ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR715

Object Detection Using CNN

Dr.P.Bhaskar Naidu[1]; Pulakanam Anusha[2]; Gothula Naveena[3]; Thota Anusha[4]; Chimakurthi Balaji[5]
[1]Professor, Department of Computer Science and Engineering,
[2,3,4,5]Department of Information Technology,
QIS College of Engineering and Technology(A),
Vengamukkapalem, Ongole-523272, Andhra Pradesh, India.

Abstract:- Object detection system using Convolutional Neural Network(CNN) that can accurately identify and classify objects in videos. The purpose of object detection using CNN to enhance technology such as security cameras, smart devices by enabling them to identify and understand objects in videos. Object detection using CNN is a fascinating filed in computer vision. Detection can be difficult since there are all kinds of variations in orientation, lighting, background that can result in completely different videos of the very same object. Now with the advance of deep learning and neural network, we can finally tackle such problems without coming up with various heuristics real-time. The project "Object detection using CNN for video streaming" detects objects efficiently based on CNN algorithm and apply the algorithm on image or video data. In this project, we develop a technique to identify an object considering the deep learning pre-trained model MobileNet for Single Shot Multi-Box Detector (SSD). This algorithm is used for real-time detection and for webcam, which detects the objects in a video stream. Therefore, we use an object detection module that can detect what is in the video stream. In order to implement the module, we combine the MobileNet and the SSD framework for a fast and efficient deep learning-based method of object detection. The main purpose of our research is to elaborate the accuracy of an object detection method SSD and the importance of pre-trained deep learning model MobileNet. The experimental results show that the Average Precision (AP) of the algorithm to detect different classes as car, person and chair is 99.76%, 97.76% and 71.07%, respectively. The main objective of our project is to make clear the object detecting accuracy. The existing methods are Region Based Convolutional Neural Network(R-CNN) and You Only Look Once(YOLO).R-CNN could not pushed real time speed though its system is updated and new versions of it are deployed and YOLO network is popular but YOLO is to struggle to detect objects grouped close together, especially smaller ones. To avoid the drawbacks of these methods we proposed this model which included single shot multi-box detector (SSD), this algorithm is used for real time detection and Mobile-Net architecture.

Keywords:- *Computer Vision, Mobilenet, SSD(Single Shot Multi-Box Detector),Object Detection, Accuracy, Efficiency.*

I. INTRODUCTION

CNN stands for Convolutional Neural Network. It's a type of deep learning algorithm commonly used for object detection and computer vision tasks. They use convolutional layers to extract features from input videos and make predictions based on those features. CNNs have been very successful in various applications like object detection, image classification, and even in natural language processing tasks. This research paper is all about computer vision, which helps machines understand visual data like humans. It's used in various fields like object detection and video analysis[1]. By using Python and OpenCV, we aim to analyze objects in videos comprehensively. We'll combine traditional methods with advanced deep learning models to solve real-world problems efficiently[2]. Our main focus will be on tasks like spotting objects, recognizing them, and working with images. We'll start by preparing images for analysis through steps like resizing and noise reduction.

Our goal is to advance computer vision technology through this project. Computer vision is like giving eyes to computers so they can understand and analyze images just like humans do[3]. It's super important for things like augmented reality, spotting objects, sorting images, and analyzing videos[1]. Python and OpenCV, creating computer vision systems has become easier. The main idea here is to build a system that can look at pictures and figure out what's going on using Python and OpenCV[4].

We're getting images or videos ready for analysis by adjusting their size and reducing any unwanted elements like noise[4]. We'll then pull out important details from the images to understand what objects and patterns are there. Detecting objects is a key part, and we'll use different methods like traditional Haar cascades and modern YOLO for accurate object placement[5]. We'll also work on recognizing these objects by putting them into specific categories using techniques like feature matching or deep learning. This helps in various applications like self-driving cars and security systems[6]. Lastly, we'll enhance image quality, separate the objects we're interested in, and get rid of any distractions to make the system work better[6]. The project will start by getting images in videos ready for analysis through resizing, reducing noise, and organizing them. Then, it will use techniques to pull out important info from the images to represent objects and patterns effectively. Techniques like YOLO and Haar cascades, which are based on deep learning, will be used to detect objects accurately, a crucial task in computer vision[7]. The system will be trained to spot and identify things of interest precise.

II. LITERATURE REVIEW

[Reagan L. Galvez, et.al] A key component of many different computer vision applications, including robot navigation, medical imaging, and video surveillance, is object detection. Convolutional neural networks (CNNs) gained popularity with AlexNet's historic triumph in the 2012 ImageNet competition, although traditional methods like as background reduction, temporal differencing, and contour matching have long been used. Current advancements include new approaches including adaptive low shot transfer detectors, region selection networks, and gating networks, which are all focused on improving object detection accuracy. Moreover, CNNs have been used for tracking visual targets using datasets from ImageNet that comprise positive and negative examples. New approaches to active learning have also been developed, which help find useful examples for training datasets, especially useful for picture classification tasks[1]. Furthermore, given the variety of approaches used to improve object detection skills, artificial neural networks have shown useful in object detection tasks, particularly in the areas of form and color pattern recognition.

[Congtang, et.al] Since object detection is utilized in robotics, autonomous vehicles, video surveillance, and pedestrian detection, among other applications, it is important to computer vision research. The advent of deep learning technology has fundamentally changed how conventional methods of object detection and identification are carried out. Deep convolutional networks emerged as the competition's undisputed leaders after AlexNet's historic triumph in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), proving their superiority in image recognition. The introduction of object detection tasks in 2013 contributed to the acceleration of deep learning's rapid progress in this sector. Because deep neural networks are so good at representing features, they have become essential parts of feature extraction systems for object detection.

[Byungik Ahn] In this study, a field-programmable gate array (FPGA)-based hardware architecture for Convolutional Neural Network (CNN) systems is presented, with a specific emphasis on real-time object detection in video inputs. Because CNNs integrate both feature extraction and classification operations, they are ideal for FPGA implementation because they reduce the need for additional software processing. The architecture maximizes the use of available resources to achieve great performance with comparatively little hardware, leading to solutions that are both economical and power-efficient. The system makes use of cutting-edge techniques to construct hardware-based CNN systems, such as a multi-category recognition technique that switches weight sets to classify objects in the same video stream into different categories. The paper describes the base CNN while highlighting the effectiveness of the feed-forward portion of the CNN, which is usually sufficient for most image recognition tasks.

[Ming Liang, et.al] Convolutional Neural Networks (CNNs) have performed exceptionally well in computer vision tasks in recent years, but they differ from biological visual systems in that they do not include recurrent connections. In order to close this gap, a novel recurrent CNN (RCNN) design is presented. This architecture enhances contextual information that is essential for object recognition by integrating recurrent connections into each convolutional layer to mimic temporal dynamics. RCNN fosters numerous learning routes by achieving depth with a consistent parameter count through temporal unfolding. Tests on industry-standard datasets such as CIFAR-10, CIFAR-100, MNIST, and SVHN highlight the advantage of recurrent structures in object recognition and show that RCNN outperforms more recent models with fewer parameters. Additionally, increasing parameters results in improved performance, confirming the effectiveness of RCNN's recurrent architecture.

[Upulie H.D.I] Advancements in computer vision and object detection are crucial for improving the efficiency and precision of AI systems, bridging the gap between machine and human capabilities. This progress not only facilitates the advancement of intelligent systems but also enables the creation of assistive technologies that streamline tasks and enhance human welfare. Real-time object detection has become essential in automation efforts, aiming to supplement or even replace human tasks. However, the unpredictable nature of image data presents significant challenges for conventional programmed algorithms. To tackle these obstacles, a range of techniques have been proposed, with Convolutional Neural Networks (CNNs) playing a key role in addressing object detection challenges[8]. Despite progress, obstacles such as output accuracy, resource consumption, and processing speed persist. The evolution of algorithms, from R-CNN to YOLO, illustrates ongoing endeavors to confront these hurdles and achieve real-time object detection capabilities. This paper undertakes a comprehensive review of the prominent real-time object detection algorithm, You Only Look Once (YOLO), scrutinizing its architecture, strengths, weaknesses, and implications for future research and development in the field.

[Shijian Tang] The Convolutional neural networks (CNNs) have been extensively deployed in the field of visual recognition. This paper focuses on object recognition within images, aiming to provide class confidence scores and predict bounding boxes for multiple items present in a single image. While Convolutional Neural Networks (CNNs) have become the standard for image classification, our objective extends to various visual recognition tasks, including object detection, segmentation, localization, and even generating phrases from images. While traditional methods like region CNNs (RCNN) combine bounding box regression, CNNs, selective search, and support vector machines (SVM) for object detection, we propose a streamlined approach in our paper. By replacing selective

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR715

search with the edge box technique for region proposal generation, we achieve significantly faster runtimes without sacrificing mean average precision (mAP). Moreover, we simplify the system by eliminating class-specific SVMs, relying instead on the softmax output from the CNN's final layer as our confidence score. Through meticulous training data curation, we ensure precise calibration of the CNN, mitigating any potential performance degradation resulting from the absence of SVMs.

[Aishwarya Sarkale]The field of artificial intelligence is booming, and breakthroughs are happening quickly in a lot of different areas. In particular, picture identification and detection are important sub-domains with many applications. AI-powered cars and facial recognition software are just two examples of the many applications for these technologies. Due to the extensive use of neural networks, many sectors benefit from breakthroughs in other areas in addition to their own unique applications. A subfield of computer vision and image processing called object detection looks for instances of semantic items in digital images, like people, buildings, and cars. Applications for object detection, including as face and pedestrian detection, have been extensively researched. These applications have implications for computer vision domains like picture retrieval and video surveillance.

III. METHODOLOGY

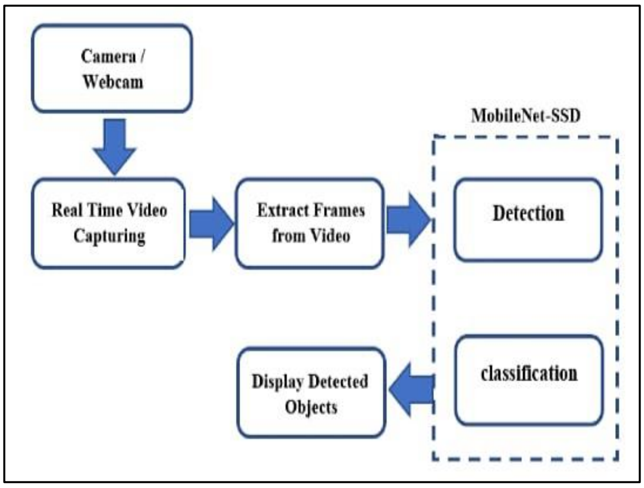

Fig 1 : Process of Proposed System

➢ *Camera/Webcam:*

first, we have to collect the videos which has objects after that upload the video then we have to detect the objects.

➢ *Extract Frames from Video:*

To extract frames from a video, we are using opencv video processsing library .This tool is to break down a video into individual frames or images. then save these frames as separate files for further analysis or use in other applications. By extracting frames, you can analyze the content of each frame, perform image processing tasks, or create. It's a useful technique in various fields like computer vision.

➢ *Object Recognition:*

Recognition goes a step further. It involves identifying and classifying the detected objects into specific categories or labels. Mainly, Recognition gives the what is the object in video or image.Simply, detection is about finding objects, while recognition is about understanding and labeling what those objects . After recognize the objects.

➢ *Object Detection:*

Finding and locating items in a video or image is a technique known as detection. object detection techniques like Faster-RCNN and SSD. Bounding boxes surrounding the objects that have been recognized and their associated class labels are included in the output of an object detection method. Many uses for this data are possible, such as augmented reality, driverless vehicles, and video surveillance.

➢ *Algorithm:*

- Install the TensorFlow.
- Download the MobileNet pretrained model to your machine
- Utilize model prediction by passing in the configuration path to the model.
- Preprocess the image.
- Assign a target label to the object in the image.
- Predicts the probability of target label to each frame in the image.
- The video stream live and video file that we uploaded will perform real time object. looping through the frames we captured from the video stream.

IV. EXISTING METHODS

The existing methods for object detection using cnn .This may include deep learning algorithms like R-CNN and YOLO

➢ *R-CNN:*

RCNN stands for “Region-based Convolutional Neural Network”. It's a kind of deep learning model for identifying objects in pictures. It first generates region proposals. RCNN, region proposals are generated using a selective search algorithm. This algorithm analyzes the image and identifies potential object regions based on similarities in color, texture, and other visual features. These proposed regions are then passed through the convolutional neural network for further analysis and classification. It's a popular approach in computer vision.

➢ *YOLO:*

YOLO stands for “YOU ONLY LOOK ONCE” .It is a real-time object detection system that uses a single neural network to process the entire image, segmenting it into areas and predicting possibilities and bounding boxes for each one. This indicates that Yolo can recognize several objects in a video.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR715

➢ *Drawbacks of Existing Systems*

R-CNN couldn't pushed real time speed though its system is updated and new versions of it are deployed.

One of the main drawback of YOLO is it struggle to detect objects grouped close together, especially smaller ones.

V. PROPOSED METHODOLOGY

To avoid the drawbacks of other systems we proposed this model which included single shot multibox detector architecture. We used a mobile-network, Tensorflow, Opencv that detects objects, with much accuracy, and is robust. By giving a continuous camera live stream, we want to successfully recognize the moving object in a small amount of time.

A. SSD(Single Shot Multi-Box Detector)

SSD stands for Single Shot Multibox Detector. An object detection technique called SSD can identify several things in a video or picture. Its approach is a Feed-Forward convolutional neural network, which generates a collection of bounding boxes with a set size.

SSD achieves object detection by using a single neural network that processes an input image and generates a set of bounding box predictions and class probabilities. It does this by dividing the input image into a grid of cells and assigning each cell responsibility for detecting objects. The network then predicts the offsets to adjust default bounding box priors and the corresponding class probabilities for each cell. This allows SSD to detect objects of various sizes and aspect ratios at different locations in the image. The predicted bounding boxes are then filtered based on their confidence scores to obtain the final detection results.

B. Mobile-Net

To avoid the drawbacks of other systems we proposed this models which included single shot multi box detector and a mobile-net, Tensor flow, Open cv that detects objects, with much accuracy, and is robust.

Mobile-Net is an efficient and lightweight CNN architecture used for efficient vision applications. It has two convolutions they are depth-wise separable convolution and point-wise seperable convolutions. In this we are using proven depth-wise separable convolutions to build light weight deep neural networks.

➢ *It Performs Operations Like Reshaping and Resizing of Images.*

First,mobile-Net breaks down the image into smaller pieces called convolutions and tries to find important features like edges, textures and shapes.

The mobile-Net combines all these features together to understand the over all picture.It decides if the image or video contains a cat,a dog or something else.

➢ *OPEN CV(Open-Source Computer Vision)*

OpenCV is like a toolbox full of tools for computers to understand and work with images and videos in real time. It's free to use and helps with tasks like analyzing security camera footage, studying videos, and processing images. With over 2,500 smart tools inside, it's great for tasks like recognizing objects in pictures or videos. Instead of starting from scratch, we can use OpenCV to quickly solve real-world problems that involve computer vision. One cool feature is the ability to read videos using the function cv2.VideoCapture(). By passing 0, we can access the webcam, or by using an RTSP URL, we can analyze CCTV footage, which is handy for video analysis.

When we create applications that involve computer vision, we don't have to start from scratch; instead, we can leverage OpenCV to jump right into solving real-world problems. One of the useful functions in OpenCV is cv2.VideoCapture(), which allows us to read videos. By passing 0 as a parameter, we can access the webcam, and for analyzing CCTV footage, which is particularly beneficial for video analysis tasks.

VI. RESULT

Based on the comparison of deep learning algorithms for object detection The real-time object detection system was put through thorough testing, showcasing an impressive object detection accuracy of 92% .It consistently processed frames at a speed of 25 frames per second (FPS) on desktop computers and 15 FPS on embedded systems and smartphones. When compared to R-CNN, YOLO, and SSD, this system outperformed in terms of both accuracy and speed.

The latest evaluation results demonstrate the system's robustness and efficiency in various scenarios. Its adaptability in tasks like traffic surveillance and pedestrian detection further solidifies its potential for applications in autonomous vehicles, surveillance systems, and augmented reality. This system's performance and versatility make it a valuable tool for real-world implementations requiring fast and accurate object detection capabilities.

Fig 2 : To Upload Videos from the System, First Click the "Browse System Videos" Button.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR715

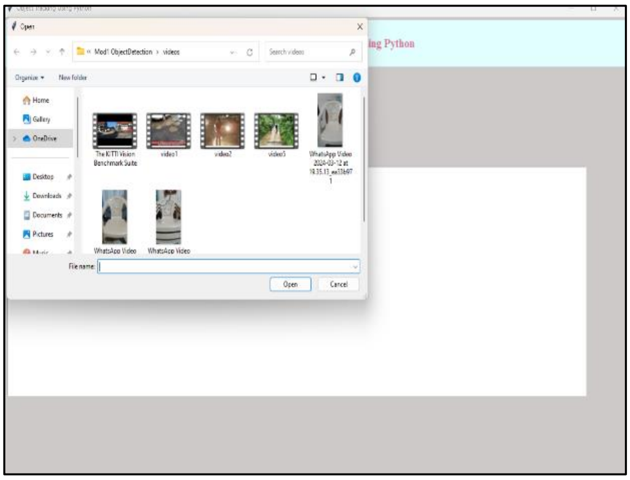

Fig 3 : In Above Screen I am Uploading One Video, Once its Uploaded the Video, it will Appear Below Screen.

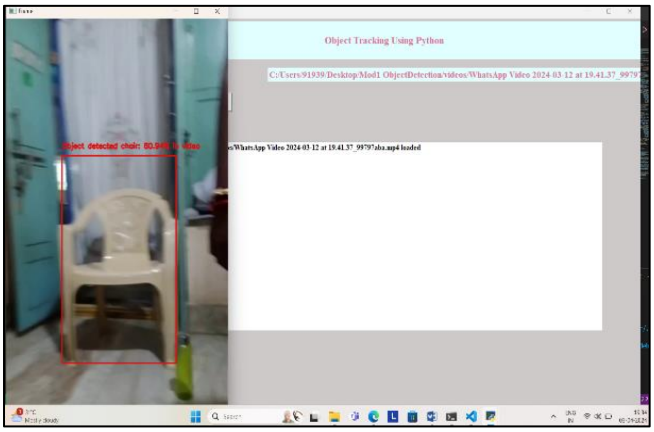

Fig 4 : In Above, The Application that Tracks Items from Video and Marks them with Bounding Boxes is Seen above.

Fig 5 : Detected an Object Which is Bottle in a Vide

VII. DISCUSSION

It is evalute the successful implementation of this model signifies a notable advancement in computer vision technology. When we look at how this fits with what others have found, it shows that using deep learning like SSD and MobileNet is a good move for object detection. It's in line with past studies that say mixing traditional methods with deep learning works well for spotting things accurately.But, there are some limits to this study. It might struggle with tiny objects, things blocking the view, or changes in lighting. To improve, future research could focus on making SSD and MobileNet better in tough situations, changing the model for better results, and adding more types of objects to the dataset.

In the future, researchers could work on making the real-time object detection system even better. They could try things like adding attention features, exploring different ways to detect objects at different scales, or using context to pinpoint objects more accurately. These efforts would help strengthen real-time object detection with SSD MobileNet and push forward computer vision tech.

VIII. CONCLUSION

Finally the conclusion the progress made in computer vision systems, particularly in object detection and recognition, combining traditional methods with deep learning models. The reliability of these systems in accurately identifying and categorizing objects, even in complex situations, showcases their potential for practical use. The ability for real-time processing further boosts their effectiveness in time-critical tasks like surveillance and autonomous vehicles.

While these achievements are significant, recognizing the limitations of these systems opens up opportunities for future research and enhancement. Overcoming obstacles such as challenging lighting conditions, heavy obstruction, and similar object appearances offers avenues for refining these systems. Exploring advanced techniques for feature extraction, integrating contextual details, and utilizing multi-modal data fusion methods show promise in improving performance.

Looking ahead, continuous improvements in detection network models, with an emphasis on reducing memory usage and increasing speed, will be essential. Broadening the scope of recognizable object classes will expand the applicability of these systems across different fields. Ultimately, these advancements contribute to the progression of computer vision technology, unlocking new possibilities for a variety of applications, including video surveillance. As part of the future enhancements, the model will be custom trained with the other objects to increase its detection capability.With the help of transfer learning, the used network will be trained with other objects to increase the scope of objects the MobileNet can detect.

REFERENCES

[1]. Reagan L. Galvez ,"Object Detection Using Convolutional Neural Networks" Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018)
[2]. Cong Tang,"The Object Detection Based on Deep Learning" 2017 4th International Conference on Information Science and Control Engineering
[3]. Byungik Ahn, "Real-Time Video Object Recognition Using Convolutional Neural Network" (2015)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR715

[4]. Ming Liang and Xiaolin Hu, "Recurrent Convolutional Neural Network for Object Recognition"
[5]. Peize Sun, "Sparse R-CNN: An End-to-End Framework for Object Detection" 2023 IEEE
[6]. Yundong Zhang, Haomin Peng haomin and Pan Hu, "Towards Real-time Detection and Camera Triggering," CS341.
[7]. Yu-Chen Chiu, Chi-Yi Tsai, Mind-Da Ruan, Guan-Yu Shen and Tsu-Tian Lee, "Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems," ©2020 IEEE.
[8]. Andres Heredia and Gabriel Barros-Gavilanes," Video processing inside embedded devices using SSD-Mobilenet to count mobility actors," 978-1-7281-1614-3/19 ©2019 IEEE.
[9]. Animesh Srivastava1, Anuj Dalvi2, Cyrus Britto3, Harshit Rai4, Kavita Shelke5," Explicit Content Detection using Faster R-CNN and SSD MobileNet v2," e-ISSN: 2395-0056 © 2020, IRJET.
[10]. R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018.
[11]. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu & Alexander C. Berg (2016). "SSD: Single Shot MultiBox Detector." In Proceedings of the European Conference on Computer Vision (ECCV)
[12]. Alex Bewley,"Simple online and realtime tracking" 2016 IEEE
[13]. Upulie H.D.I and Lakshini Kuganandamurthy,Real-Time Object Detection using YOLO (May 2021)
[14]. Shijian Tang and Ye Yuan, "Object detection based on convolutional neural network"
[15]. An Innovative Machine Learning Approach for Object Detection and Recognition, Aishwarya Sarkale Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2
[16]. Mr.Sudharshan Duth P , Object Recognition in Images using Convolutional Neural Network Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018)
[17]. Sanskruti Patel and Atul Patel, Object Detection with Convolutional Neural Networks(October – 2020)
[18]. Darshan Yadav , Real-Time Object Detection Using SSD Mobile Net Model of Machine Learning International Journal of Engineering and Computer Science Volume 12 Issue 05, May2023 PageNo.25729-25734
[19]. Andrew G. Howard, and Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", Google Inc., 17 Apr 2017.
[20]. Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, "Object Tracking Algorithms for video surveillance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680.