WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

REAL TIME OBJECT DETECTION BY USING DEEP LEARNING

Manish Yewange[1], Ramkishan Kamble[2], Kanishka Gaikwad[3], Sagar Maske[4], Prof.D.T.Salunke[5]

JSPM Rajarshi Shahu College Of Engineering Pune-411033

ABSTRACT

Object detection has received a significant attention from researchers in recent years because of its close relationship with video analysis and image understanding. Handcrafted features and shallow trainable architectures are the foundations of traditional object detection methods. By constructing complex ensembles that combine multiple low-level image features with high-level context from object detectors and scene classifiers, their performance can easily plateau. With the rapid advancement of deep learning, more powerful tools that can learn semantic, high-level, and deeper features are being introduced to address the issues that plague traditional architectures. In terms of network architecture, training strategy, and optimization function, these models behave differently. We present a review of deep learning-based object detection frameworks in this paper.Object recognition tasks, on the other hand, are more difficult to assess, consume more energy, and necessitate more computing power than image categorization. To address these issues, a novel technique for real-time object detection applications is being developed in order to improve the detection process' accuracy and energy efficiency.

Keywords *- CNN algorithm, framework, data base images or videos*

1. INTRODUCTION

Human computer interaction is an important application of object detection using CNN which is a challenging and an interesting problem. It can be used to build more smatter and accurate robots with an ability of better understanding the objects. There are many other real life applications of this paper such as surveillance cameras used at highways to prevent over speeding, interactive game development and driverless cars. Object detection which includes location detection and detecting the categories of various objects present in one single image, nearly two thousand regions are proposed to contain an object in the image which are called the proposed regions. Several machine learning and feature extraction algorithms have been developed for object detection tasks. Numerous handcrafted feature extraction techniques for object detection such as SVM(Support Vector Machines), SGD(Stochastic Gradient Descent), Convolutional Neural Support Vector Machines (CNSVMs) or an integration of multiple features have been proposed. Due to the success of Convolutional Neural Network (CNN) in image classification tasks[3] it has also been used in object detection tasks. Contrasting to conventional Computer Vision(CV) systems and other machine learning tasks where each feature must be defined beforehand manually, here in CNN, in automatically learns to extract features from the predefined database of features. The CNN is combined with neural network classifiers which are feed forward to make the CNN network trainable on the dataset all over. CNN requires a very large number of trained data to speculate well enough. The capability of CNN to work on larger datasets even with a limited computational power increases the value of CNN and making it.

More likely to be used. However, this is not the case always, sometimes we need to detect objects with a limited number of features. Although SIFT[6] and other traditional detection techniques give a less accurate result than CNN, they require only a small amount of datasets to speculate. The conventional methods have their own limitations such as their modeling capacities are bound which stay unchanged for different sources of data.

2. LITERATURE SURVEY

[1]Real-Time Objects Recognition Approach for Assisting Blind People. [Jamal S. Zraqou Wissam M. Alkhadour and Mohammad Z. Siam, Multimedia Systems Department, Electrical Engineering Department, Isra University, Amman-Jordan Accepted 30 Jan 2017, Available online 31 Jan 2017, Vol.7, No.1] Blind assistance is promoting a widely challenge in computer vision such as navigation and path finding. In this paper, two cameras placed on blind person's glasses, GPS free service, and ultra-sonic sensor are employed to provide. the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. The two cameras are necessary to generate the depth by creating the disparity map of the scene, GPS service is used to create groups of objects based on their locations, and the sensor is used to detect any obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. The proposed method for the blind aims at expanding possibilities to people with vision loss to achieve their full potential. The experimental results reveal the performance of the proposed work in about real time system.

[2]Object Detection Combining Recognition and Segmentation [Fudan University, Shanghai, PRC, yfshen@fudan.edu.cn University of Pennsylvania,3330 Walnut Street, Philadelphia, PA19104 Liming Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen.] We develop an object detection method combining top-down recognition with bottom-up image segmentation. There are two main steps in this method: a hypothesis generation step and a verification step. In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust to object deformation and background clutter. The improved Shape Context is used to generate a set of hypotheses of object locations and figure ground masks, which have high recall and low precision rate. In the verification step, we first compute a set of feasible segmentations that are consistent with top-down object hypotheses, then we propose a False Positive Pruning(FPP) procedure to prune out false positives. We exploit the fact that false positive

regions typically do not align with any feasible image segmentation. Experiments show that this simple framework is capable of achieving both high recall and high precision with only a few positive training examples and that this method can be generalized to many object4 - Microsoft COCO Common Objects in Context.

Seyed Yahya Nikouei et.al [3] Human objects detection, behavior recognition and prediction in smart surveillance fall into that category, where a transition of a huge volume of video streaming data can take valuable time and place heavy pressure on communication networks. It is widely recognized that video processing and object detection are computing intensive and too expensive to be handled by resource-limited edge devices. Inspired by the depthwise separable convolution and Single Shot Multi-Box Detector (SSD), a lightweight Convolutional Neural Network (L-CNN) is introduced in this paper. By narrowing down the classifier's searching space to focus on human objects in surveillance video frames, the proposed L-CNN algorithm is able to detect pedestrians with an affordable computation workload to an edge device. A prototype has been implemented on an edge node (Raspberry PI 3) using open CV libraries, and satisfactory performance is achieved using real-world surveillance video streams.

Mohannad Farag et.al [4] in this work The movement of SCARA robot is guided by deep learning-based object detection for grasp task and edge detection-based position measurement for place task. Deep Convolutional Neural Network (CNN) model, called KS S net, is developed for object detection based on CNN Alexnet using transfer learning approach. SCARA training dataset with 4000 images of two object categories associated with 20 different positions is created and labeled to train KSS net model. The position of the detected object is included in prediction result at the output classification layer. This method achieved the state-of-the-art results at 100% precision of object detection, 100% accuracy for robotic positioning and 100% successful real-time robotic grasping within 0.38 seconds as detection time.

Edward Rzaev et .al [5] In this review Object detection is one of the most active research and application areas of neural networks. In this article we combine FPGA and neural networks technologies to solve the real-time object recognition problem. The article discusses the integration of the YOLOv3 neural network on the DE10-Nano FPGA. Slightly worse indicators of the main metrics (mAP, FPS, inference time) when operating a neural network on a De10-Nano board in comparison with more expensive solutions based on GPUs, are offset by differences in the cost and dimensions of the FPGA board used. Based on the results of the study of various methods for converting neural networks to FPGA.

Shaoqing Ren et. al [6] In this work, we introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectless scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection. We further merge RPN and Fast R-CNN into a single network by sharing their convolutional features using the recently popular terminology of neural networks with 'attention' mechanisms, the RPN component tells the unified network where to look. For the very deep VGG-16 model.

Di Guo et,al [7] Grasp an object from a stack of objects in realtime is still a challenge in robotics. This requires the robot to have the ability of both fast object discovery and grasp detection: a target object should be picked out from the stack first and then a proper grasp configuration is applied to grasp the object. In this paper, we propose a shared convolutional neural network (CNN) which can simultaneously implement these two tasks in real-time. The processing speed of the model is about 100 frames per second on a GPU which largely satisfies the requirement. Meanwhile, we also establish a labeled RGBD dataset which contains scenes of stacked objects for robotic grasping.

Chandan G, et. al [8] in this review Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since past few years. 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. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance.

Zhihao Chen et. al [9] In this review , we will introduce our object detection, localization and tracking system for smart mobility applications like traffic road and railway environment. Firstly, an object detection and tracking approach was firstly carried out within two deep learning approaches: You Only Look Once (YOLO) V3 and Single Shot Detector (SSD). A comparison between the two methods allows us to identify their applicability in the traffic environment. Both the performances in road and in railway environments were evaluated. Secondly, object distance estimation based on Mono depth algorithm was developed.

3. PROPOSED SYSTEM

The proposed method is to help detect object to person in detecting the obstacle in front of them i.e. car, person, traffic sign etc. as a camera based assistive object detection technique. The implemented idea involves person, car and traffic sign detection from image taken by camera and produced sound after detection of object.

4. CLASSIFICATION

Image classification refers to the task of extracting information classes from two or many class of image. Features extracted bywavelet transform and by using ultraviolet rays are feed to classifier so that classifier, here CNN algorithm, should be able to classify the note and detect object/obstacle.

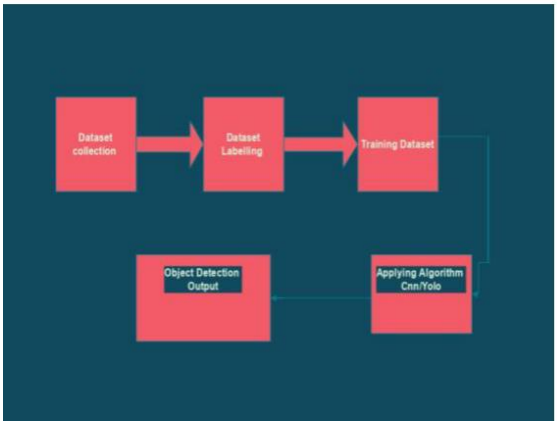

Fig 1. Proposed system block diagram

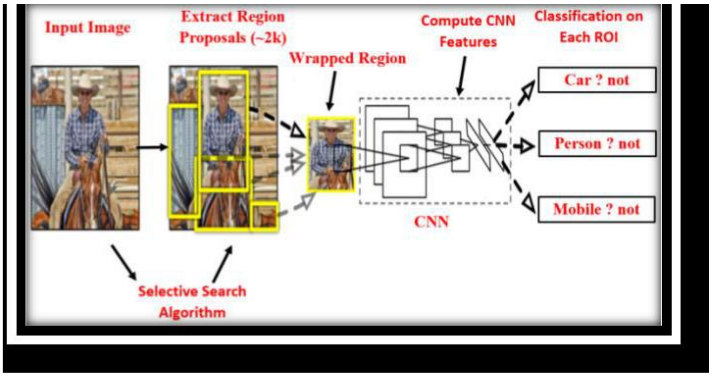

Fig 2. CNN working structure

Assuming the infirmity is gentle, the framework will propose specific drugs; assuming the condition is moderate, the framework will suggest that the client visit a specialist in the event that the manifestations don't improve; and assuming that the illness is serious, the framework will suggest that the client see a specialist right once. Furthermore, the framework makes dietary and exercise proposals in light of the illness.

A) INPUT IMAGE:

Camera captures image of note or object and is feed as input to the Python for further processing

B) PRE-PROCESSING:

Pre-processing images commonly involves removing low-frequency background noise, normalizing the intensity of the individual particles images, removing reflections, and masking portions of images. Image pre-processing is the technique of enhancing data images prior to computational processing

C) FEATURE EXTRACTION:

(a) Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved.
(b) We are using Wavelet transform to extract features like RMS value, average, entropy of image.

A.CNN Algorithm

Over the most recent decade, counterfeit neural organizations have made some amazing progress. Profound layered convolutional neural organizations (CNN) have exhibited top tier execution on an assortment of undertakings. an assortment of AI issues, including picture acknowledgment.

CNN is a counterfeit neural organization with a one of a kind geography, as found in Figure 2. Dark scale or RGB pictures (3 channels) are frequently utilized as CNN input information (1 channel). A few convolutional or pooling layers follow the information layer (with or without enactment capacities). To handle arrangement issues, at least one full association (FC) layers are normally utilized. In view of the info picture, the last layer creates forecast qualities (like back likelihood or probability) for K different kinds of articles. in. An initiation capacity can be applied to each CNN layer to control how much result esteem that is spread to the following layer. For middle layers, the redressed straight unit is utilized (ReLU)

$$f_k(\boldsymbol{z}) = \frac{\exp(z_k)}{\sum_{\kappa=1}^{K} \exp(z_\kappa)},$$

(1)

The I-th unit in the l-th transitional layer gets all iR signals as an aggregate. Meanwhile, the delicate max work is normally used to give probabilistic results to the last layer.

$$f(a_i^l) = \max(0, a_i^l),$$

(2)

It's actually significant that z is a K-layered vector, with zki addressing the amount of signs got by the last layer's k-th unit. Since the capacity is non-negative and meets the unit aggregate prerequisite (kef(z) = 1), the outcome infers a class back likelihood that an information has a place with the k-th class. Subsequently, by involving the delicate max work in the result layer, CNN can be utilized as likelihood assessors for object arrangement assignments. One of the remarkable properties of CNNs is that they have sequential various element portrayals that are naturally organized in each convolutional layer through preparing utilizing given named cases. Standard dimensionality decrease draws near (like as PCA) will see each element portrayal independently, in spite of this novel situation. A convolutional layer is a sort of layer that is utilized to consolidate at least two Apply.

I. POOLING:

The convolutional layer's element map is subsampled in a pooling layer prior to being given to the following convolutional layer in conventional CNNs. The pooling layer works by supplanting a little fix in the element map with the outline measurement for that district. The notable max-pooling layer, for instance, changes the information fix over to a solitary worth, the amount of all qualities inside that fix. Other pooling choices incorporate interpretations. In the event that a model's results don't change when the information is made an interpretation of, it is supposed to be invariant to interpretation.

In exercises that need shifted input sizes, pooling is once in a while required. In a CNN for order, for instance, the result layer is generally a completely connected layer with a set information size. When managing pictures of fluctuating goal or size, go-between pooling layers are important to proficiently decrease the picture to a decent size prior to conveying it to the result layer. This can be refined by requiring the last pooling layer to deliver a predetermined number of results by cutting the contribution to a comparing number of districts and providing a synopsis measurement for every one of these areas. Pooling can be utilized spatially, along a similar component map, or across highlight maps that have been made utilizing various portions

II. TRAINING:

A CNN can be prepared similarly as some other profound feed forward network. The parts of the model are fitted utilizing stochastic angle plunge using the back proliferation calculation for registering inclinations in the wake of setting up an errand explicit misfortune capacity like cross-entropy. Cross-approval is ordinarily used to alter hyper boundaries such the quantity of layers, part aspects, and pooling layer widths.

5. RESULT AND DISCUSSION

This project's output shows the detected objects with a rectangle box around them and a label on top showing the object's name and thus the accuracy with which it was detected. It can reliably extract any number of items present during a single picture.

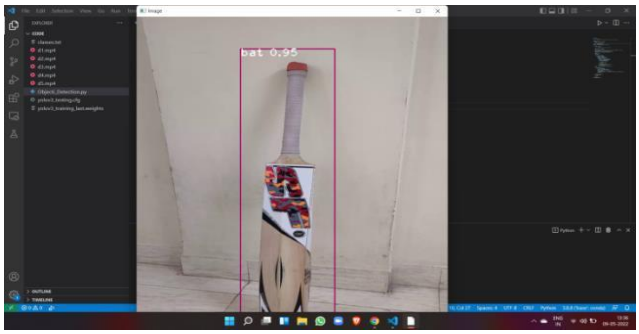

Fig 5.1 Bat detectected

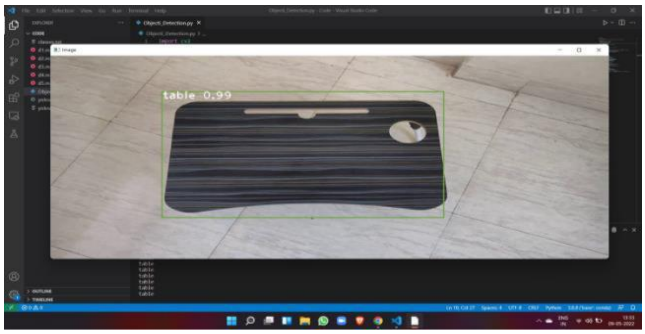

Fig 5.2 Table detected

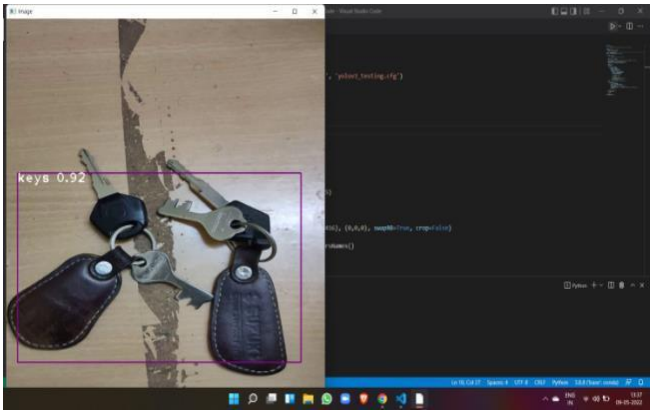

Fig 5.3 Key detected

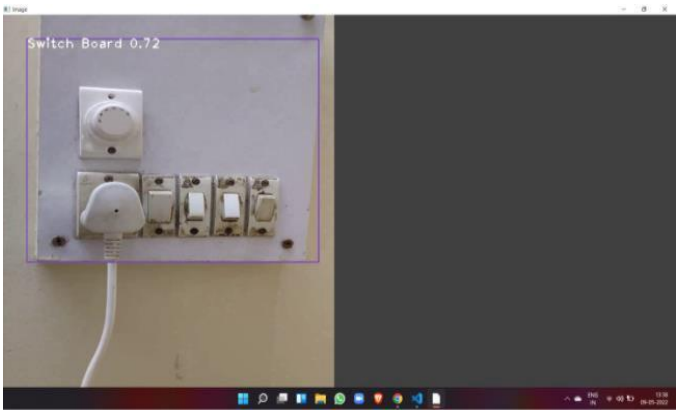

Fig 5.4 Switch board detected

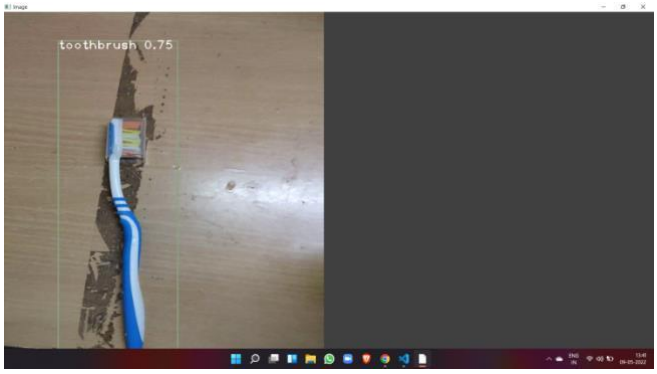

Fig 5.5 Tooth brush detected

In the above images, such as the result of our design system "Real time object detection," we have observed that we have collected various output results by giving a real-time object and there are 99.9 percent of detected objects with their specific names, such as: bat, table, key, switch board, tooth brush. We achieve 99.9 percent accuracy in detecting real-time objects by using CNN classifier. Object detection is a computer vision technology that locates items in pictures or movies.

6. CONCLUSION

Main Objective of CNN algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance. This real time analysis of the ecosystem can yield great results by enabling security, order and utility for any enterprise.

REFERENCES

[1] Real-Time Objects Recognition Approach for Assisting Blind People. Jamal S. Zraqou Wissam M. Alkhadour and Mohammad Z. Siam,2017

[2] Object Detection Combining Recognition and Segmentation[Fudan University, Shanghai, Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen] 2018

[3] Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN Seyed Yahya Nikouei†, Yu Chen†, Sejun Song‡, Ronghua Xu†, Baek-Young Choi‡, Timothy R. Faughnanξ 018

[4] Real-Time Robotic Grasping and Localization Using Deep Learning-Based Object Detection Technique Mohannad Farag,2nd Abdul Nasir Abd Ghafar Mohammed Hayyan ALSIBAI2019

[5] Neural Network for Real-Time Object Detection on FPGA Edward Rzaev Moscow, Russian Federatio Anton Khanaev Moscow, Russian Federation Aleksandr Amerikanov 2021

[6] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun2016

[7] Object Discovery and Grasp Detection with a Shared Convolutional Neural Network Di Guo, Tao Kong, Fuchun Sun and Huaping Liu 2016 IEEE

[8] Real Time Object Detection and Tracking Using Deep Learning and Open CV Real Time Object Detection and Tracking Using Deep Learning and Open CV

[9] Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility Zhihao Chen 2nd Redouane Khemmar 3rd Benoit Decoux 2019 IEEE