Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

Object Identification using Deep Learning

Suraj Kunwar
BE ELE NHITM
Thane, India

Suhas Waghmare
Asst. Professor
NHITM Thane, India

Saurabh Dudwadkar
BE ELE NHITM
Thane, India

Digvijay Deshmukh
BE ELE NHITM
Thane, India

Siddhant Survase
BE ELE NHITM
Thane, India

***Abstract*—The application of object detection via deep learning proves to be a vital fragment of deep learning technology, characterized by its distinct and vivid methods of feature learning and representation, which overpowers other traditional object detection methods. The article postulated in this paper firstly discusses the whereabouts of traditional learning methods in object detection, and then builds itself up to elaborating the emergence of object detection methods based on deep learning methods and all the subsequently derived new-world object detection and identification methods via deep learning. The paper explains further on the methods of framework designs and the principles on which the models function, while also analyzing its performance in real-time and the accuracy of detection. The paper also addresses the challenges in deep learning object detection and renders solutions with respect to some references.**

I. INTRODUCTION

In this paper we will discuss various techniques of object detection and will use the most efficient among them. Object detection is a process of identifying the presence of objects in an image. It can be done by using a pre-trained object detector or by training the detector for a specific task. It is a very common task in computer vision and statistical data analysis. The paper will start with an introduction to neural networks, followed by some examples of object detection. Thereafter the contents of the paper will be split up discussing different patterns used in object detection including CNN, R-CNN, Fast R-CNN, Faster R-CNN, YOLO and deep learning approaches to solving it for end-to-end models. Neural networks are models created to simulate the human brain and its general functions, playing a big role in many different fields of science. In pattern recognition, neural networks have been applied to image classification, object detection and face recognition. They are made up of a number of neuron-like units arranged in layers. Each layer is responsible for processing data from the previous layer, which is why they are hierarchical structures. The most important layers, on which all others depend on, are called the 'hidden' layers. There can be any number of hidden layers (1 or more) and there can be any depth (number of neurons). The parameters that describe the connections between neurons in a neural network are called 'weights'. The weights in object detection are the parameters that are used to train the neural network. The weights are usually initialized randomly and then updated during training. The weights in object detection can be updated using a variety of different methods, such as gradient descent, stochastic gradient descent, or momentum.

II. BRIEF OVERVIEW OF DEEP LEARNING

Deep learning is a subset of machine learning, which makes it possible to derive meaning from data. Deep learning uses artificial neural networks that contain many layers and nodes. An artificial neural network is a mathematical construct that takes in information from the layer before it and applies some function to it, eventually outputting another piece of information for the next layer to see in order to teach the network about patterns. The more layers and subsequent nodes in a deep learning neural network, the deeper into abstractions a given algorithm can go. Deep learning networks are a subset of artificial neural networks, which themselves are a subset of machine learning algorithms. Machine learning algorithms take in information, whether that be data or observations, and apply some function to it so that the information can be used to make accurate predictions.

What is deep learning? To answer this question, we must look back at the current landscape of machine learning and artificial intelligence (AI). Linear regression was the first algorithm used to predict continuous values. It was discovered in 1805 by Pierre Laplace and Joseph Fourier. This algorithm has been used not only for prediction but also for analysis. The problem with using linear regression for prediction is the fact that it assumes a linear relationship between input variables and output values. In other words, it takes into account the relationship between two input variables but assumes a linear relationship is present. To solve this problem, we turned to neural networks. The first neural network was created by Frank Rosenblatt in 1963. It took input values and output values as binary numbers and compared numerical patterns to determine results. This was called perceptrons and was used for pattern recognition.

Deep learning is not a new algorithm but rather a way of thinking about machine learning. To put it simply, deep learning is a subset of machine learning that creates meaningful neural networks. A deep learning neural network, according to some people (e.g., Deeplearning4j), is one that contains 100 layers or nodes. However, it is also possible for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

there to be exactly one neuron at each layer but with a very large number of hidden neurons at the end which are not yet trained but are used to make predictions given input. The reason this method isn't typically called deep learning is because these "deep" networks can have an arbitrary number of layers and nodes, so long as the prediction functions at each node can be computed linearly using simple mathematical equations and linear algebraic techniques (landscape of machine learning and AI. The deep learning package is part of the Python programming language, which means that this functionality can be easily accessed from any framework with Python support.

III. EMERGENCE OF OBJECT DETECTION BASED ON DEEP LEARNING

The object detection method adopts modes such as region selection, feature extraction and classification. The region selection process is based on a particular strategy depending on the desired results whereas feature extraction can be achieved by CNN (convolutional neural network) and classification is done by traditional SVM or the special neural network. DNN and overfeat are the earlier typical modes of deep learning applied on object detection which draw ups the curtain for applying deep learning in object detection The object detection by DNN has designed a complex region selection strategy, which makes use of hierarchical structure to generate multiple binary receptive fields of convolutional neural network to do region selection, which increases the computation cost and does not consider the memory. The DNN has shown that the global structure of objects can be detected by using the method, but it does not work for objects that are broken or deformed.

The method based on Overfeat will be more flexible in describing the spatial structure of objects, but it is difficult to resolve the boundary between different objects. After all, the main goal of an object detection algorithm is to detect an individual object and its category (e.g., pedestrian, vehicle). A general approach is to treat each part of an object as a separate instance. Neurons together with their nonlinear functions as features. The Overfeat adopts the strategy that extracts the feature from each convolutional layer together and then use the support vector machine to classify it. It only adopts a local region selecting strategy and cannot deal with complex scenes containing several objects. Different from DNN and Overfeat, recent CNNS (Convolutional Neural Networks) [25] for object detection have used an effective feature extraction method based on selective search and obtained a good performance in PASCAL VOC 2007 competition.

Most of the earlier object detection methods are based on deep learning by employing the sliding window operation adopted by overfeat and this method would result in problems related to data explosion also the model accuracy was not satisfactory. The models which came into existence after this model tried to solve this by making the current model better or putting forth new ideas. The models which currently existed had their drawbacks and to overcome this drawback leads to emergence of new model design of object detection based on deep learning.

IV. DIFFERENT MODELS

A. R-CNN

Classification and localization are the two most important task that are involved in object detection. R-CNN stands for region-based Convolutional Neural Network. Region proposals are the key concept behind R-CNN. Region proposals are used to localize objects within an image. The key to region proposals is the convolutional layer that is used to detect the object. The key concept behind R-CNN is to use a larger network with a smaller network. In order to do this, they need to figure out which parts of the image belong in the object and which parts don't. Region proposal uses Conditional Random Fields (CRF), which are maps for classifying an image. CRF recognizes the location of objects in an image by finding regions that have similar values for certain pixels within an image. This means it works best on images with known objects because it has been preprocessed. In order to make it work on images without an object, they need to gather that information as well. This is done by training on two types of data; background and foreground. The background images are ones that contain no objects in them. The foreground images are preprocessed and contain the object of interest. There are three main steps for building region proposals: The first step is generating a set of high-quality region proposals. The second step is finding a suitable model that can detect the object from the image. Finally, integrating the model with a multi-stage framework to achieve better detection results at higher speeds than existing methods. But it still takes a huge amount of time to train the network and it cannot be implemented real time as it takes around 47 seconds for each test image.

B. Fast R-CNN

Due to some of the drawbacks possessed by R-CNN a faster algorithm was developed called Fast R-CNN. The approach is similar to the R-CNN algorithm. Here, the input image is fed to CNN instead of the region proposals and a convolutional feature map is generated. From the convolutional feature map, regions of proposals are identified and are wrapped into squares and thereby using a Roi pooling layer they are reshaped into a fixed size so that it can be fed into a fully connected layer. From the RoI feature vector, a layer called SoftMax layer is used to predict the class of the proposed region and also the offset values of the bounding boxes. Due to the fact that we don’t have to feed 2000 region proposals to the convolutional neural network every time Fast R-CNN is faster than R-CNN.

C. Faster R-CNN

In the R-CNN family of algorithms, the evolution between algorithms was usually in terms of computational efficiency, reduction in test time and improvement in performance. R-CNN and Fast R-CNN both used region proposal algorithms, e.g., the selective search algorithm which requires around 2 seconds per image and runs on CPU computation. This time was reduced from 2s to 10ms by Faster R-CNN which makes use of another convolutional network (RPN) to generate the region proposals. Along with reducing the time it also allows the region proposal stage to share layers with the following detection stages, causing an overall improvement in feature representation. Faster R-CNN refers to the detection pipeline

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

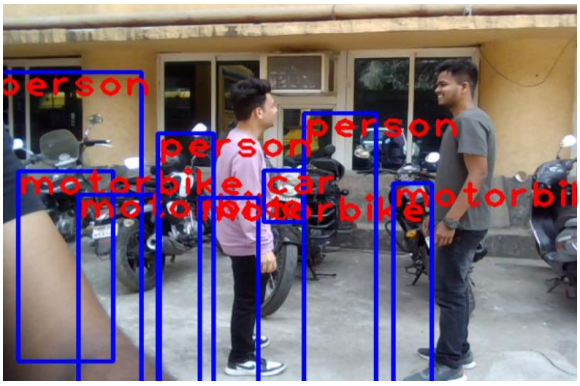
which comprises of RPN which is used as region proposal algorithm, and Fast R-CNN as a detector network.

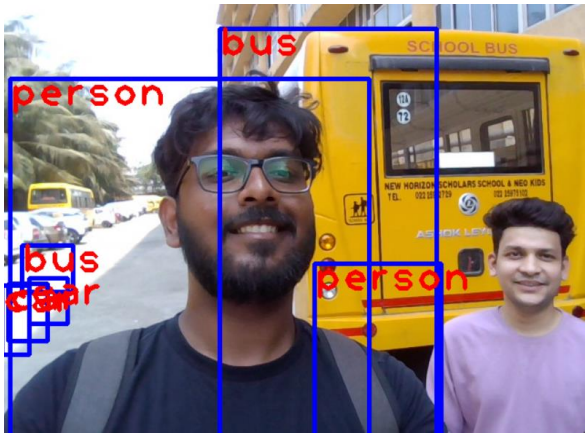
D. YOLO

YOLO is the state of art object detection algorithm and it is so fast that it can process 45 frames per second and it is almost a standard way of detecting objects in the field of computer vision. YOLO stands for You only look once and it out performed all the previously mentioned object detection algorithms. Earlier algorithms made use of a two stage approach for object detection while it gave accurate results it slowed down the process as it had to make iterations over the same image. Compared to the other object detection algorithms YOLO proposes the use of an end- to-end neural network that makes predictions of bounding boxes and class probabilities all at once.

The YOLO algorithm divides the input image into N grids, each of these grids have MxM dimension. All of these N grids are responsible for detection and localization of the object it contains. Correspondingly bounding boxes are predicted by these grids along with object label and probability of the object being present in the cell. As both detection and recognition are handled by the cells the computation time is greatly reduced but it brings a lot of predictions which are duplicate of each other due to multiple cells predicting the same object with different bounding box predictions. This issue is dealt with by the use of Non-Maximal Suppression. In Non-Maximal Suppression, all the bounding boxes that have low probability scores are suppressed. Yolo is able to achieve this by choosing the highest probability score. Following this, the bounding boxes having the largest Intersection over Union with the current high probability are suppressed. This step is continued till we obtain the final bounding box.

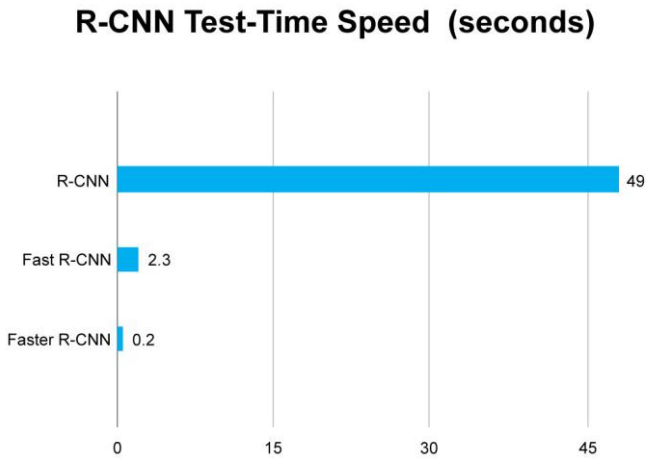
V. RESULT





V. CONCLUSION

This paper firstly introduces all the classical object detection methodologies and compares it with deep learning methods. Then it compares all the existing models based on R- CNN families and states their advantages, disadvantages over each other and also explain why they can't be used in real time due to their large computational time and lastly it discusses about YOLO which we have implemented due to its phenomenal computational speed. In order of magnitude yolo is (45 frames per second faster) than other object detection algorithms.



REFERENCES

[1] D. Erhan, C. Szegedy, A. Toshev, et al, "Scalable object detection using deep neural networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2155-2162. 727

[2] A. Borji, M. M. Cheng, H. Jiang, et al, "Salient object detection: A benchmark," IEEE Transactions on Image Processing, vol. 24, Dec 2015, pp. 5706-5722.

[3] Y. Tian, P. Luo, X. Wang, et al, "Deep learning strong parts for pedestrian detection," 2015 IEEE International Conference on Computer Vision, 2015, pp. 1904-1912.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

[4] P. AHMADVAND, R. EBRAHIMPOUR AND P. AHMADVAND, "HOW POPULAR CNNS PERFORM IN REAL APPLICATIONS OF FACE RECOGNITION, 2016 24TH TELECOMMUNICATIONS FORUM (TELFOR), 2016, PP.1-4.

[5] W. OUYANG, X. WANG, X. ZENG, ET AL, "DEEPID-NET: DEFORMABLE DEEP CONVOLUTIONAL NEURAL NETWORKS FOR OBJECT DETECTION," 2015 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, PP. 2403-2412.

[6] P. M. MERLIN, D. J. FARBER, "A PARALLEL MECHANISM FOR DETECTING CURVES IN PICTURES," IEEE TRANSACTIONS ON COMPUTERS, VOL. C-24, JAN 1975, PP. 96-98.

[7] N. SINGLA, "MOTION DETECTION BASED ON FRAME DIFFERENCE METHOD," INTERNATIONAL JOURNAL OF INFORMATION & COMPUTATION TECHNOLOGY, VOL. 4, NO. 15, 2014, PP. 1559-1565.

[8] D. S. LEE, "EFFECTIVE GAUSSIAN MIXTURE LEARNING FOR VIDEO BACKGROUND SUBTRACTION," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, MAY 2005, PP. 827-832.

[9] B. K. P. HORN, B. G. SCHUNCK, "DETERMINING OPTICAL FLOW," ARTIFICIAL INTELLIGENCE, VOL. 17, 1981, PP. 185-203.

[10] J. L. BARRON, D. J. FLEET, S. S. BEAUCHEMIN, ET AL, "PERFORMANCE OF OPTICAL FLOW TECHNIQUES," 1992 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1992: PP. 236-242.