International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Object Detection using Deep Learning

K Lekshmi Sasidharan

School of Computer Science and Application, REVA University, Bangalore, India

---

Abstract - *The extensive applications of object detection in robotics, self-riding automobiles, scene understanding, video surveillance etc triggered huge studies in the area of computer vision. Being in the middle of these applications, visual recognition structures which include image classification, localization and detection have a high priority these days. Due to the remarkable upgradation in neural networks particularly in deep learning, the visual recognition structures have attained an exceptional performance. Object detection is such a domain witnessing equisite achievement in computer vision. Here this paper ymbolizes the function of deep learning techniques primarily based totally on convolutional neural network for the object detection. Deep learning strategies for modern day object detection are assessed in this paper.*

Key Words: deep learning, convolution neural network, object detection, visual recognition, computer vision.

I. INTRODUCTION

Gartner's 2018 trends in technology states that Artificial Intelligence could be extensively used trend amongst the industries and so the Computer vision ! . Industries primarily based totally on automation, markets, medical domains, protection and surveillance sectors are maximum probable domains significantly with the use of computer vision. It is forecasted that the CV marketplace could reach 33.3B USD in 2022 fostering the outstanding emergings in the domain consumer, robotics, and machine learning.

Technologies in deep learning has turn out to be a buzzword in recent times because of the brand new effects acquired withinside the area of image classification, object detection, natural language processing. The motives in the back of the trending of deep learning are 2-folded, viz. massive availability of datasets and effective Graphics Processing Units. As deep learning wants massive datasets and effective sources to carry out the training part, each necessities have already been done in this present era. Fig. 1 suggests the upsurge of Deep Learning with Computer Vision within the current lustrum.

Image classification, being the extensively researched domain with side the area of computer vision has finished outstanding effects in world-spread competitions as ILSVRC, PASCAL VOC, and Microsoft COCO with the assistance of deep learning in those competitons. Motivated with the aid of using the effects of image classification, deep learning had been made for object detection and deep learning primarily based on object detection has additionally finished the state-of-the-effects. We are having the goal to evaluate deep learning strategies primarily based on convolutional neural network (CNN) for object detection. The splendor of convolutional neural networks is that they no longer rely upon manually created characteristic extractors or filters. Rather, they teach in line with each sec from the raw pixel degree as much as very last item categories.

Fig. 1. Upsurge of deep learning for computer vision over the recent lustrum from March 2013 to January 2018 (Created by Google Trends)

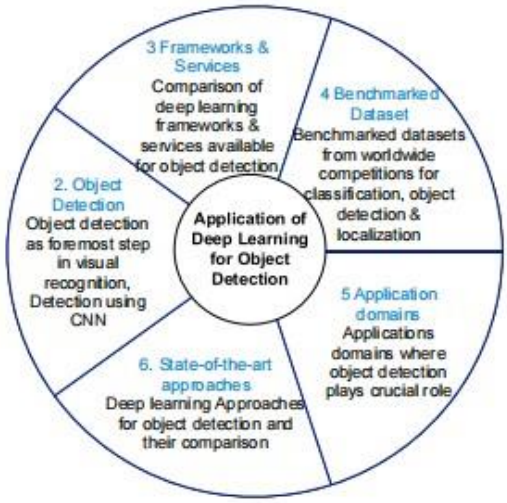

Fig. 2. Organization of the paper

Deep neural architectures handles most complicated models successfully than the other shallow networks.

CNNs are much less correct for smaller information however it displays significant/ record breaking accuracy at the big image datasets. But, CNNs require big quantity of

classified and named datasets to carry computer vision associated tasks such as recognition, type and detection.

II. OBJECT DETECTION

2.1. Object detection as the first step

Object detection is the technique of figuring out the instance of the class with which the item belongs and estimating the area of the object through the outputting of the bounding container across the object. Detecting one example of a class from an photoimage is referred to as single class object detection, while detecting the classes of all object present withinside the image is called multi class object detection. Different demanding situations which include partial/complete occlusion, varying illumination conditions, poses, scale, and so on are had to be dealt with at the same time as doing the object detection. As shown in the fig3, object detection is the first step in any visual recognition activity.

2.2. Object detection using CNN

Deep Convolutional Neural Networks had been significantly used for object detection. CNN is a kind of feed-forward neural network and works on precept of weight sharing. Convolution is an integration displaying how one characteristic overlaps with other characteristic and is a mix of these two capabilities being multiplied. Fig. 4 indicates layered structure of CNN for item detection. Image is convolved with the activation function to get the feature extraction graphs. To lessen the spatial complexity of the network, extraction maps are handled with inner layers to get abstracted feature maps. This manner is repeated for the favored no. of time and consequently the feature extraction maps are created. Eventually, those function maps are processed with absolutely related layers to get output of image recognition displaying self assurance rating for the anticipated class labels. For ameliorating the complexity of the network and decrease the wide variety of parameters, CNN employs distinct forms of pooling layers as shown in the table 1. Inner layers are translation-invariant. Activation maps are fed as an trigger to the pooling layers. They perform on every patch within the decided map.

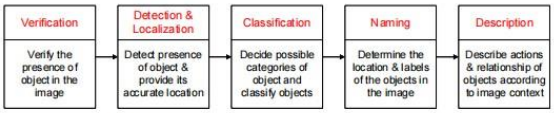

Fig. 3. Object detection as foremost step in visual recognition activity

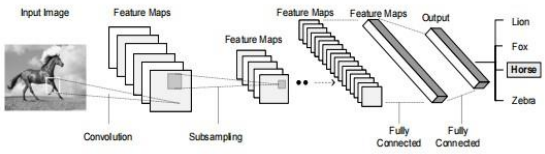

Fig. 4. Use of Convolutional neural network for object detection

Table 1. Pooling layers used for object detection.

Pooling layer	Description
Max pooling	It is widely used pooling in CNNs. It takes maximum value from the selected image patch and place in the matrix storing the maximum values from other image patches.
Average pooling	This pooling averages the neighborhood pixels.
Deformation pooling [40]	Deformable pooling has ability to extract deformable properties, geometric constraints of the objects.
Spatial pyramid pooling [53]	This pooling performs down-sampling of the image and produces feature vector with a fixed length. This feature vector can be used for object detection without making any deformations on the original image. This pooling is robust to object deformations.
Scale dependent pooling [54]	This pooling handles scale variation in object detection and helps to improve the accuracy of detection.

III. FRAMEWORKS AND SERVICES OF OBJECT DETECTION

The listing of deep learning frameworks which are existing till date is exhaustive. We have cited a few important deep learning frameworks in the tab 2. The frameworks are studied from the factor of view of showing features, interface, help for deep learning of version viz. convolutional neural network, recurrent neural network , Restricted Botltzmann Machine and Deep Belief Network and in support for Multi-node parallel execution, developer of the architecture. Tab 3 displays the listing of offerings which may be used for the object detection. These offerings may be availed via the APIs cited in the above mentioned table.

IV. APPLICATION DOMAINS OF OBJECT DETECTION

Object detection is relevant in a wide variety of domain names starting from defense (surveillance), human computer interaction, robotics, transportation, retrieval, etc. Sensors which are used for continual surveillance generate petabytes of photograph facts in a very few hours. These information is decreased to geospatial facts and incorporated with different facts to get clean data of the current scenario. This procedure includes object detection to tune entities like people, cars and suspicious gadgets from the raw imagery facts [12]. Spotting and detecting the wild animals withinside the territory of sterile zones like industrial area, detecting the ones parked in limited regions also are a few implications of object detection. Detecting the unattended may be very vital utility of object detection. For self sufficient driving, detecting objects on the way might play a critical role. Detection of defective electric powered wires whilst the photograph is captured from drone camera is likewise utility of the object detection. Detecting the drivers' drowsiness on the motorway on the way to keep away from accidents can be done through object detection.

The necessities of aforementioned implementation range in line with the use case. Object detection analytics may

be accomplished offline, or online or close to actual time. Other elements like occlusions, rotation invariance, interclass and intra-class variation, and multi-pose object detection want to be taken into consideration for the object detection scheme.

V. DEEP LEARNING BASED APPROACHES OF OBJECT DETECTION

The figure compares deep learning techniques for item detection that is beneficial for the study network to work similarly withinside the area of deep learning of primarily base in object detection. Szegedy et al. pioneered the usage of deep CNN for item detection [13] through modeling item detection as a regression problem. They have changed ultimate layer withinside the AlexNet [2] with regression layer for item detection. Both the obligations of detection and localization have been completed the use of object mask regression. DeepMultiBox [14] prolonged the method of [13] to locate multiple gadgets in an photograph. How the CNN learns the characteristic is a prime issue. The undertaking of visualizing the CNN capabilities is finished by Zeiler et al. [15]. They implemented each CNN and deconvolution procedure for visualisation of capabilities. This method outdoes[2]. They have additionally justified that overall performance of deep version is stricken by the intensity of the network. Overfeat version [16] applies Sliding window method primarily based totally on multiscaling for mutually appearing classification, detection and localization. Girshick et al. [17] proposed deeper version of the model primarily based totally on Region proposals. In this method, photograph is divided into small areas after which deep CNN is used for purchasing characteristic vectors. Features vectors are used for classification through linear SVM. Object localization is finished the use of bounding-field regression. On the same lines, [18] used regionlets for usual item detection no matter context information. They designed Support Pixel Integral Image metric to extract capabilities primarily based totally histogram of gradients, covariance capabilities and sparse CNN.

Earlier then the deep learning, item detection become ideally completed by the use of deformable part approach [19]. Deformable part model approach plays multi-scale primarily based object detection and localization. Based at the standards of this version, Ouyang et al. [20] positioned forth pooling layer for coping with the deformation homes of items for the sake of detection.

Method	Working	Features	Reference
Deep saliency network	CNNs are used for extracting the high-level and multi-scale features.	It is challenging to detect the boundaries of salient regions due to the fact that pixel residing in the boundary region have similar receptive fields. Due to this, network may come with inaccurate map and shape of the object to be detected.	[48-51]
Generating image (or pixels)	This method is used when the occurrence of occlusions and deformations is rare in the dataset.	This method generates new images with occlusions and deformations only when training data contains occurrences of occlusions and deformations.	[55]
Generating all possible occlusions and deformations	In this method, all sets of possible occlusions and deformations are generated to train the object detectors.	This method is not scalable since deformations and occlusions incur large space.	[56], [57]
Adversarial learning	Instead of generating all deformations and occlusions, this method use adversarial network which selectively generates features mimicking the occlusions and deformations which are hard to be recognized by the object detector.	As this method generates the examples on-the fly, it is good candidate to be applied in real time object detection. As it selectively generates the features, it is also scalable.	[58]
Part-based method	This method represents object as collection of local parts and spatial structure. This method exhaustively searches for multiple parts for object detection.	This method addresses the issue of intra-class variations in object categories Such variations occur due to variation in poses, cluttered background, partial occlusions	[39]
CNN with part-based method	In this method, deformable part model is used for modelling the spatial structure of the local parts whereas CNN is used for learning the discriminative features.	This method handles the issue of partial occlusions. But requires multiple CNN models for part based object detection. Finding out the optimal number of parts per object is also challenging.	[59-61]
Fine-grained object detection method	This methods works on annotated object parts during training phase. Part-localization is the fundamental component used in testing phase.	This method has capability to figure out the differences in inter-class objects at finer level. And they work more on discriminative parts compared to generic object detection methods.	[62-65]

For object detections and localization, Huang et al. [21] proposed task-pushed progressive part localization (TPPL) framework. Spatial Pyramid pooling layer and swarm optimization method is used for detecting the item withinside the photograph area. Zhu et al. placed forth hybrid technique primarily based totally on segmentation and context modeling for item detection [22] through using Markov Random Field. The use of multi-scale fashions and context fashions is performed in [23] for joint object detection and localization.

Approaches stated in [13-23] have centered on object detection with aim of keeping the accuracy of detection. The methods stated in [24-27] have centered on close to

real time detection of item through keeping the trade-offs from most of the overall performance metrics. Saliency-stimulated methods are stimulated from human vision which has the functionality to pick the important data from the complicated photograph. These methods comply with the idea of contrasts withinside the photograph. Deep learning executed first rate overall performance in salient object detection [28-31]. Girshick et al. prolonged the preceding work of region of interest pooling from [23] through introducing multi-scale training and multi-task training for quicker object detection in [24]. As region of interest pooling works exhaustively in every photograph area, it's so far computationally very expensive. To alleviate this problem, a technique primarily based totally on area proposal community is placed forth in [25] which makes use of convolutional network for simultaneous localization and detection.

For a quicker object detection, rather than the usage of dense CNN, Kim et al. [26] used shallow CNN withinside the method –PVANET. This method works in pipeline structure withinside the consecutive steps of local concept generation, characteristic extraction, and categorisation primarily based on region of interet. "You Only Look Once (YOLO)" [27] is a popular and broadly used framework for item detection in our environments because of its feature of scanning the photograph in the simplest once even as training and testing it out for inferring the data at context and look degree.

As image categoriation datasets own massive quantity of traning data as compared to object detection data, for harnessing the effect of massive facts to be had with image categoriation and use it for item detection, Redmon and Farhadi implemented the hierarchical classification technique. Their method particularly YOLO9000 is the improved version of YOLO framework [27] and plays detection for around 9000 item detection at actual time. YOLO9000 uses the technique for combining the unique dataset (that are inherently now no longer intended for item detection) and join the training method wherein the version is educated on each ImageNet dataset and Microsoft COCO dataset.

It is anticipated for item detection structures to robustly carry out item detection invariant to illumination, occlusions, deformations and intra-class variations. As occlusions and deformations comply with long-tail statistical distribution, there are possibilities that datasets pass over the uncommon occlusions and deformations of objects. This hinders the overall performance of item detection structures. Therefore, Wang et al. [58] placed forth the method primarily based totally on adversary community wherein community selectively generates the capabilities of occlusions and deformations that are difficult to be identified through item detector. They used Spatial Dropout Network and Spatial transformer Network primarily based totally on antagonistic network to generate occlusion and deformation capabilities respectively.

Fine-grained item detection calls for locating the diffused variations amongst inter-class object classes. Fang et al. placed forth co-incidence layer for integrating CNN with element-primarily based totally techniques. The coincidence layer encodes the co-incidence among numerous elements detected through the neurons. This layer does now no longer want part-degree annotation as required in elementary region-primarily based totally models and generates the co-incidence capabilities the usage of singlestream network.

As assessed from the literature, deep learning strategies specially primarily based totally on convolutional neural networks are relevant to each generic item detection and fine-grained item detection and localization. CNNs being the base point of item detection strategies are very beneficial for robotically studying the capabilities used for object detection.

VI. CONCLUSION AND FUTURE DIRECTIONS

Object detection is taken into consideration as the fundamental step in deployment of self riding automobiles and robotics. In this paper, we demystified the function of deep mastering strategies primarily based totally on CNN for object detection. Deep learning architecture and offerings for object detection also are there withinside the paper. Benchmarked datasets for item localization and detection launched in international competitions also are covered. The point to the domain names wherein item detection is relevant has been mentioned. State-of-theartwork in deep learning primarily based totally object detection strategies have been assessed and as compared. Future instructions may be said as follows. Due to infeasibility of people to manner massive surveillance facts, there is a want to bring facts towards the sensor in which facts are generated. This could end result into actual time detection of objects. Currently, item detection structures are small in size only having 1-20 nodes of clusters having GPUs. These structures need to be prolonged to address actual time complete movement video producing frames at 30 to 60 according to each second. Such item detection analytics need to be incorporated with different equipments which uses the data fusion. The most important trouble is the way to integrate processing right into a centralized, effective GPU for processing facts received from numerous servers concurrently and plays close to actual time detection analysis. To make the most of the representational strength of deep learning, massive datasets over the dimensions of 100TB are essential. More than 100 M images are required to

educate the self-riding automobiles[32]. Deep learning libraries need to be augmented with prototyped environments in an effort to offer paramount throughput and productiveness handling huge linear algebra primarily based operations. The datasets of image type are broadly to be had as compared to that of object detection, the techniques may be devised through which datasets supposed for different responsibilities aside from object detection could be relevant for use for object detection. Existing techniques are advanced thinking about object detection as essential hassle to be solved. There is scope of growing new layout mechanisms able to providing "Object Detection as a Service" in complicated applications inclusive of drone cameras, computerized riding of automobiles, robots navigating the regions inclusive of planets, deep sea bases, and commercial plant life in which excessive stage of precision in sure responsibilities is expected.

REFERENCES

[1] Lin, Tsung-Yi et al. (2014). "Microsoft Coco: Common Objects in Context." In European Conference on Computer Vision, 740–55.
[2] Deng, Jia et al. (2009). "Imagenet: A Large-Scale H ierarchical Image Database." In Computer Vision and Pattern Recognition, 248–55.
[3] Torralba, Antonio, Rob Fergus, and William T Freeman. (2008). "80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition." IEEE transactions on pattern analysis and machine intelligence, **30(11):** 1958–70.
[4] Krizhevsky, Alex, and Geoffrey Hinton. (2009). "Learning Multiple Layers of Features from Tiny Images." Thesis ch.3
[5] Wah, Catherine et al. (2011). "The Caltech-Ucsd Birds-200-2011 Dataset."
[6] Welinder P., Branson S., Mita T., Wah C., Schroff F., Belongie S., Perona, P. (2010). "Caltech-UCSD Birds 200". California Institute of Technology. CNS-TR-2010-001. 2010
*Ajeet Ram Pathak et al. / Procedia Computer Science 132 (2018) 1 706–1717*1716[27] Griffin, Gregory, Alex Holub, and Pietro Perona. (2007). "Caltech-256 Object Category Dataset."
[8] Russakovsky, Olga et al. (2015). "ImageNet Large Scale Visual Recognition Challenge." Int. Journal of CV, **1 15(3):** 211–32.
[9] Everingham, Mark et al. (2015). "The Pascal Visual Object Classes Challenge: A Retrospective." Int. journal of CV, **111(1):** 98–136.
[10] Russell, Bryan C, Antonio Torralba, Kevin P Murphy, and William T Freeman. (2008). "LabelMe: A Database and WebBased Tool for Image Annotation." International journal of computer vision, **77(1–3):** 157–73.
[11] Xiao, Jianxiong et al. (2010). "Sun Database: Large-Scale Scene Recognition from Abbey to Zoo." In CVPR, 3485–92.
[12] Chang, Wo L. (2015). NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements. [13] Szegedy, Christian, Alexander Toshev, and Dumitru Erhan. (2013). "Deep Neural Networks for Object Detection." In Advances in Neural Information Processing Systems, 2553–61.
[14] Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. (2014). "Scalable Object Detection Using Deep Neural Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2147–54. [15] Zeiler, Matthew D, and Rob Fergus. (2014). "Visualizing and Understanding Convolutional Networks." In European Conference on Computer Vision, 818–33.
[16] Sermanet, Pierre et al. (2013). "Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks." arXiv preprint arXiv:1312.6229.
[17] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. (2014). "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–87.
[18] Wang, Xiaoyu, Ming Yang, Shenghuo Zhu, and Yuanqing Lin. (2015). "Regionlets for Generic Object Detection." IEEE transactions on pattern analysis and machine intelligence, **37(10)**: 2071–84.
[19] Felzenszwalb, Pedro F, Ross B Girshick, David McAllester, and Deva Ramanan. (2010). "Object Detection with Discriminatively Trained Part-Based Models." IEEE transactions on pattern analysis and machine intelligence,**32(9)**: 1627–45.
[20] Ouyang, W et al. (2015). "DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection." In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2403–2412.
[21] Huang, Chen, Zhihai He, Guitao Cao, and Wenming Cao. (2016). "Task-Driven Progressive Part Localization for FineGrained Object Recognition." IEEE Transactions on Multimedia, **18(12)**: 2372–83.
[22] Huang, Chen, Zhihai He, Guitao Cao, and Wenming Cao. (2016). "Task-Driven Progressive Part Localization for FineGrained Object Recognition." IEEE Transactions on Multimedia, **18(12)**: 2372–83.
[23] Ohn-Bar, Eshed, and Mohan Manubhai Trivedi. (2017). "Multi-Scale Volumes for Deep Object Detection and Localization." Pattern Recognition, **61:** 557–72.
[24] Girshick, Ross. (2015). "Fast R-Cnn." arXiv preprint arXiv:1504.08083.

[25] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. (2017). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE transactions on pattern analysis and machine intelligence **39(6):** 1137–49.
[26] Kim, Kye-Hyeon et al. (2016). "PVANET: Deep but Lightweight Neural Networks for Real-Time Object Detection." arXiv preprint arXiv:1608.08021.
[27] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–88.
[28] Liu, Nian, and Junwei Han. (2016). "Dhsnet: Deep Hierarchical Saliency Network for Salient Object Detection." In Computer Vision and Pattern Recognition
[29] Li, Xi et al. (2016). "Deepsaliency: Multi-Task Deep Neural Network Model for Salient Object Detection." IEEE Transactions on Image Processing, **25(8)**: 3919–30.
[30] Wang, Lijun, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. (2015). "Deep Networks for Saliency Detection via Local Estimation and Global Search." In Computer Vision and Pattern Recognition (CVPR), 183–92.
[31] Li, Guanbin, and Yizhou Yu. (2016). "Deep Contrast Learning for Salient Object Detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 478–87.
[32] Bojarski, Mariusz et al. (2016). "End to End Learning for Self-Driving Cars." arXiv preprint arXiv:1604.07316.
[33] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." IEEE transactions on pattern analysis and machine intelligence **37(9):** 1904–16.
[34] Yang, Fan, Wongun Choi, and Yuanqing Lin. (2016). "Exploit All the Layers: Fast and Accurate Cnn Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2129–37.