# Object Detection using Deep Learning

**Pranita Jadhav[1], Vrushali Koli[2], Priyanka Shinde[3], Dr. M.M. Pawar[4]**

*[1,2]Graduate Student, SVERI's College of Engineering, Pandharpur*
*[3]Professor, SVERI's College of Engineering, Pandharpur*

---

**Abstract:** *The object recognition based on deep learning is an important application in deep learning technology which is distinguished by its strong capability of feature learning and feature representation compared with the traditional object recognition methods. The paper introduces the classical methods for object recognition and expounds the relation and difference between the classical methods and the deep learning methods in object detection. It introduces the emergence of the object recognition methods based on deep learning and elaborates the most typical methods nowadays in the object detection via deep learning. The paper focuses on the framework design and the working principle of the models and examines the model performance in the real-time and the accuracy of detection. Ultimately, it discusses the challenges in the object detection based on deep learning and offers some solutions for reference.*

***Keywords: Object detection, Deep learning, Framework design, Model analysis, Performance analysis***

## Introduction

Object detection has already been the remarkable research direction and the focus in the computer vision which can be applied in the automatic car, robotics, video surveillance and pedestrian detection. The exposure of deep learning technology has changed the traditional ways of object identification and object detection. The deep neural network has the vigorous feature depiction capacity in image processing and is usually used as the feature extraction module in object detection. The deep learning models don't required special hand engineered features and can be designed as the classifier and regression device. Therefore, the deep learning technology is of significant prospect in the object detection. The problem statement of object detection is to determine where objects are actually located in a given image (object localization) and detect it. So the pipeline of traditional object detection models mainly divided into three stages: informative region selection, feature extraction and detection.

1. Informative region selection: As various objects appear in any positions of the image and have different aspect ratios or sizes. It is a compulsory task 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 huge number of candidate windows, it is computationally expensive and produces too many redundant windows. However, if only a possibility of unsatisfactory regions in the given image.

2. Feature Extraction: To identify different objects, we need to extract visual features which can provide a semantic and robust representation. This is due to the features which can produce representations associated with complex cells in human brain. However, due to the diversity of appearances, illumination conditions and backgrounds, it's difficult to manually design a robust feature descriptor for perfectly describing all kinds of objects.

3. Detection: A detector distinguishes a target object from all the other categories and to make the presentations more hierarchical and informative for visual recognition.

Usually, the Supported Vector Machine (SVM) and Deformable Part-based Model (DPM) are good choices. Among these classifiers, the DPM is a flexible model by combining object parts with deformation cost to handle severe deformations. In DPM, with the support of a graphical model, carefully designed low-level features and kinematically inspired part decompositions are combined and discriminative learning of graphical models permits for building high-precision part- based models for a variety of object classes.

## Literature Review

**CeLiet. al[1]** presents an object detector based on deeplearning of smallsamples. The proposed nodel uses the semantic relevance of objects to improve the accuracy of weak feature objects in complex scenarios. **Cong Tanget. al[2]** focuses on the framework design and the working principle of the models and analyzes the model performance in the real-time and the accuracy of detection. **Christian Szegedyet.al[3]** presents a simple and yet powerful formulation of object detection as a regression problem to object bounding box masks. It defines a multi-scale inference procedure that produces high- resolution object detections at a low cost by a few network applications. **XiaogangWanget.al[4]** provides a overview of deep learning and focus on the applications in object recognition, detection, and segmentation which are the key challenges for computer vision and have numerous applications to images and videos.**ShuaiZhanget.al[5]** proposes a framework for achieving tasks in a nonoverlapping multiple camera network. A new object detection algorithm using mean shift (MS) segmentation is introduced and objects are further separated with

The help of depth information derived from stereo fixed number of sliding window templates are applied there is vision. It is also possible for supervised learning in implementing the problem using Decision trees or more likely SVM in deep learning which is implemented in **Malay Shahet.al[6]**. **XinyiZhouet. al[7]** deals with the field of computer vision mainly for the deep learning in object detection task. There is a simple summary of the datasets and deep learning algorithms used in computer vision. **Zhong-QiuZhaoet.al[8]** provides a detailed review of deep learning based object detection frameworks which handle different sub-problems, such as clutter and low resolution, with different degrees of modifications on R-CNN.**Sandeep Kumaret.al [9]** deals with the easynet model where the detection predictions with a Single network is possible. The easynet model looks at the whole image at test time so the predictions are informed by global context.**Adami Fatima Zohraet. al[10]** focuses on detection and recognition of vehicles from a video stream. This method gives better results in terms of sprecision, detection and classification where accuracy of 99.2% is achieved.

**Introduction to Deep Learning**

The deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for decision making. It is a subset of machine learning in artificial intelligence which has networks capable of learning unsupervised from data.

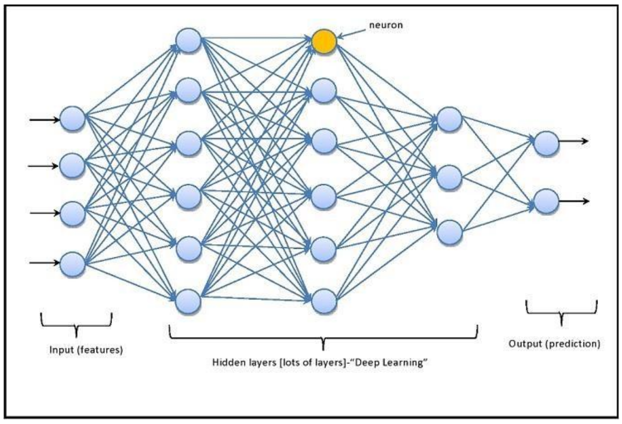


Fig.1 Architecture of Deep Learning

*A.CNN:* The CNN is nothing but Convolution Neural Network. A CNN includes an input, an output layer and multiple hidden layers. The hidden layers of a CNN have a series of convolutional layers which convolve with a multiplication or other dot product. The activation function is generally a RELU layer which is followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers as their inputs and outputs are masked by the activation function and final convolution. It is generally a sliding dot product or cross-correlation. These CNN layers convolve the input and pass the result to the next layer.

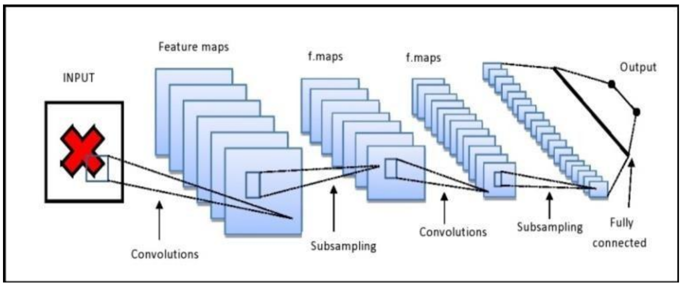


Fig.2 CNN

*B.YOLO:* The YOLO (You Only Look Once) uses deep learning and convolutional neural networks (CNN) for object detection. It only needs to "see" each image once. It allows YOLO to be one of the fastest detection algorithms. It can detect objects in real time up to 30 FPS. For the detection, the image is divided in a grid of S*S (left image). Each cell will predict N possible bounding boxes and the level of probability of each one of them. This means S*S*N boxes are calculated.

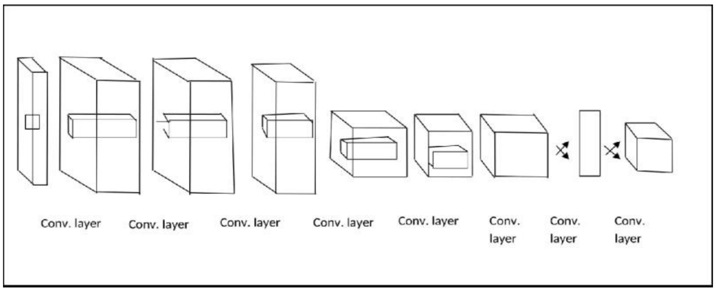


Fig.3 YOLO

*C.R-CNN: The* R-CNN is nothing but Region based CNNs. The R-CNN model first selects several proposed regions from given image and then label their categories and bounding boxes. It uses a CNN to perform forward computation to extract features from each proposed area. Then we use the features of each proposed region to predict their categories and bounding boxes.

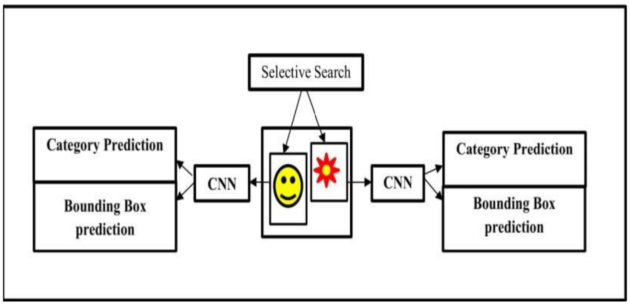


Fig.4R-CNN

Methodology

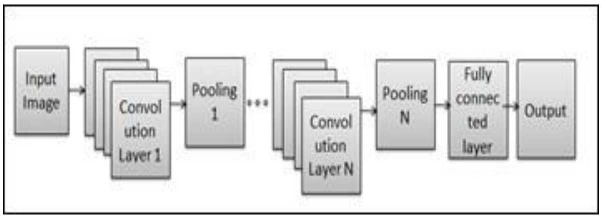


Fig.5 Basic Architecture of CNN

Basic architecture of CNN model consists of

A. Input Image

B. Convolution Layer

C. Pooling Layer

D. Fully Connected Layer

A. Input image

- The input image is the image given to the model to check the output by performing various functions on it. It is given to the block named as convolution layer.

B. Convolution Layer

- Convolution is the first layer to extract features from the given input image. Convolution restores the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation which takes two inputs such as image matrix and a filter.

C. Pooling Layer

- The section of pooling layers reduces the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling that reduces the dimensionality of each map without changing the important information. Spatial pooling can be of different types:

1) Max Pooling

2) Average Pooling

3) Sum Pooling

Max pooling take the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

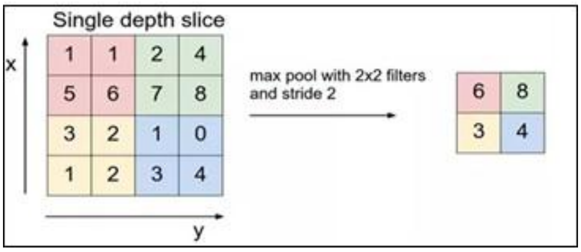


Fig.6 Pooling

D. Fully Connected Layer

- The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like neural network.

- The fully connected layer combines all the features together to create a model. Finally an activation function such as softmax or sigmoid is used to classify the outputs as cat, dog, car, truck etc.

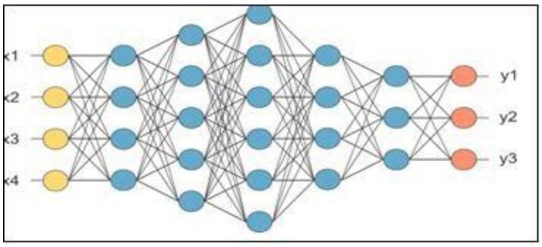


Fig. 7 After pooling layer, flattened as FC layer

E. Output

- The output image provides the result of the model. On the basis of the fully connected layer the output is carried out.

**Result**

The performance of a model for object detection is evaluated using the precision and recall across each of the best matching bounding boxes for the known objects in the image.


Fig.5 Object Detection result in case of animals

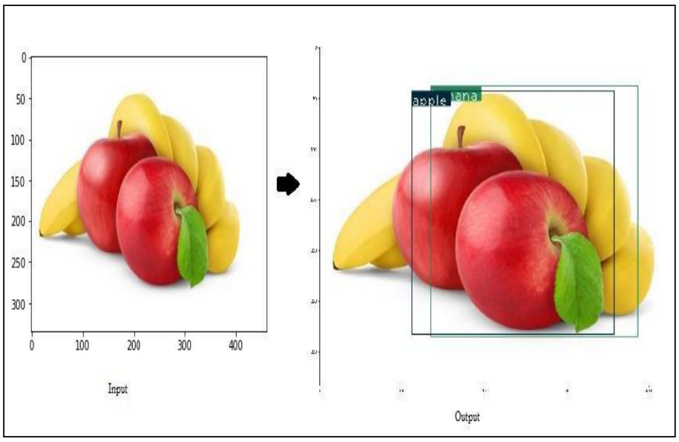


Fig.6 Object Detection result in case of fruits

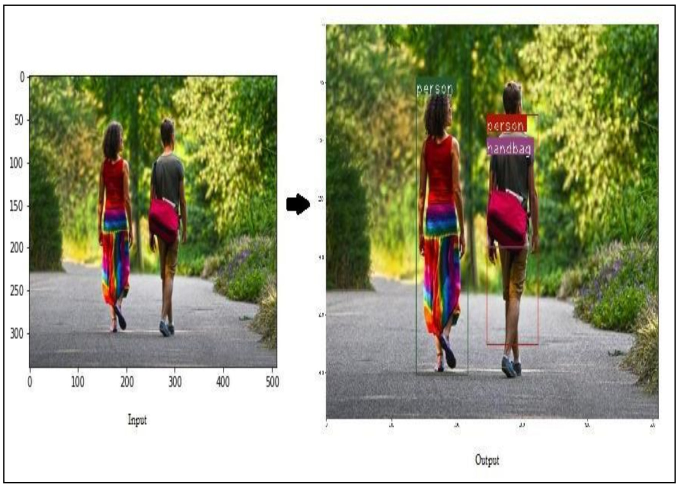


Fig.7 Object Detection result in case


Fig.8 Object Detection result in case of vehicles

## Conclusion

This paper firstly introduces the classical methodologies of object detection in deep learning. Then it clarifies the ideas of model design and the limitations of deep learning method by overviewing the early object detection methods based on deep learning. Afterwards, it elaborates on the common object detection model based on deep learning. Finally, this paper makes a further analysis of the challenges in object detection based on deep learning, and offers some solutions for reference.

## Acknowledgment

We give special thanks to our respected guide Dr. M.M. Pawar for her constant interest and constant encouragement throughout the completion of our project.

We are also equally indebted to our principal Dr. B.P. Ronge for his valuable help whenever needed.

We express deep gratitude to all staff members who lend us their valuable support and co-operation to enable us to complete our project successfully.

## References

[1] Ce Li, Yachao Zhang and YanyunQu, “Object Detection Based on Deep Learning of Small Samples,” International Conference, pp.1-6, March 2018.

[2] Cong Tang, YunsongFeng, Xing Yang, Chao Zheng and Yuanpu Zhou, “The Object Detection Based on Deep Learning,” International Conference, pp.1-6, 2017.

[3] Christian Szegedy, Alexander Toshev and DumitruErhan, “Deep Neural Networks for Object Detection,” IEEE, pp.1-9, 2007.

[4] Xiaogang Wang, “Deep Learning in Object Recognition, Detection, and Segmentation,” IEEE, pp.1-40, Apr. 2014.

[5] Shuai Zhang, Chong Wang and Shing-Chow Chan, “New Object Detection, Tracking, and Recognition Approaches for Video Surveillance Over Camera

Network," IEEE SENSORS JOURNAL, vol. 15, no.69, pp. 1-13, May 2015.

[6] Malay Shah and Prof. RupalKapdi, "Object Detection Using Deep Neural Networks," International Conference, IEEE, pp.1-4, 2017.

[7] Xiao Ma, Ke Zhou and JiangfengZheng, "Photo-Realistic Face Age Progression/Regression Using a Single Generative Adversarial Network," Neurocomputing, Elsevier B.V., pp.1-16, July 2019.

[8] Zhong-Qiu Zhao, PengZheng, Shou-taoXu and Xindong Wu, "Object Detection with Deep Learning: A Review," IEEE, pp.1-21, 2019.

[9] Sandeep Kumar, AmanBalyan and ManviChawla, "Object Detection and Recognition in Images," IJEDR, pp.1-6, 2017.

[10] Adami Fatima Zohra, SalmiKamilia, Abbas Faycal and SaadiSouad, "Detection And Classification Of Vehicles Using Deep Learning," International Journal of Computer Science trends and technology(IJCST), vol. 6, pp. 1-7, 2018.