Exploration and Learning of Object Detection using Deep Learning

Ms. Kalyani D. Bawangarh

[#1]*Assistant Professor, Department of Computer Science, Dr. S. C. Gulhane Prerna College of Commerce, Science, and Arts, Nagpur, MS. (India)*

Abstract **This review gives an overview of the evolution of DL with a focus on image segmentation and object detection in convolutional neural networks (CNN) [1]. The problem discussed in this paper is object detection using deep neural networks, especially convolution neural networks. Object detection was previously done using only conventional deep convolution neural networks whereas using regional-based convolution networks increases the accuracy and also decreases the time required to complete the program [2]. Rather than using object identification techniques, it analyses the different features of images and generates an intelligent and effective comprehension of images, similar to how human vision works. In this study, we will begin with a brief introduction to deep learning and well-known object identification systems such as CNN (Convolutional Neural Network), R-CNN, RNN (Recurrent Brain Network), Faster RNN, and YOLO (You Only Look Once). Then, we focus on the suggested object detection model architecture**.

Keywords- Deep learning, Object Detection, Convolutional Neural Network, YOLO (You Only Look Once)

I. INTRODUCTION

Object detection is inextricably linked to the field of computer vision. Object recognition allows you to recognize instances of various items in photos, movies, or video recordings. A fundamental computer vision problem known as object detection is locating and recognizing things within a frame of an image or a video. Some of the real-world applications include autonomous driving, monitoring security, transportation surveillance, robotic Some of the real-world applications include autonomous driving, monitoring security, transportation surveillance, and robotic vision and more [3]. Convolutional neural networks (CNNs), in particular, have transformed the field of object detection by enabling extremely precise and effective detection techniques.

Although the human eye can distinguish a given visual, including its content, location, and neighboring visuals, by interacting with it, computer vision-enabled robotic devices can be slow and imprecise at times. Any advancement in this subject will boost efficiency and performance, possibly creating the path for more intelligent systems analogous to humans. As a result, technologies like recent advances that allow individuals to perform jobs with little to no conscious effort will undoubtedly make our lives easier.

Object detection is a critical component of self-driving car perception systems. These systems use object detection to identify and track pedestrians, vehicles, road signs, traffic lights, and other obstacles around the vehicle. This information is crucial for making driving decisions and ensuring the safety of passengers and pedestrians. Computer vision and object detection are very important and crucial fields in machine learning, and they are expected to help unleash the hidden potential of general-purpose robotic systems in the future.

Artificial neural networks are the foundation of the machine learning subfield known as deep learning. It can recognize intricate links and patterns in data. We don’t have to explicitly program everything in deep learning. Due to improvements in processing power and the accessibility of massive datasets, it has grown in popularity recently. Because it is built on deep neural networks (DNNs), often referred to as artificial neural networks (ANNs). These neural networks are built to learn from massive amounts of data and are modeled after the structure and operation of organic neurons in the human brain. Because of the advancement of Object detection in Deep Learning, it can be further classified into two models [4] Model based on region proposal; and [5] Model based on regression/Classification.

Over time, numerous techniques have been put forth to address the problem of object identification. These methods are centered on providing solutions at various phases. These central steps specifically include object detection, location, classification, and recognition. Along with the long-term growth of current technology, these Techniques have been dealing with challenges including output accuracy, resource cost, processing speed, and complexity issues. With the development of the primary Convolutional Neural Network (CNN) algorithms during the 1990s roused by Yann LeCun et al. [6] and very important research and innovations like AlexNet [7], CNN algorithms have been fit for giving answers for the item recognition issue in various methodologies. CNN algorithms are suitable for providing solutions to the item recognition problem using a variety of techniques. Algorithms with an optimization focus are designed to make human interaction easier while increasing recognition and detection speed and accuracy.

II. OVERVIEW OF OBJECT DETECTION

Object detection is a computer vision task that involves identifying and localizing objects within an image or a video. Convolutional Neural Networks (CNNs) have revolutionized object detection by enabling the development of highly accurate and efficient detection systems. Some of the popular object detection algorithms are Region-based Convolutional Neural Networks (RCNN), Faster-RCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy [8]. Here's an overview of object detection using CNNs:

A. *Problem Statement:* Object detection involves two primary tasks: classifying the objects present in an image and accurately localizing their positions with bounding boxes. This is different from image classification, which only assigns a single label to an entire image.

B. *CNN Basics:* Convolutional Neural Networks (CNNs) are a class of deep learning models designed to process and extract features from grid-like data, such as images. CNNs are composed of layers that learn hierarchical representations of visual features.

C. *Key Components of Object Detection using CNNs:*

1) *Convolutional Layers:* These layers perform convolutions on the input image to detect low-level features like edges, textures, and corners.

2) *Pooling Layers:* Pooling layers reduce the spatial dimensions of the feature maps while retaining important information. Common pooling operations include max pooling and average pooling. Spatial pooling can be of different types: Max Pooling, Average Pooling, and Sum Pooling [9].

3) *Fully Connected Layers:* These layers are often used at the end of the network to make class predictions and refine object localization.

D. *Object Detection Architectures:*

1) *Single Shot MultiBox Detector (SSD):* SSD is a popular architecture that predicts object classes and bounding box coordinates at multiple scales within a single forward pass. It combines predictions from different layers to handle objects of various sizes.

2) *Faster R-CNN:* Faster R-CNN introduced Region Proposal Networks (RPNs), which generate potential object proposals before refining their locations and classes. This two-stage approach achieves high accuracy but is slower than SSD.

3) *YOLO (You Only Look Once):* YOLO divides the input image into a grid and predicts bounding boxes and class probabilities directly using a single network pass. YOLO is known for its real-time processing capabilities. YOLO is an object detection algorithm entirely different from the district-based algorithms seen previously [10].

4) *EfficientDet:* An evolution of object detection architectures, EfficientDet focuses on achieving high accuracy while maintaining efficiency. It uses a compound scaling method to balance model size and computational cost.

E. Training Object Detection Models:

1) Loss Functions: Object detection models are trained using a combination of classification loss (e.g., softmax cross-entropy) and localization loss (e.g., smooth L1 loss) to ensure accurate object detection and bounding box localization.

2) Data Augmentation: Techniques like random cropping, flipping, and color jittering are applied to increase the diversity of the training data and improve model generalization.

F. Evaluation Metrics:

1) Intersection over Union (IoU): Measures the overlap between predicted and ground-truth bounding boxes.

2) Mean Average Precision (mAP): A commonly used metric that calculates precision-recall curves for different object categories and averages the results to provide an overall performance measure.

G. Transfer Learning: Pre-trained CNN models (e.g., VGG, ResNet, etc.) trained on large image datasets like ImageNet can be fine-tuned for object detection tasks. This transfer learning approach helps improve training efficiency and detection performance.

H. Applications: Object detection using CNNs has numerous applications, including autonomous vehicles, surveillance, robotics, medical imaging, and more.

III. CONVOLUTION NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNN) are a sort of multi-layer neural network designed to extract visual patterns from pixel images. CNN refers to 'convolution' as the mathematical function. It is a form of linear operation in which two functions are multiplied to produce a third function that expresses how the shape of one function can be affected by the other. Simply put, two pictures represented by two matrices are multiplied to provide an output that is used to extract information from the image. CNN is similar to other neural networks, but they add a layer of complexity to the equation because they use a sequence of convolutional layers. CNN cannot work in the absence of convolutional layers.

A. Typical CNN Architecture- The ConvNet's task is to reduce the size of the images while maintaining essential components for making accurate predictions. This is crucial for developing an architecture that can learn features and scale to enormous datasets.

The three layers that make up a convolutional neural network, or ConvNets in short, are its building blocks.

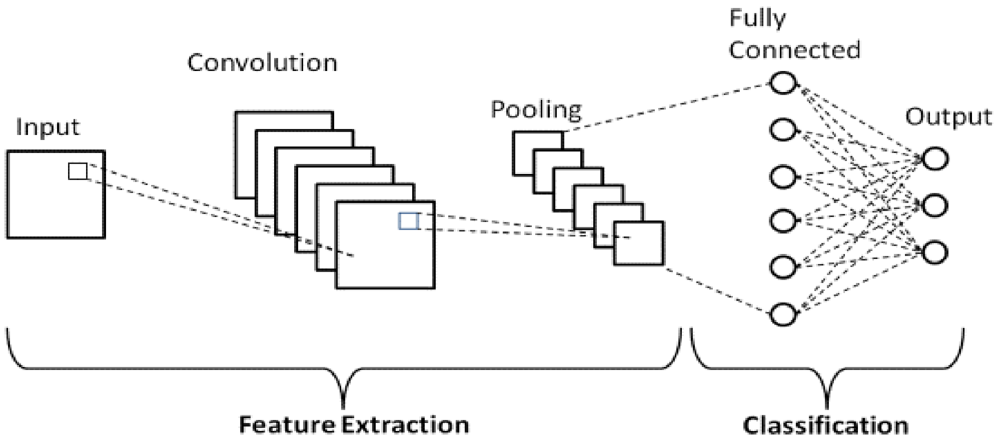


Figure 1: CNN Layer Diagram

1) *Convolutional Layer (CONV):* This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered a matrix of pixel values. Consider the following 5x5 image whose pixel values are either 0 or 1. There's also a filter matrix with a dimension of 3x3. Slide the filter matrix over the image and compute the dot product to get the convolved feature matrix. Convolutional layers also contain the Non-linear activation function, which is a significant component in addition to convolution. A non-linear activation function is applied to the outputs of linear processes like convolution. Although they are mathematical representations of biological neuron activity, smooth nonlinear functions like the sigmoid or hyperbolic tangent (tanh) function have previously been used. The most popular non-linear activation function today is the rectified linear unit (ReLU). f(x) max(0, x)

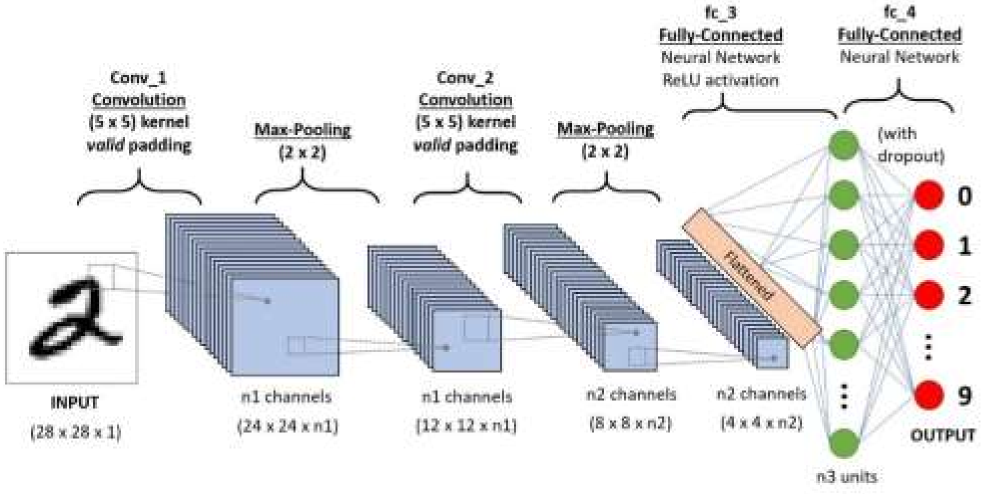


Figure 2: Convolutional Layer

2) *Pooling Layer (POOL):* This layer is in charge of reducing dimensionality. It aids in reducing the amount of computing power required to process the data. Pooling can be divided into two types: maximum pooling and average pooling. The maximum value from the area covered by the kernel on the image is returned by max pooling. The average of all the values in the part of the image covered by the kernel is returned by average pooling.

3) Fully Connected Layer (FC): We flattened our matrix into a vector and fed it into the layer we refer to as the FC layer, which is fully connected and resembles a neural network. All the features are

combined to form a model in the fully linked layer. To categorize the outputs as cat, dog, automobile, truck, etc., an activation function like softmax or sigmoid is utilized in the last step.

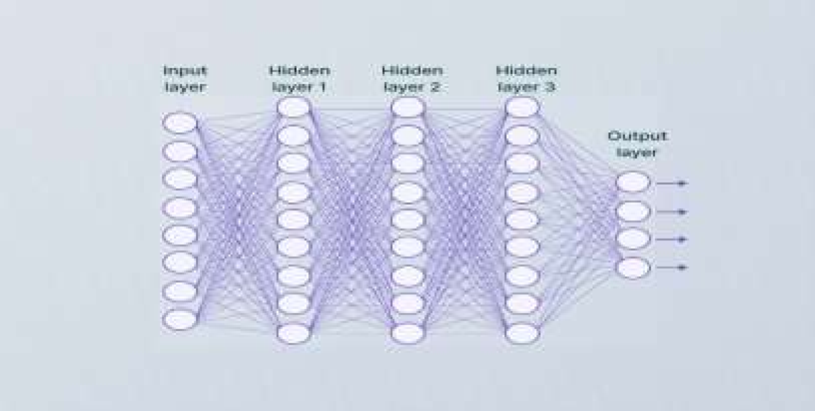
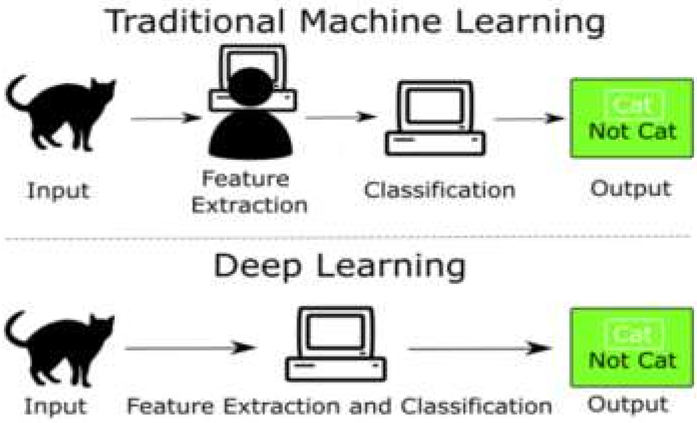


Figure 3: Fully Connected Network

IV. RESULT

The output image provides the result of the model. Based on the fully connected layer the output is carried out. The performance of a model for object detection is evaluated using precision and recall across each of the best matching bounding boxes for the known objects in the image.



A.

Figure 5: Image Extraction and Classification

CONCLUSION

In this paper, the investigation of item identification is a significant undertaking inside the domain of PC vision and profound learning. Through this study, we have acquired experiences in the strategies and procedures that empower machines to recognize and find objects inside pictures and video outlines. Object location has extensive applications across ventures and areas, reshaping how we collaborate with innovation and our current circumstance

REFERENCES

[1] A. Boukerche, & H. Zhijun. "Object detection using deep learning methods in traffic scenarios" ACM Computing Surveys (CSUR), vol. 54.2, 2021, pp. 1-35.

[2] Shah, Malay & R. Kapdi "Object detection using deep neural networks", International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, (2017).

[3] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection" IEEE, vol. 7, 2019, pp.128837–128868.

[4] J. Redmon, S. Divvala, R. Girshick & A. Farhadi, "You only look once: Unified, real-time object detection,", 2016, doi: 10.1109/CVPR.2016.91.

[5] Wu, R. B. Research on Application of Intelligent Video Surveillance and Face Recognition Technology in Prison Security. China Security Technology and Application., vol. 6, 2019, pp.16-19.

[6] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," IEEE 2nd Int. Conf. Big Data Anal. ICBDA, 2017, pp. 721– 724, doi: 10.1109/ICBDA. 8078730.

[7] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," J. Phys. Conf. Ser., vol. 1004, no. 1, 2018, doi: 10.1088/1742-6596/1004/1/012029.

[8] G. Chandan, A. Jain, and H. Jain. "Real-time object detection and tracking using Deep Learning and OpenCV " International Conference on Inventive Research in computing applications(ICIRCA), IEEE, 2018.

[9] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, vol. 07 no. 09, 2020.

[10] Tomar, Pradyuman, and S. Haider "A Study on Real-Time Object Detection using Deep Learning" INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), 2022.

[11]https://www.researchgate.net/publication/336805909/figure/fig1/AS:817888827023360@1572011300751/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26.ppm

[12]https://saturncloud.io/images/blog/a-cnn-sequence-to-classify-handwritten-digits.jpg

[13]https://assets-global.websitefiles.com/5d7b77b063a9066d83e1209c/ 627d1225cb1b3d197840427a_60f040a887535b932a3b2b6e_cnn-hero%2520(1).png

*[14]https://miro.medium.com/v2/resize:fit:1400/format:webp/1*IhtfcoUyjMR33Jq217YRRg.png*