwater imagery object detection, Road Segmentation in satellite imagery, Wheat detection in the field, Trash Detection and many more
G-RCNN	Pramanik et al.	2021	• Improved version of Fast RCNN and Faster RCNN	Multi-object detection and tracking in videos

4. Conclusion

Object detection is a key ability of computer vision systems and the current research and technologies making great progress in many directions. In this paper, many research papers have been reviewed, found that lots of excellent work have been done in the field of object detection and deep learning (neural network). technology growing rapidly but still some obstacles to robust object detection, few of them solved but not up to the mark is the finding of this paper, such as real-time object detection with high accuracy and efficiency remains challenging, viewing the object from different angles may look completely different which is also a challenge knows as Viewpoint variation, deformation object detection challenge in which object (like, human body) can change the shape and difficult to detect correctly, Occlusion in which objects are overlapped or unclear and difficult to identify correctly, efficiency or speed, computational power and many more. So that still we have much room for improvement in this area.

References

[1] D.G. Lowe, Object recognition from local scale-invariant features, in: Proc. Seventh IEEE Int. Conf. Comput. Vis., 1999: pp. 1150–1157 vol.2. https://doi.org/10.1109/ICCV.1999.790410.

[2] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR 2001, 2001: p. I–I. https://doi.org/10.1109/CVPR.2001.990517.

[3] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05, 2005: pp. 886–893 vol. 1. https://doi.org/10.1109/CVPR.2005.177.

[4] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded Up Robust Features, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Comput. Vis. – ECCV 2006, Springer, Berlin, Heidelberg, 2006: pp. 404–417. https://doi.org/10.1007/11744023_32.

[5] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-Based Convolutional Networks for Accurate Object Detection and Segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 142–158. https://doi.org/10.1109/TPAMI.2015.2437384.

[6] R. Girshick, Fast R-CNN, in: 2015: pp. 1440–1448. https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html (accessed September 14, 2021).

[7] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, in: Adv. Neural Inf. Process. Syst., Curran Associates, Inc., 2015. https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html (accessed September 14, 2021).

[8] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, in: 2017: pp. 2961–2969. https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html (accessed September 14, 2021).

[9] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, in: 2016: pp. 779–788. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html (accessed September 14, 2021).

[10] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, in: 2017: pp. 7263–7271. https://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html (accessed September 14, 2021).

[11] J. Redmon, A. Farhadi, YOLOv3: An Incremental Improvement, ArXiv180402767 Cs. (2018). http://arxiv.org/abs/1804.02767 (accessed September 14, 2021).

[12] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, ArXiv200410934 Cs Eess. (2020). http://arxiv.org/abs/2004.10934 (accessed September 14, 2021).

[13] ultralytics/yolov5, Ultralytics, 2021. https://github.com/ultralytics/yolov5 (accessed September 14, 2021).

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single Shot MultiBox Detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Comput. Vis. – ECCV 2016, Springer International Publishing, Cham, 2016: pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, in: 2017: pp. 2980–2988. https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html (accessed September 14, 2021).

[16] S. Zhang, L. Wen, X. Bian, Z. Lei, S.Z. Li, Single-Shot Refinement Neural Network for Object Detection, in: 2018: pp. 4203–4212. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Single-Shot_Refinement_Neural_CVPR_2018_paper.html (accessed September 14, 2021).

[17] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, Cascade object detection with deformable part models, in: 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2010: pp. 2241–2248. https://doi.org/10.1109/CVPR.2010.5539906.

[18] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, J. Shi, FoveaBox: Beyound Anchor-Based Object Detection, IEEE Trans. Image Process. 29 (2020) 7389–7398. https://doi.org/10.1109/TIP.2020.3002345.

[19] J. Guo, J. Wang, R. Bai, Y. Zhang, Y. Li, A New Moving Object Detection Method Based on Frame-difference and Background Subtraction, IOP Conf. Ser. Mater. Sci. Eng. 242 (2017) 012115. https://doi.org/10.1088/1757-899X/242/1/012115.

[20] F. Particke, R. Kolbenschlag, M. Hiller, L. Patiño-Studencki, J. Thielecke, Deep Learning for Real-Time Capable Object Detection and Localization on Mobile Platforms, IOP Conf. Ser. Mater. Sci. Eng. 261 (2017) 012005. https://doi.org/10.1088/1757-899X/261/1/012005.
[21] D. Lin, X. Shen, C. Lu, J. Jia, Deep LAC: Deep Localization, Alignment and Classification for Fine-Grained Recognition, in: 2015: pp. 1666–1674. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Lin_Deep_LAC_Deep_2015_CVPR_paper.html (accessed September 14, 2021).
[22] T. Malisiewicz, A. Gupta, A.A. Efros, Ensemble of exemplar-SVMs for object detection and beyond, in: 2011 Int. Conf. Comput. Vis., 2011: pp. 89–96. https://doi.org/10.1109/ICCV.2011.6126229.
[23] X.-C. Yin, X. Yin, K. Huang, H.-W. Hao, Robust Text Detection in Natural Scene Images, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 970–983. https://doi.org/10.1109/TPAMI.2013.182.
[24] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks, ArXiv13126229 Cs. (2014). http://arxiv.org/abs/1312.6229 (accessed September 14, 2021).
[25] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov, Scalable Object Detection using Deep Neural Networks, in: 2014: pp. 2147–2154. https://openaccess.thecvf.com/content_cvpr_2014/html/Erhan_Scalable_Object_Detection_2014_CVPR_paper.html (accessed September 14, 2021).
[26] K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.
[27] D. Yoo, S. Park, J.-Y. Lee, A.S. Paek, I. So Kweon, AttentionNet: Aggregating Weak Directions for Accurate Object Detection, in: 2015: pp. 2659–2667. https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Yoo_AttentionNet_Aggregating_Weak_ICCV_2015_paper.html (accessed September 14, 2021).
[28] S. Gidaris, N. Komodakis, Object Detection via a Multi-Region and Semantic Segmentation-Aware CNN Model, in: 2015: pp. 1134–1142. https://openaccess.thecvf.com/content_iccv_2015/html/Gidaris_Object_Detection_via_ICCV_2015_paper.html (accessed September 14, 2021).
[29] A. Ghodrati, A. Diba, M. Pedersoli, T. Tuytelaars, L. Van Gool, DeepProposal: Hunting Objects by Cascading Deep Convolutional Layers, in: 2015: pp. 2578–2586. https://openaccess.thecvf.com/content_iccv_2015/html/Ghodrati_DeepProposal_Hunting_Objects_ICCV_2015_paper.html (accessed September 14, 2021).
[30] T. Kong, A. Yao, Y. Chen, F. Sun, HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection, in: 2016: pp. 845–853. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Kong_HyperNet_Towards_Accurate_CVPR_2016_paper.html (accessed September 14, 2021).
[31] Z. Cai, Q. Fan, R.S. Feris, N. Vasconcelos, A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Comput. Vis. – ECCV 2016, Springer International Publishing, Cham, 2016: pp. 354–370. https://doi.org/10.1007/978-3-319-46493-0_22.
[32] T-CNN: Tubelets With Convolutional Neural Networks for Object Detection From Videos, (n.d.). https://ieeexplore.ieee.org/abstract/document/8003302/ (accessed September 14, 2021).
[33] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, X. Xue, DSOD: Learning Deeply Supervised Object Detectors From Scratch, in: 2017: pp. 1919–1927. https://openaccess.thecvf.com/content_iccv_2017/html/Shen_DSOD_Learning_Deeply_ICCV_2017_paper.html (accessed September 14, 2021).
[34] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, A.C. Berg, DSSD : Deconvolutional Single Shot Detector, ArXiv170106659 Cs. (2017). http://arxiv.org/abs/1701.06659 (accessed September 14, 2021).
[35] T. Kong, F. Sun, A. Yao, H. Liu, M. Lu, Y. Chen, RON: Reverse Connection With Objectness Prior Networks for Object Detection, in: 2017: pp. 5936–5944. https://openaccess.thecvf.com/content_cvpr_2017/html/Kong_RON_Reverse_Connection_CVPR_2017_paper.html (accessed September 14, 2021).
[36] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable Convolutional Networks, in: 2017: pp. 764–773. https://openaccess.thecvf.com/content_iccv_2017/html/Dai_Deformable_Convolutional_Networks_ICCV_2017_paper.html (accessed September 14, 2021).
[37] L. Tychsen-Smith, L. Petersson, DeNet: Scalable Real-Time Object Detection With Directed Sparse Sampling, in: 2017: pp. 428–436. https://openaccess.thecvf.com/content_iccv_2017/html/Tychsen-Smith_DeNet_Scalable_Real-Time_ICCV_2017_paper.html (accessed September 14, 2021).
[38] P. Zhou, B. Ni, C. Geng, J. Hu, Y. Xu, Scale-Transferrable Object Detection, in: 2018: pp. 528–537. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhou_Scale-Transferrable_Object_Detection_CVPR_2018_paper.html (accessed September 14, 2021).
[39] H. Hu, J. Gu, Z. Zhang, J. Dai, Y. Wei, Relation Networks for Object Detection, in: 2018: pp. 3588–3597. https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Relation_Networks_for_CVPR_2018_paper.html (accessed September 14, 2021).
[40] H. Law, J. Deng, CornerNet: Detecting Objects as Paired Keypoints, in: 2018: pp. 734–750. https://openaccess.thecvf.com/content_ECCV_2018/html/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.html (accessed September 14, 2021).

[41] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, D. Lin, Libra R-CNN: Towards Balanced Learning for Object Detection, in: 2019: pp. 821–830. https://openaccess.thecvf.com/content_CVPR_2019/html/Pang_Libra_R-CNN_Towards_Balanced_Learning_for_Object_Detection_CVPR_2019_paper.html (accessed September 14, 2021).
[42] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, C.C. Loy, D. Lin, Hybrid Task Cascade for Instance Segmentation, in: 2019: pp. 4974–4983. https://openaccess.thecvf.com/content_CVPR_2019/html/Chen_Hybrid_Task_Cascade_for_Instance_Segmentation_CVPR_2019_paper.html (accessed September 14, 2021).
[43] Y. Li, Y. Chen, N. Wang, Z. Zhang, Scale-Aware Trident Networks for Object Detection, in: 2019: pp. 6054–6063. https://openaccess.thecvf.com/content_ICCV_2019/html/Li_Scale-Aware_Trident_Networks_for_Object_Detection_ICCV_2019_paper.html (accessed September 14, 2021).
[44] Multi-scale Positive Sample Refinement for Few-Shot Object Detection | SpringerLink, (n.d.). https://link.springer.com/chapter/10.1007/978-3-030-58517-4_27 (accessed September 14, 2021).
[45] M. Tan, R. Pang, Q.V. Le, EfficientDet: Scalable and Efficient Object Detection, in: 2020: pp. 10781–10790. https://openaccess.thecvf.com/content_CVPR_2020/html/Tan_EfficientDet_Scalable_and_Efficient_Object_Detection_CVPR_2020_paper.html (accessed September 14, 2021).
[46] H. Zhang, H. Chang, B. Ma, N. Wang, X. Chen, Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Comput. Vis. – ECCV 2020, Springer International Publishing, Cham, 2020: pp. 260–275. https://doi.org/10.1007/978-3-030-58555-6_16.
[47] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, C. Wang, P. Luo, Sparse R-CNN: End-to-End Object Detection With Learnable Proposals, in: 2021: pp. 14454–14463. https://openaccess.thecvf.com/content/CVPR2021/html/Sun_Sparse_R-CNN_End-to-End_Object_Detection_With_Learnable_Proposals_CVPR_2021_paper.html (accessed September 14, 2021).
[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper With Convolutions, in: 2015: pp. 1–9. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html (accessed September 14, 2021).
[49] A. Pramanik, S.K. Pal, J. Maiti, P. Mitra, Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking, IEEE Trans. Emerg. Top. Comput. Intell. (2021) 1–11. https://doi.org/10.1109/TETCI.2020.3041019.

Authors' Profiles

Diwakar, received his bachelor's degree in Computer Science and application from University of Allahabad, India and master's degree from J.K Institute of applied physics and technology, Allahabad University. He is currently pursing his Ph.D. from Babasaheb Bhimrao Ambedkar University Lucknow, India. His area of research interests include Computer vision, Digital Image processing, Deep learning, real time Object detection, Neural Networks, Artificial Intelligence.

Dr. Deepa Raj, Working as associate professor in the Department of Computer Science at Babasaheb Bhimrao Ambedkar University Lucknow (A Central University). She did her post-Graduation from J.K Institute of applied physics and technology, Allahabad University and Ph.D. from Babasaheb Bhimrao Ambedkar University Lucknow in the field of software Engineering. Her field of interest is Software Engineering, Digital Image Processing, Computer Graphics and Algorithm analysis. She has attended lots of National and International conference and numbers of research papers published in her field.

How to cite this paper: Diwakar, Deepa Raj, " Recent Object Detection Techniques: A Survey", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.2, pp. 47-60, 2022.DOI: 10.5815/ijigsp.2022.02.05