A Complete Analysis on the applications and challenges in Object Detection models.

S. Loganayagi [1], Dr.D.Usha[2]

[1]*Ph.D Research Scholar, Department of Computer Science, Mother Teresa Women's University, Kodaikanal*

[2]*Assistant Professor, Department of Computer Science, Mother Teresa Women's University, Kodaikanal*

Abstract

Object Detection is a visionary technique wrt computer in locating or identifying the existence of objects present in images and videos. This technology in Artificial Intelligence helps the computer for visualisation and identification. An image consists of various objects and each application focusses on specific objects for example face detection application for face finding and a control system created for traffic focusses the vehicles. It can also be used in capturing the item of notice and thus it improves the time of execution. It involves image classification and object localization and can be attained by applying either ML or DL such as Viola Jones face detector, AlexNet etc. It has a lot of applications to fit in, it is used in autonomous driving, agriculture field, crowd counting , healthcare sector and so on. This paper focuses on the complete understanding of Object detection models, areas of applications and challenges.

Keywords: Object detection, face detector, AlexNet, autonomous driving.

1. Introduction

Object detection majorly could be utilized for autonomous driving (Chen et al., 2017), face recognition (Yang and Nevatia, 2016), image classification (Jia et al., 2014; Krizhevsky et al., 2021), and behavior analysis of humans (Cao et al., 2017). Identifying objects in the image through an object identification algorithm focuses upon image resolution with classifier as the primary technique in NN models namely ResNet-50, InceptionNet, VGGNet-16, AlexNet, and other models along with face detection and the feature extraction algorithms (Abdulrazzaq and Dino, 2020). The object detection algorithms such as KNN, MultiLayer Perceptron, Naïve Bayes, Linear Regression, Random Forest, Support Vector Machine, decision tree, and others are most commonly utilized by researchers for higher accuracy in ANN (artificial NN), CNN, DNN and other machine learning (ML) based models.

2. Object Detection

The study by Kumar et al., (2017) focused exclusively on object recognition and identification/ detection model, in images. Authors argued, in image recognition and detection, the object detecting algorithm-based models purely rely upon pattern recognition, learning, and matching procedures through feature-based or appearance-based methods. Image intensity, brightness, number (single or multiple) of objects, color, size, pixels, contrast, and other features are extracted from input images and are later resized (pre-processed) according to the researcher's model adaptation.

Authors Deng et al., (2020) studied multiple object identification in deep learning models that uses two-stage image recognition algorithms. The authors Jadhav et al., (2020) found differences between the customary and contemporary deep learning models. According to both studies, the obtained findings by Deng et al., traditional detection models include a few stages such as feature extraction, feature classification, window sliding, pre-processing, feature selection, and post-processing. They classified one-stage and two-stage algorithms in object detection. Authors categorized models like SPP-Net (spatial pyramid pooling) by He in 2015, R-CNN (regions with CNN) by Girshick in 2014, Faster-RCNN by Ren, and Fast-RCNN by Girshick (2015) as one-stage algorithms. Similarly, models such as Yolo-v1 by Redmon in 2016,2017 and 2018 , SSD (single shot detector) by Liu in 2016, and Yolo-v4 by Bochkovskiy in 2020 are classified as two-stage algorithms. The first stage (single-stage) relies on region proposal whereas the second stage (two-stage) relies on regression. Thus they concluded rather than adopting single-stage algorithms two-stage is efficient; however, it lacks in multi-category detection of objects, and thus when developing multi-objects detection models the single-stage is efficient and rapid for dependence datasets.

3. Object Detection Methods

The methods involved in the detection of object includes ML and DL. In ML the features has to be definite with the usage of any of the available methods and then using other techniques to do classification. But in Deep learning technique , the entire recognition method does not require the features that do classification.

a)Machine Learning based object detection

Machine learning refers in making the machines to think like humans. It is the process of usage of algorithms in anlyzing ,learning and determining things based on the given data. It uses image-based feature extraction technique to extract the feature(Fig 1a). The various machine learning techniques/methods are as follows:

1. **Viola Jones face detector**
2. **Histogram of Oriented Gradients(HOG) features**
3. **Scale Invariant Feature Transform(SIFT)**

- **Viola Jones face detector**

The face detection algorithm which provided satisfactory outcomes. The face structures were hardcoded and the classifier was trained accordingly which could use to detect faces. It involves the combination of AdaBoost algorithm, feature assortment and calculation tools. At the time of analyzing videos with moving objects, it is unnecessary to apply detection to every frame. KLT algorithm a variant of tracking algorithm helps in identifying of finding bounding boxes and following amongst edges.

Object Identification using KLT

Detector1= vision.CascadeObjectDetector

Detector1= vision.CascadeObjectDetector(model)

Detector1 = vision.CascadeObjectDetector(XMLFILE)

Detector1 = vision.CascadeObjectDetector(Name,Value)

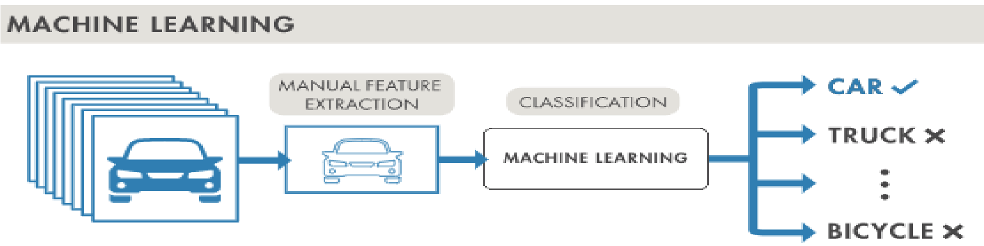

Figure 1: a. Machine Learning based Entity Recognition

- **Histogram of Oriented Gradients(HOG)**

The distribution of intensity gradients can be used to describe the appearance and shape of the object within an image. The given image is divided into cells which are connected regions. Each detector window has a descriptor, which serves as information during image recognition. Descriptor helps in training and testing of classifiers. The ML approach to object detection is similar to that of form frameworks, measure invariant change forms and advantage alignment histograms. Histogram of Oriented Gradients is implemented using 5 steps- the calculation of ramp, orientation binning, computation of descriptor blocks, wedge regularization and object acknowledgment.

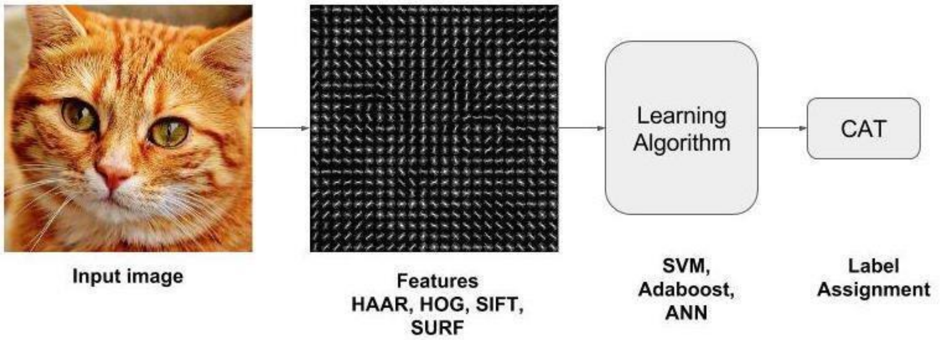

Figure 2: b. Histogram of Oriented Gradients

This method uses hand coded features that resulted in the failure with more noise and distractions in the background in the event of generalized setting(*Figure 2* b).

- **Scale Invariant Feature Transform**

SIFT algorithm is mostly used to detect and also to describe the local features in images. It is widely used in images. It can be used in image stitching, navigation , robotic mapping, object recognition, gesture recognition, 3D modeling and video tracking. The processes comprises of Difference of Gaussians(DoG)Space Generation, key points detection and feature description.

b) Deep Learning Based Object Detection

Deep learning works based on the neurons and its working present in our brain. Neural networks are cast-off in deep learning methods to achieve the result.

It is based on CNN, which utilizes the multi-layer neural networks for recognizing the visual patterns for the image(*Figure* 2).

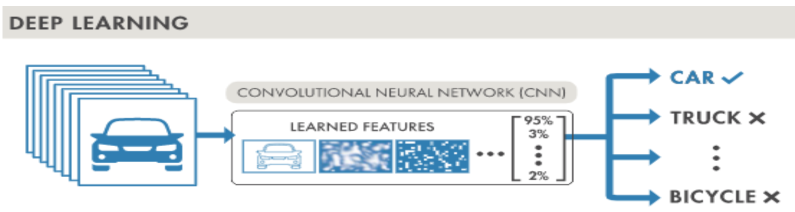

Figure 3: Deep learning base Object Detection

The frequently used DL models for edge recognition includes

- **R-CNN model family**: The Region based CNN are of three types R-CNN, Fast R-CNN, Faster R-CNN.
- **YOLO model family**: You Look Only Once acronym for YOLO, its types are YOLOv1, YOLOv2, YOLOv3.

The Steps involved in object detection

Step 1: Input is either the visual taken from image or video.

Step 2: The input is divided into sections or regions.

Step 3: Work with each section individually

Step 4: Classify into classes by passing into CNN

Step 5: Once the classification is over, the individual images can be combined and generate the original input image.

1. **AlexNet**

CNN has become standard for image classification next to kriszhevsky's CNN's performance during ImageNet.

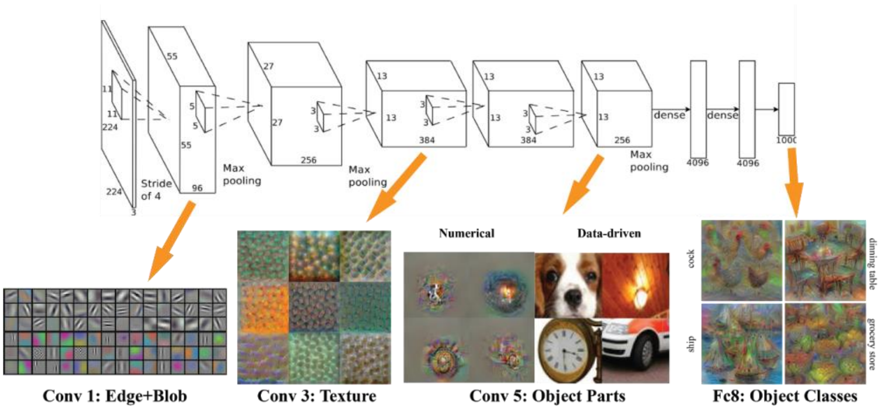

Figure 4: AlexNet

R-CNN

The R-CNN works by dividing the pictorial into chunks and focuses on the parts having sophisticated possibility of an object , whereas the YOLO focus on the whole image and identifies the bounding boxes after which it calculates the class probabilities for labelling the boxes.

	R-CNN	Fast-RCNN	Faster-RCNN
Trial period for each image	50 sec	2 sec	0.2 sec
Speediness	1x	25x	250x

Table 1: Comparison of R-CNN, Fast-RCNN, Faster-RCNN

You Only Look Once(YOLO)

It foresees the leaping boxes and session responsibilities in a single evaluation . They reframe the entity recognition similar to single regression problem. It has come with various versions for example YOLO V1, V2 V3 . YOLO V3 is the best of all three versions.

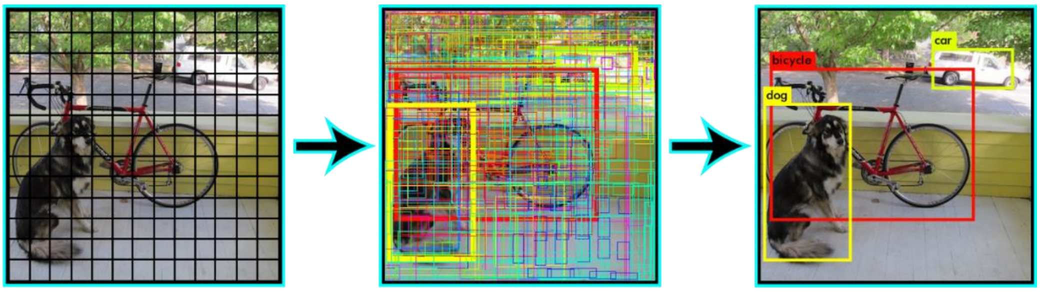

Figure 4: YOLO

Working of Object Detection

If the image is provided as input to CNN, the way in which we classify the class based on the image is known as image classification. The image classification(Fig 4) possesses a regression task which can predict the object's position with the help of bounding box(Fig 6).

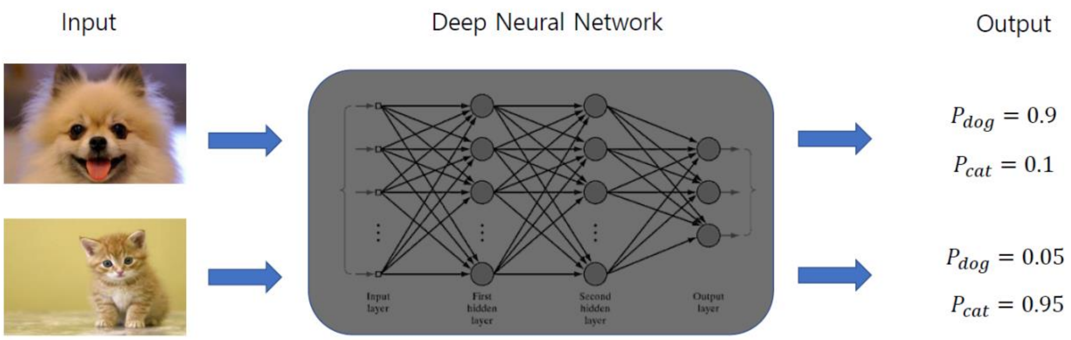

Figure 5: Image Classification

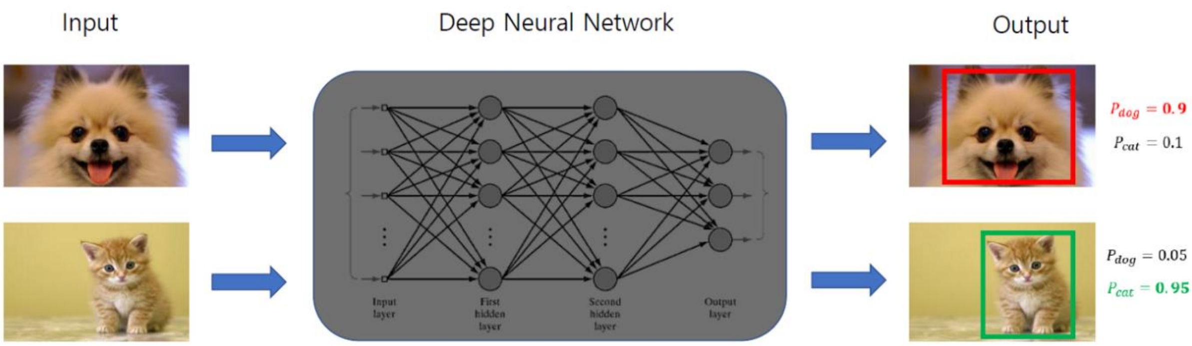

Figure 6: Object Detection along with Bounding box

4. Applications of object detection models

Authors Sunil and Gagandeep (2019), examined digital images based implementation of detection models. They affirmed that feature extraction (color, optical flow, edges, and gradient histograms) plays a vital and eminent role in object detection. Though recognition models have gained popularity, the application of models has a certain level of challenges like object rotation, lighting/ clarity, positioning, scaling method, mirroring, and occlusion condition hinders accuracy and precision of models in recognizing and identifying objects.

A study by Pathak et al., (2018) focused on and scrutinized the applications of DL-based entity recognition fashionable CNN models. They structured the features and interface according to deep learning models. Python was found to be a more efficient interface and used in major models, where Caffe, Microsoft-Cognitive Toolkit (CNTK), Theano, Keras, Tensorflow, MXnet, Chainer, Neon, Apache Singa, and Deeplearning4j are commonly used frameworks. However, the authors found that the presentation of each DL model varies in entity recognition and thus affects the outcome. Hence it was evidently concluded that utilizing the appropriate number of datasets could enhance the accuracy.

5. Challenges in object detection models

Authors Tiwari and Singhai, (2017) also had specified that feature selection is a crucial phase in object detection and application of the model. However, they also reviewed and explored other techniques and processes involved in the detection model. They found temporal differentiating, optical flow, and subtraction of background (selected frame subtracted from the background frame, pixel functioning in Gaussian probability, multimodal distribution, and test frame and median frame subtraction) as significant phases that are perceived as challenges.

6. Conclusion

Contrarily, the authors insisted that prior to developing a model, the choice of tracking method is also significant. Tracking techniques vary for a model and the accuracy rate is affected when the computational cost is lesser. Finally, the study concluded that the higher the accuracy rate, the higher the computational complexity becomes. Statistical techniques, sequential differencing, circumstantial reduction and visual movement are the major contributors in object recognition models, where the developer should consider the challenges of darker shadows, illumination changes, and object occlusion.

References:

Bochkovskiy. A, Wang. C.Y, and Liao. H.Y.M, (2020), “YOLOv4: Optimal Speed and Accuracy of Object Detection”, Computer Vision and Pattern Recognition.

Cao. Z, Simon. T, Wei. S-E and Sheikh. Y, (2017), “Realtime multi-person 2d pose estimation using part affinity fields,” in Computer Vision and Pattern Recognition, 1-9.

Chen. X, Ma. H, Wan. J, Li. B and Xia. T, (2017), “Multi-view 3d object detection network for autonomous driving,” in Computer Vision and Pattern Recognition - IEEE, 1907-1915.

Deng. J, Xuan. X, Wang, Li. Z, Yao. H and Wang. Z, (2020), “A review of research on object detection based on deep learning”, Journal Physics: Conference Series, 1684 (012028), 1-7.

Dino. H.I and Abdulrazzaq. M.B, (2020), “A Comparison of Four Classification Algorithms for Facial Expression Recognition”, Polytechnic Journal, 10(1),pp: 74-80.

Girshick. R, (2015), “Fast R-CNN”. In: Proceedings of the IEEE international conference on computer vision, Santiago, 1440-1448.

Girshick. R., Donahue. J., Darrel. T and Malik. J, (2014), “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”, In: Computer Vision and Pattern Recognition, Columbus, 580-587.

He. K.M, Zhang. X.Y, Ren. S.Q. and Sun. J, (2015), “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition”, IEEE Transactions on Pattern Analysis & Machine Intelligence, 37: 1904-1916.

Jadhav. P, Koli. V, Shinde. P and Pawar. M. M, (2020), “Object Detection using Deep Learning”, International research journal of engineering and technology (IRJET), 7(9), 275-279.

Krizhevsky. A, Sutskever. I and Hinton. G. E, (2012), “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, 1-9.

Kumar. S, Balyan. A and Chawla. M, (2017), “Object detection and recognition in images”, International Journal of Engineering Development and Research, 5(4), 1029-1034.

Liu. W, Anguelov. D, Erhan. D., et al., (2016), “SSD: Single Shot MultiBox Detector”, European Conference on Computer Vision, 21-37.

Pathak. A. R, Pandey. M and Rautaray. S, (2018), "Application of Deep Learning for Object Detection", In: International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science 132: 1706-1717.

Redmon. J, Divvala. S, Grishick. R, and Farhadi. A, (2016), "You Only Look Once: Unified, Real-Time Object Detection", In: Computer Vision and Pattern Recognition. Las Vegas, 779-788.

Redmon. J, Farhadi. A, (2017), "YOLO9000: better, faster, stronger", In: Computer Vision and Pattern Recognition, Hawaii, 7263-7271.

Redmon. J, Farhadi. A, (2018), "Yolov3: An incremental improvement", Computer Vision and Pattern Recognition, 161-168.

Sunil and Gagandeep, (2019), "Study of Object Detection Methods and Applications on Digital Images", International Journal of Scientific Development and Research (IJSDR), 4(5): 491-497.

Tiwari. M and Singhai. R, (2017), "A Review of Detection and Tracking of Object from Image and Video Sequences", International Journal of Computational Intelligence Research, 13(5): 745-765.

Yang. Z and Nevatia. R, (2016), "A multi-scale cascade fully convolutional network face detector," in 23rd International Conference on Pattern Recognition (ICPR) - IEEE, 1-6.