Various Deep Learning Methods Used in Object Detection System: Review

Shainy Bakshi[1] , Neha Bathla[2]
Department of Computer Science Engineering
Yamuna Institute of Engineering and Technology, Yamunanagar, Haryana.

Abstract- Recent years, object tracking and detection is one of the famous fields of research due to regular alter in moving of objects and changes in scene size, arrival and illumincation variations, occlusions, ego motion etc. Normally, the selection of features is the important role in object tracking. Object tracking is regared to various real-time uses such as video survillence, smile detecton, tracking of the cars and car numbers etc. To resolve the detection problem, appearance, tracking regarded to moving objects. The deep learning method main focus on the detection, localization and tracking method to filtered the video series. Additionally, a few approaches use the previous define data about the object color, text and shape etc. Detection, tracking and classification method which associates above stated metrics of object is described in this paper. The aim of the review paper is to analyse the basic functionalities of the object detection system, various applications of object detection, real-time data set used in object detection and tracking. Besides, the verification the existing problems and advise a novel method to enhance the detecting and tracking of object over digital images and video frames.

Keywords- *Object Detection, Tracking, Classification, and Deep Learning.*

I. INTRODUCTION

OD stands for Object Detection using DL (Deep Learning) and CV (Computer Vision) to work with VSs (Video Streams) and video records is give characteristics to verify the various categories of objects. OR (Object Recognition) is a vital job in IP (Image Processing) and CV (Computer Vision). It is main concerned with evaluating the verify of an objective being considered in a pictures. People can classify the purpose in real-world simply without any determinations, on different machines by it-self can't classify or detect the objects. OD is normal VPT (Visual Perception Task) and one of the main field of applications of CV (Computer Vision). It is necessary deals with searching specific objects with in a picture. OR is an unique of the main works in CV (Computer Vision). This procedure of verifying onjects in DIs (Digital Images), saved and real-time videos. OD is the computer field regarded to CV and IP that recognises and describes objects like people, car, animals from DIs and videos. This computer technology has the main power to detect one or more objects within 3D image or video. ODS has built various techniques, but OD using DL methods allows more precise for variety of object categories.

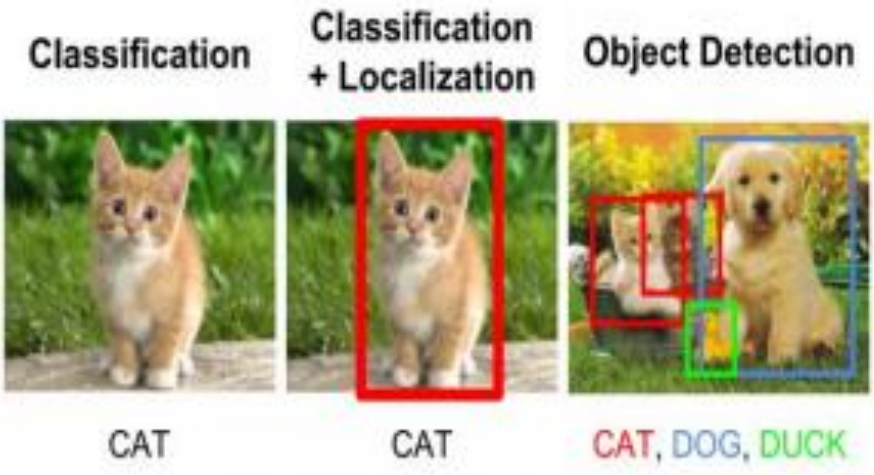


Fig 1. The basic functionality of Object Detection [1]

Above fig 1 shows the basic process or functionalities of OD system. OT (Object Tracking)[2] is the procedure of finding MOs(Moving Objects)[3] over-time using the digital camera in VSs (Video Sequences)[4]. The main idea of OT is to sub-ordinate target objects in successive video extraction frames. It needs shape and position or characteristics of objects in VFs (Video Frames). Thus, the classification is the prior phases of OT in CV use. To locate the MO in video frame, OD is the initial phase on tracking. Afterwards, object detection can be recognized as cars, person, animals, birds and other MOs. It is the main difficult job in the IP to detect the objects into successive video frames. [5] Several experiments can growth, due to difficult object motion, unstructured object shape, object detection in occlusion and scene and real-time processing necessities.

Several techniques use in object recognition like extracted properties and learning concepts to detect objects and relating to an object class. Object categories main deals with recognizing and classifying the object into particular category, whereas OD objectives at confining a main object of interest in DIs or videos. All object and category has it's own certain properties that describe themselves and distinguisg them. It is helping in detection or recognization of the similar objects in other images[6].

Classification[7]: Suppose an image area, select which of the various possible classes is current in that area.

Object locatization and detection[8]: Assume a difficult view image, select if an particular object of interest is suited, Wherever in this picture, and give precise position data on the object.

II. RELATED WORK

This section described the object detection, segmentation, and classification methods used in DIP (Digital Image Processing) field. Object Detection defines to collect the reliable jobs for verifying and authenticating objects in digital images. Object Detection (OD) is one of the most important parts in CV (Computer Vision). It seeks to position based on object features from a large category of the predefined data sets. **Zhong-Qiu et al., 2017[9]** analysed the overview on the DL (Deep Learning) and CNN (Convolutional Neural Network). Then, they focused on normal object detection architectures along with changes and reliable approach to enhance detection system performance in future. The specific detection jobs have different feature set values, it was described the related work in various methods, including salient OD (Object Detection), and face detection.

Problem Definition:

Object Detection System is to define somewhere things are positioned in a defined (Localization Object) image, and which class individual object relates to OC (Object Classification). So that the pipeline of OD structures can be separated into three different phases like (i) Data region-selection (ii) Extract the features, and (iii) Classification.

A. Data Region-Selection

Various objects have various aspect sizes; appear in locations of the image, it scans the complete image with a multi scale sliding window. Though, this plan can find-out all possible locations of the objects, it is limitations creates too various redundant windows. But, if only a static number of SWTs (Sliding Window Testing's) are realistic, un-satisfactory edges may be created.

B. Extract the features

It identifies various objects and it required to extract VFs (Vector Features) which can give a robust and semantic re-presentation. HoG, Haar and SIFT algorithm is used for features can get demonstrations connected with difficult cells in HB (Human Brain) [10,11,12]. Various problems are occurd such as illumination situations, background noises and diversity of apperances etc.

C. Object based Classification

This classification is requied to divide a target based on object from other classes, and then to create the representations more informative, semantic and hierarchical for VR (Visual Recognition). Normally, the SVM (Support Vector Machine), DPM (Deformable Part Based Model), and AB(AdaBoost) are very good classifiers. These modes are more reliable and flexible model by connecting object sections with de-formation cost to manage various de-formations. **Ajeet Ram Pathak et al., 2018[13]** discussed the world-wide applications such as (i) VS (Video Surveillance), (ii) SU (Scene Understanding) (iii) Robotics and (iv) SDSs (Self-Driving Systems) initiated enormous investigation in the field of Computer Vision(CV). These were main applications, VRS (Visual Recognition System) which includes IC (Image Classification), Image localization, and Image detecting. This article had performed a main part of DL (Deep Learning) Methods depends on CNN (Convolutional Neural Network) for OD. DL frameworked and facilities obtainable for OD were also pronounced.

Cong Tang et al., 2017[14] described the development of OD approaches depends on DL (Deep Learning) and discussed the most reliable approaches nowadays in the OD via DL. In the declaration of the approaches, the main focused on the outline project and the employed replicas and analyses the system presentation in the real-time and the correctness of the OD. Finally, it described the main limitation in the OD depends on DL and gives some explanations for reference. **Xinyi Zhou et al., 2017[15]** explained with the area of CV is mainly for the use of DL in OD job. Also, there was a easy summary of the data sets and DL methods normally used in CV (Computer Vision). The novel data set was constructed according to those normally used data sets, and select one of the network known as RCNN approach to work on this novel data set. Although, the simulation to support the understanding of these systems, and done the study of the consequences study the significance of DL technology, and the significance of the data set for DL. Table 1 describes the parameters, future work, problem statement and performance metrics.

TABLE I: VARIOUS PERFORMANCE ANALYSIS

Author Name	Year	Methods	Performance Metrics	Data set	Future scope	Problem Statement
Zhao et al., [9]	2019	Deep Learning Neural Networks	Time (sec) Fmeasure Mean Absolute Error Rate	YOLO PASCAL's ECSSD HKU-IS SOD	Valuable insights and rules for future enhancement	Overfitting High-dimensional data transformation
Lowe et al., [10]	2004	SIFT DoG	Repeatability Correct Nearest Descriptor Sensitivity	Large Dataset	Importance of Generic object classes To recognized suitable features in particular object class	Short range motion tracking Scale selection Edges Effects
Dalal et al., [11]	2005	HoG (Histogram of Gradients)	Miss Rate	MIT Pedestrain INRIA	Current linear SVM detector will used	Problem with huge dense descriptor vectors
Leinhart et al.,[12]	2002	HaaR Feature Vector	Miss Rate	-	-	-
Pathak et al., [13]	2018	CNN	Train and Test Accuracy	PASCAL	Real-Time Detection	Regression Issue
Tang C et al., [14]	2017	RCNN	-	PASCAL YOLO	Real time system using Deep Learning	Data explosion
Zhou et al.,[15]	2017	Faster RCNN and NN	Map	PASCAL VOC	Synthetic data to increase the amount of data in future	Classification issues and issue in open source data set used

III. DEEP LEARNING APPROACHES - OBJECT DETECTION SYSTEM

This algorithm concept is recycled by the system has been always enhancing, furthermore to the alters in the system model, the more is to fix some tune depends on the real system or applying few tricks to create the system presentation to improve. The well-known methods of OD is a serial of methods depends on RCNN mostly in the subsequent:

A. R_CNN Model

This model stands for regions with Convolutional neural network[16] has been the state-of-art articles in the area of OD in 2014 yrs. The main objective of this topic has altered the knowledge of OD. Advanced, the methods in various surveys on DL of OD normally inherited this objective which is the main method used for OD (Object Detection) with Deep Learning (DL). The main point of this paper is that the Convolutional Neural Network is implemented to CB (Candidate Box) to extract the feature-vector set, and the 2nd is to implement a way to efficiently sequence the huge Convolutional Neural Networks. It is used for supervised predefined on huge data set like ILSVRC and before fix some fine-tuning preparation in a particular variety on a minor data set like PASCAL VOC data set.

B. Model in Deep Learning Method SPP –Net [17]

This model is an enhancement depends on the Region-RCNN with rapidity. This planned approach a SPP (spatial pyramid pooling) model layer that eliminates limitations on the detection network fixed-size. It is only requires to execute the Convolution Layer (CL), and the use SPP model layer to excerpt characteristics, associated to the region-CNN, to escape repeat CO (Convolution Operatiom) the candidate area, mitigating the no. of convolution times. The SSP-net speed evaluating the convolution on the PASCAL VOC 2007 database by 30 to 170 times faster than the Region-CNN, and the complete speed is 24 to 64 times faster than the Region-CNN.

C. Fast RCNN

For the limitations of Region-CNN and SPP-Net, FAST RCNN [18] did the following enhancmenets: The higher detection quality mAP than Region-CNN and SPP-Net, write the LF (Loss Function) of various jobs together to get single level train procedure, in the training phase can update all the neural layers, don't require to save features in the dataset. Fast RCNN method can enhance the speed of training DNNs like VGG-16. To compared to RCNN, the rapidity for fast RCNN train process is nine times faster, and the rapidity for test is 213 time faster. The rapidity for fast RCNN train module is 3 times earlier than SPP-Net and the speed for test analysis is 10 times earlier, the accuracy value also have a certain increase.

D. Faster RCNN

Rise of SPP-Net and Fast RCNN has significantly reduced the executing time of object detection system. Moreover, the time they gross for the region proposed approach is too long, and the job of getting region proposal method is a blockage. Faster RCNN [19] defines a solution to this issue by changing tranditional practices, to use a DN to evaluate a proposal approach. The experimental analysis chooses the faster RCNN system. Table 1 defines the comparative analysis of mAP (Mean Average Precision) of above 4 types of network model on the VOC-2007 database.

TABLE II. MEAN AVERAGE PRECISION (mAP) OF DIFFERENT NETWORK ON THE VOC-2007 DATABASE.

Network/System	RCNN	SPP-NET	Fast RCNN	Faster RCNN
VOC 2007 mAP[20]	0.66	0.631	0.669	0.732

IV. DATASET USED IN OBJECT DETECTION SYSTEM

The database is one of the basics of DL(Deep Learning), for various investigators to achieve the data to carry-out the simulation just by themselves is main issue, so it requires various open-source database for everyone to value. Few normally used databases in CV (Computer Vision) is the following: (i) PASCAL (ii)YOLO and (iii) VOC 2007.

A. IMAGE-NET

This data set [21] has more than fortheen million digital images covering more than 20k classes. There are more than a million images with clear category comments and comments of object positions in the digital image. This data set is world wide used data sets in the area of DL (Deep Learning). Mostly research work like image classification, position and detection depends on this database. This data set is detailed and is simple to use. It is worldwide used in the area of CV (Computer Vision) research, and has become the specific data set of the recent DL of image-domain to test analysis method performance. There is a well defined main limitation known "Image-Net ILSVRC (International computer vision challenge" [22] depends o n the data set.

B. PASCAL

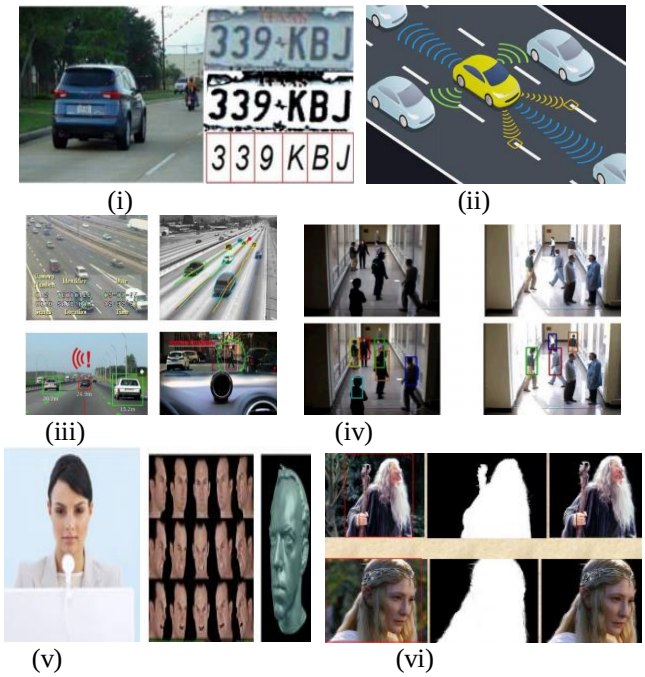
This data set gives standard image databases from object class recognition and detection and gives a common set of simulation tools for opening the databases and comments. This data set comprises twenty categories and has limitation depends on this database. This data set main challenge is no longer available after 2012, but it is database is of very good quality and well-defined, and allows calculation and comparative of various techniques. Because the sum of information of the PASCAL data set is small size, compared to the Image Net database, very appropriate for investegators to the test analysis network programs[23].

V. APPLICATIONS AREA OF OBJECT DETECTION SYSTEM

The object detection system is realted in various fields altering from VS (Video Survillence), HCI (Human Computer Inter-action), AI with Robotics, Transportation, Information Retrieval, and so on. Devices used for determined survillence produce peta bytes of picture information in few hrs. These information are optimized to geo spatial information and integrated with other information to grow clear view of recent scene. In this procedure includes OD[24] to trail things such as a vehicles, people and apprehensive objects from the rare imagery information. Detecting the animals in the area of disinfected such as detecting the vehicle parked in limited fields are also few uses of ODS (Object Detection System). Detecting is very important use of OD (Object Detection). OD analytics could achieved off-line, online or clostest real-time. The other factors such invariance, interclass, intraclass, rotation in-variance and occlusions, and multiple purposes OD required to be measured for OD (Object Detection).

The main applications are defines as follow:

- Character Recognition in Vehicles
- Vehicle Tracking
- Self-driving Vehicles
- Moving and Tracking Objects
- Face Identification
- Object Extraction
- Smile and acitivity Recognition etc.



(i) (ii) (iii) (iv) (v) (vi)

Fig 2. Various Applications used in Object Detection System[24]

VI. CONCLUSION AND FUTURE SCOPE

In this review paper, analysis on various detection, tracking and recognition of object methods, feature extraction, localization and Classification approach which depends on the video frame and several tracking tools. These type of methods are used towards increase the OD (Object Detection) with novel objectives. Due to its great culture capability and benefits in managing with scale transformation, occlusion and DL based OD (Object Detection) has been a research hot-spot in current years. It give a brief on DL depends on OD systems which manage various sub issues like clutter, minimum resolution, and occlusion. The analysis on object detection systems which gives basic functionalities, various applications are explained in section V. It described the deep learning concepts and methods depends on CNN, RCNN, Faster RCNN and etc. Deep Learning concepts and facilities are available for OD are also described in this paper. Object Detection data sets used for localization and classification released in world-wide and described in this review paper.

VII. REFERENCES

[1] Vahab, A., Naik, M. S., Raikar, P. G., & SR, P. (2019). Applications of Object Detection System.

[2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

[3] Viola, P., & Jones, M. (2001). Robust real-time object detection. *International journal of computer vision*, *4*(34-47), 4.

[4] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[5] Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep neural networks for object detection. In *Advances in neural information processing systems* (pp. 2553-2561).

[6] Papageorgiou, C., & Poggio, T. (2000). A trainable system for object detection. *International journal of computer vision*, *38*(1), 15-33.

[7] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), 1627-1645.

[8] Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE transactions on image processing*, *24*(12), 5706-5722.

[9] 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.

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

[11] 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.

[12] 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.

[13] Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Application of deep learning for object detection. *Procedia computer science*, *132*, 1706-1717.
[14] Tang, C., Feng, Y., Yang, X., Zheng, C., & Zhou, Y. (2017, July). The object detection based on deep learning. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)* (pp. 723-728). IEEE.
[15] Zhou, X., Gong, W., Fu, W., & Du, F. (2017, May). Application of deep learning in object detection. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)* (pp. 631-634). IEEE.
[16] Gidaris, S., & Komodakis, N. (2015). Object detection via a multi-region and semantic segmentation-aware cnn model. In *Proceedings of the IEEE international conference on computer vision* (pp. 1134-1142).
[17] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), 1904-1916.
[18] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
[19] 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).
[20] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, *88*(2), 303-338.
[21] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.
[22] Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., & Fei-Fei, L. (2012). Imagenet large scale visual recognition competition 2012 (ILSVRC2012). net. org/challenges.
[23] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2007). The PASCAL visual object classes challenge 2007 (VOC2007) results.
[24] Chang, W. L., & Fox, G. (2015). *NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements* (No. Special Publication (NIST SP)-1500-3).