Several Methods for Object Detection System Using Deep Learning: A Review

Gurpreet Singh Panesar
Assistant Professor, Chandigarh University, Gharuan, Punjab (India).

Abstract: Computer vision is excelling in the field of segmentation, feature extraction, and object detection from image data. The object detection is gaining immense interest from a different application such as healthcare, traffic monitoring, surveillance, robotics etc. The ability to detect the object more precisely is an important factor due to its application in sensitive domains. Over the past few years, researchers have strived to cope up with this challenge. Due to significant development in neural networks especially deep learning, these visual recognition systems have attained remarkable performance. Object detection is one of these domains witnessing great success in computer vision. This study presents a review of object detection approach considered using Convolutional Neural Network (CNN), Fast R-CNN, and RCNN model. The CNN is used in all three methods (salient, objectness, and category-specific) of object detection. Deep learning frameworks and the platforms that are popular for the object detection task are also reviewed.

Keywords: Object Detection System, CNN, Computer-vision, Deep learning models using Image Classification.

I. INTRODUCTION

Recently, computer vision has been extensively researched in the area of object detection for industrial automation, consumer electronics, medical imaging, military, and video surveillance. It is predicted that the computer vision market will be worth $50 billion by the end of 2020[1]. People look at a picture and immediately realize what objects are inside the picture, where they are, and how they are associated. On the other hand, if calculations for picture preparing could be exact and quick enough, the PCs would most likely perform autonomous driving without particular sensors, and assistive gadgets would probably pass on continuous scene data to clients. Similarly, if these calculations could provide Deep Learning errands with high proficiency and superb execution like individuals do then only it would be recognized as genuine artificial intelligence. Hence, the main objectives of picture preparing are the classification of objects, localization of objects, and detection of objects; and the key difficulties are precision, execution time, processing speed, and financial feasibility [2].

Object detection is a characteristic expansion of the classification issue. Given a picture and a lot of object classes, the objective of object detection is to decide if the picture contains any objects of the predetermined classes, as well to show where in the picture these objects are found. While in the classification of objects, it just concerns the nearness or non-nearness of these objects, and to semantic division, that looks to characterize singular pixels as being or not being a piece of an object of the predetermined classes. Classification of objects [3] is henceforth a subtask of object detection, which thus is a subtask of the semantic fragmentation. The process of classifying objects recognizes the likelihood of an object in a picture while localization of objects implies recognizing the area of an object in the picture. The algorithms used for the localization of the objects provide the coordinates of the area of an object regarding the picture.

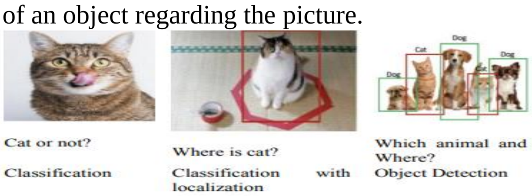

Fig 1. The normal CV (Computer Vision Tasks) [4]

The investigators had searched several areas of OD (Object Detection) such as;

- Face Detection System
- Text Detection System
- Pedestrian Detection System
- Vehicle or Number Plate Detection System
- Surveillance System for verification, and medical IA (Image Analysis) etc.

The problem definition of object detection is to determine where objects are located in a given image (object localization) and which category each object belongs to (object classification). So the pipeline of traditional object detection models can be mainly divided into three stages:

- Informative region selection
- Feature extraction and
- Classification.

A. Information Region Selection [3]: As different objects may appear in any positions of the image and have different aspect ratios or sizes, it is a natural choice to scan the whole image with a multi-scale sliding window. Although this exhaustive strategy can find out all possible positions of the objects, its shortcomings are also obvious. Due to a large number of candidate windows, it is computationally expensive and produces too many redundant windows. However, if only a fixed number of sliding window templates is applied, unsatisfactory regions may be produced.

B. Feature Extraction: To recognize different objects, we need to extract visual features which can provide a semantic and robust representation. SIFT [5], HOG [6] and Haar-like [7] features are the representative ones.

C. Classification: Besides, a classifier is needed to distinguish a target object from all the other categories and to make the representations more hierarchical, semantic and informative for visual recognition. Usually, the Supported Vector Machine (SVM)[8], CNN[9], and DNN[10].

The deep learning boosted the growth of computer vision and produced state-of-the-art results in image recognition, feature extraction, and object detection. The deep learning requires a significant amount of dataset to train and a high computation power. The graphical processing units fulfil the requirement of computer vision to process efficiently. In the field of computer vision, object detection is an important task. The object detection is a process in which the instances of the objects are detected for a particular class in an image. The object detection is trending due to its applicability in a broader area [11].

II. LITERATURE REVIEW

Athira M V et al., 2020 [12] surveyed of several works developed so far in the field of image classification and object detection and a relative study of different methods. Survey is divided in three sub areas as Machine Learning based approach, Deep Learning based approach and object detection for night vision applications. ***Chinthakindi Balaram Murthy et al., 2020 [1]*** detailed survey on recent advancements and achievements in object detection using various deep learning techniques. Several topics have been included, such as Viola–Jones (VJ), histogram of oriented gradient (HOG), one-shot and two-shot detectors, benchmark datasets, evaluation metrics, speed-up techniques, and current state-of-art object detectors. Detailed discussions on some important applications in object detection areas, including pedestrian detection, crowd detection, and real-time object detection on Gpu-based embedded systems have been presented.

Zhong-Qiu Zhao et al., 2019 [3] provided a review of deep learning-based object detection frameworks. The review begins with a brief introduction on the history of deep learning and its representative tool, namely, the convolutional neural network. Then, we focus on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance further. As distinct specific detection tasks exhibit different characteristics, we also briefly survey several specific tasks, including salient object detection, face detection, and pedestrian detection. Experimental analyses are also provided to compare various methods and draw some meaningful conclusions. Finally, several promising directions and tasks are provided to serve as guidelines for future work in both object detection and relevant neural network-based learning systems. ***Usha Mittal et al., 2019 [4]*** discussed the human brain takes less than a minute to identify the location of object inside the image as well as recognize it as soon as it sees to it; but machine needs time and large amount of data to do the same task. Deep neural network based on convolution neural network gives high accuracy and great results in object detection and classification. To train deep neural networks, large amount of data such as (images and videos) and time is required. As computational cost of computer vision is very high, transfer learning technique, where a model trained on one task is reused on another related task, gives better results. Authors have proposed various deep learning based algorithms for object detection and classification like Region based Convolutional neural network, Fast Region based Convolutional neural network, and CNN (Convolutional Neural Network).

III. THE BASIC ARCHITECTURE OF OBJECT DETECTION SYSTEM

The intent is to figure out real-world object instances like cats, bicycles, telephones, various flowers, and humans in real-time videos or still images. It paves the way for object recognition, localization, and detection of single/multiple objects within a video frame or an image with a much better interpretation of an image as a whole [13]. Difficult challenges such as occlusion and irregular lighting conditions should be handled carefully while performing object detection.

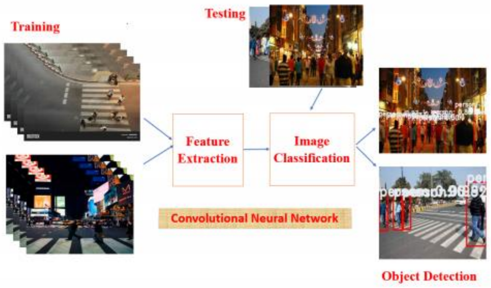

Fig 2. The Basic Block Diagram of OD (Object Detection) System [13]

Fig 2 shows the basic block diagram of object detection. The application of object detection covers wide areas, such as medical imaging, security, video surveillance, self-driving vehicles, robot vision, and facial recognition. Fig 3 shows various approaches available in object detection.

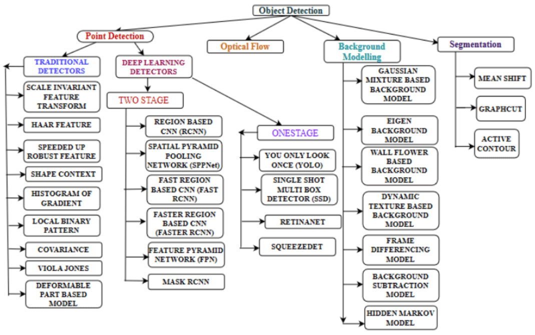
Object Detection
Point Detection
Optical Flow
Background Modelling
Segmentation
TRADITIONAL DETECTORS
DEEP LEARNING DETECTORS
SCALE INVARIANT FEATURE TRANSFORM
HAAR FEATURE
SPEEDED UP ROBUST FEATURE
SHAPE CONTEXT
HISTOGRAM OF GRADIENT
LOCAL BINARY PATTERN
COVARIANCE
VIOLA JONES
DEFORMABLE PART BASED MODEL
TWO STAGE
REGION BASED CNN (RCNN)
SPATIAL PYRAMID POOLING NETWORK (SPPNet)
FAST REGION BASED CNN (FAST RCNN)
FASTER REGION BASED CNN (FASTER RCNN)
FEATURE PYRAMID NETWORK (FPN)
MASK RCNN
ONESTAGE
YOU ONLY LOOK ONCE (YOLO)
SINGLE SHOT MULTI BOX DETECTOR (SSD)
RETINANET
SQUEEZEDET
GAUSSIAN MIXTURE BASED BACKGROUND MODEL
EIGEN BACKGROUND MODEL
WALL FLOWER BASED BACKGROUND MODEL
DYNAMIC TEXTURE BASED BACKGROUND MODEL
FRAME DIFFERENCING MODEL
BACKGROUND SUBTRACTION MODEL
HIDDEN MARKOV MODEL
MEAN SHIFT
GRAPHCUT
ACTIVE CONTOUR

Fig 3. Several Methods of OD (Object Detection) [1]

There are multiple ways to detect objects and these are done using the Viola–Jones (VJ) object detector [14-15], the feature-based object detector [5-16], HOG features using a support vector machine (SVM) classification object detector [17], and object detection-based deep learning techniques. Fig 3 shows various approaches available in object detection.

IV. OBJECT DETECTION METHODS USING DEEP LEARNING

There are some machine learning algorithms which don't utilize deep learning but all deep learning algorithms come under machine learning algorithms. Estimation models supported by Deep learning are made out of various hidden layers which help in learning data representation along with consideration at each level. Deep learning basically deals with the deep neural network algorithms where deep refers to the number of hidden layers and its main objective is to resolve the learning problems by copying the functioning of the human brain. Deep learning utilized by the system has been always improving, notwithstanding the adjustments in the system structure, the more is to do some tune dependent on the first system or apply some trap to make the system execution to upgrade. The various algorithms used for object detection and classifications are as follows:

- R-CNN
- Fast RCNN
- Contour Based Method
- CNN Model

A. R-CNN Model: Region-based Convolutional network firstly perform region search as the name suggest then classification is done. Region search is a process to locate an object in an image. One of the methods for region search is the selective search but later in 2012, an alternate method was developed [18] which are known as the exhaustive search method. It begins by computing smaller parts or regions in an input image and after that, it clubs them in a hierarchical structure. Hence the entire image is contained in the final group or hierarchical structure. There are two factors that are color space and similarity metrics which are considered when the detected regions are grouped. Thus, the final image is formed by grouping smaller regions which result in a certain number of region proposals.

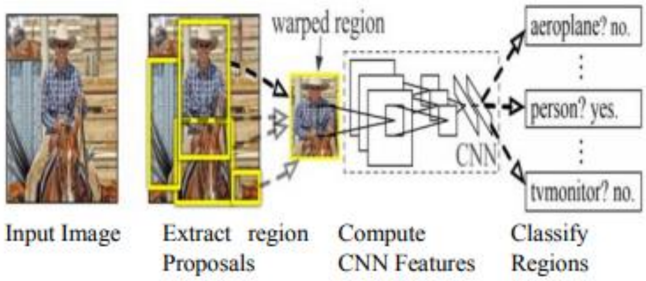

Fig 4. R-CNN Framework Model [18]

In R-CNN, region proposals are detected using a selective search approach and then deep learning is applied for the detection of objects in those detected region proposals. Then CNN is used so to match the input size of the CNN individual region proposals are resized which helped in extracting the feature vectors of 4096-dimensions. To predict the probability for each class various classifiers are used which take these feature vectors as an input. Then the probability for detecting the objects using these feature vectors for individual classes which contain pretrained support vector machine (SVM) is predicted. To minimize the error in object localization, linear regression can be used in region proposal which modifies the shapes and sizes of the bounding boxes.

B. Fast R-CNN Model: Fast Region-based Convolutional Network (Fast R-CNN) was developed [19] in 2015, this is similar to RCNN in some aspects but its main objective is to minimize the time consumption required for evaluating each region proposals that are related to a large number of models. In fast R-CNN, the entire image is considered as an input for the CNN which utilizes many Convolutional layers as compared to the R-CNN where each region proposal requires CNN. The selective search approach is implemented on the feature maps generated by CNN to detect the region of interests (RoI). To get the accurate region of interest along with the length and breadth as main criteria, the RoI pooling layer should be used to reduce the size of the feature maps. The output from individual RoI layer is given as an input to the fully-connected layers which produces a feature vector as an output. Then a softmax classifier along with these feature vectors is used to detect the objects as well as uses a linear regression to modify the localization of objects.

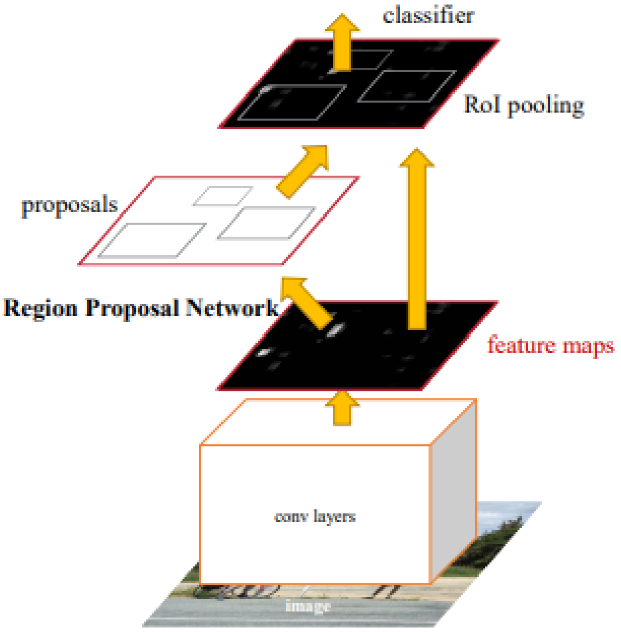

Fig 5. Fast R-CNN Framework [20]

C. Contour Based Method [21]: contour extraction has to cope with inherent problems originating mainly from changing lighting conditions and environment texturing when applied to real images. Using depth images avoids these problems but introduces other challenges, e.g., the reliable generation of range images. The segmentation of objects in natural environments is simplified by using range images, since they are distinguishable if the spatial distance between object and background is sufficiently high. The contour extraction of objects from range images is relatively simple because scan points belonging to the same object show smooth changes in their distance values. At object borders, discontinuities emerge that cause an edge in the range image that in turn is identified by segmentation algorithms. Nevertheless, all standing objects cannot easily be separated from floor. It overcomes this problem by introducing a novel technique to identify the ground based on local gradients. Contour or silhouette based object recognition in range images are view dependent. Hence, view invariant recognition is reached by generating several object views from the original object and training different classifiers.

D. CNN Model [22]: Convolutional neural network (CNN) is a class of deep, feed-forward artificial neural network that has been utilized to produce an accurate performance in computer vision tasks, such as image classification and detection. CNNs are like traditional neural network, but with deeper layers. It has weights, biases and outputs through a nonlinear activation. The neurons of the CNN are arranged in a volumetric fashion such as, height, width and depth. CNN architecture, it is composed of convolutional layer, pooling layer and fully connected layer. Convolutional layer and pooling layer are typically alternated and the depth of each filter increases from left to right while the output size (height and width) are decreasing. The fully connected layer is the last stage which is similar to the last layer of the conventional neural networks.

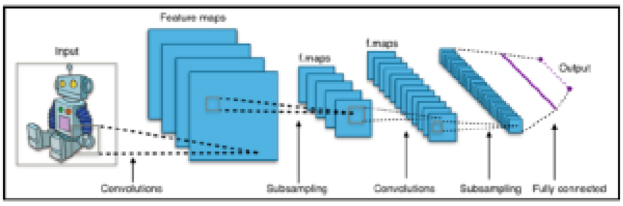

Fig 6. CNN Framework [22]

Table 1 discussed the several deep learning methods are used in Object detection system. It shows the metrics, demerits and uses of the deep learning methods. Table 2 described the summary review of various articles studied in deep learning concepts and datasets etc.

TABLE 1. COMPARISON ANALYSIS [4]

Methods	Merits	Demerits	Uses
Region-CNN [18]	High Detection Quality	Large number of data, processing and energy etc	Large Data set used
Fast RCNN+Vgg16 [19]	Quality and speed detection are enhanced	Highly cost	Quality and speed is a major concern as it results the R-CNN model.
Faster R-CNN Vgg [19]	Region proposal network	Speed is a major issue	It is faster than R-CNN
YOLO [4]	It is faster outcomes DPM and R-CNN Model	Maximum localization errors	It may be used for Detection of large objects

Table 2. Summary of the Review

Datasets	Methods	Metrics	Score Value
PASCAL VOC 2007 [23]	FR-CNN +VGG+ZF	mAP	73.5 per cent
PASCAL VOC+MS COCO [24]	Scale Transfer dense net	mAP	81 per cent
PASCAL VOC 2007 +2011 , MS COCO 2017 [25]	Hierarchical Object	mAP	78.6 per cent
PASCAL VOC 2007[26]	Region Proposed Network and FR-CNN	mAP	58.7 per cent
PASCAL VOC 2007[27]	Deep Object representation and Local Linear Regression	mAP	81 per cent

V. CONCLUSION AND FUTURE SCOPE

Object detection is considered as foremost step in deployment of self-driving cars and robotics. In this paper, we demystified the role of deep learning techniques based on CNN for object detection. Deep learning frameworks and services available for object detection are also discussed in the paper. Benchmarked datasets for object localization and detection released in worldwide competitions are also covered. The pointer to the domains in which object detection is applicable has been discussed. State-of-the-art deep learning based object detection techniques have been assessed. CNN is designed to process the image and any data modalities that can be represented in the form of multiple arrays. CNN and its extensions such as R-CNN and Faster R-CNN are applied with great success to detection, segmentation, and recognition of objects and regions in images. In the reviewed literature, the use of CNN is reflected as a core function, and many proposed models have their version of filters attached to the convolutional network. From the review summary mentioned above in most of the proposed the datasets used be the benchmark.

Future directions can be stated as follows. Due to infeasibility of humans to process large surveillance data, there is a need to bring data closer to the sensor where data are generated. This would result into real time detection of objects. Currently, object detection systems are small in size having 1-20 nodes of clusters having GPUs. These systems should be extended to cope with real time full motion video generating frames at 30 to 60 per second. Such object detection analytics should be integrated with other tools using data fusion.

VI. REFERENCES

[1] Murthy, C. B., Hashmi, M. F., Bokde, N. D., & Geem, Z. W. (2020). Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review. *Applied Sciences*, *10*(9), 3280.

[2] Du, J. (2018, April). Understanding of object detection based on CNN family and YOLO. In *Journal of Physics: Conference Series* (Vol. 1004, No. 1, p. 012029). IOP Publishing.

[3] Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, *30*(11), 3212-3232.

[4] Mittal, U., Srivastava, S., & Chawla, P. (2019, June). Review of different techniques for object detection using deep learning. In *Proceedings of the Third International Conference on Advanced Informatics for Computing Research* (pp. 1-8).

[5] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

[6] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

[7] Lienhart, R., & Maydt, J. (2002, September). An extended set of haar-like features for rapid object detection. In *Proceedings. international conference on image processing* (Vol. 1, pp. I-I). IEEE.

[8] Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine learning*, *20*(3), 273-297.

[9] Cai, Z., Saberian, M., & Vasconcelos, N. (2015). Learning complexity-aware cascades for deep pedestrian detection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3361-3369).

[10] Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). Scalable object detection using deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2147-2154).

[11] Kamate, S., & Yilmazer, N. (2015). Application of object detection and tracking techniques for unmanned aerial vehicles. *Procedia Computer Science*, *61*, 436-441.

[12] Athira, M. V., & Khan, D. M. (2020, July). Recent Trends on Object Detection and Image Classification: A Review. In *2020 International Conference on Computational Performance Evaluation (ComPE)* (pp. 427-435). IEEE.

[13] Jalled, F., & Voronkov, I. (2016). Object detection using image processing. *arXiv preprint arXiv:1611.07791*.

[14] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE.

[15] Viola, P., & Jones, M. (2001, July). Robust real-time face detection. In *null* (p. 747). IEEE.

[16] Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision* (Vol. 2, pp. 1150-1157). IEEE.

[17] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

[18] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *38*(1), 142-158.

[19] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

[20] *Faster R-CNN | ML - GeeksforGeeks*. GeeksforGeeks. (2020). Retrieved 5 December 2020, from https://www.geeksforgeeks.org/faster-r-cnn-ml/.

[21] Stiene, S., Lingemann, K., Nuchter, A., & Hertzberg, J. (2006, June). Contour-based object detection in range images. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (pp. 168-175). IEEE.

[22] Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R. R. P., & Maningo, J. M. Z. (2018, October). Object detection using convolutional neural networks. In *TENCON 2018-2018 IEEE Region 10 Conference* (pp. 2023-2027). IEEE.

[23] Chu, W., & Cai, D. (2018). Deep feature based contextual model for object detection. *Neurocomputing*, *275*, 1035-1042.

[24] Zhou, P., Ni, B., Geng, C., Hu, J., & Xu, Y. (2018). Scale-transferrable object detection. In *proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 528-537).

[25] Wang, J., Tao, X., Xu, M., Duan, Y., & Lu, J. (2018). Hierarchical objectness network for region proposal generation and object detection. *Pattern Recognition*, *83*, 260-272.

[26] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

[27] Wu, Q., Li, H., Meng, F., Ngan, K. N., & Xu, L. (2017, July). Blind proposal quality assessment via deep objectness representation and local linear regression. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1482-1487). IEEE.