Detection and Recognition of Objects and Providing Purchase links using APIs

Sainag Shetty[1], Amitesh Sharma[2], Abhishek Singh[3], Tanmay Sawaji[4], Dnyaneshwar Dhangar[5]
BE Student[1, 2, 3, 4], Assistant Professor[5]
Department of Computer Science
Rajiv Gandhi Institute of Technology, Andheri, Mumbai, India

Abstract:
The proposed system is designed to the help the online buyers and the online window-shoppers. The idea is that a lot of people today look to buy goods over the internet, but they need to know the name of the item they wish to purchase and enter the name in the field provided by the online sellers. This project aims to reduce the burden from the buyers. In the proposed system, the user only needs to click the image of the object he/she wishes to look up. The system will detect the object using the detection algorithm. After the object is detected, the system performs an API call requesting a list of products from the online vendor. The API returns the result in a JSON format. The system formats the result in a manner that will be easily accessible to the user.

Keywords: Object Detection, Object Recognition, Convolutional Neural Networks, Region-based CNN, Faster R-CNN, Application Programming Interface, Flipkart API

I.   INTRODUCTION

To search for an online product today, it takes a lot of effort; taking in the user input which are the keywords and then select from the myriad of products that are presented. Before this the buyer also need to choose an online vendor like Amazon, Flipkart, and so on. It would be a lot hassle free for one if the buyer could see the product; take a picture of it and get the list of similar products from all the vendors at one place. The project aims to make the job for the online job easier. The project uses machine learning to detect the object that the user takes the picture of. After which, it uses APIs to search the different vendor sites for the detected object, and provides a list of the products. Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance. In our project, we will detect stationary and computing items using a camera feed. Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension, as is the case in data mining applications, machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets. In our system, where the object detection is used to find the price and purchase links for the detected object(s), we propose to build a system which can recognize everyday

objects with reasonable accuracy. We intend to choose a framework amongst some of the more accurate techniques available such as YOLOv2 and Region-based Convolution Neural Networks. Application Programming Interface (API) is a set of subroutine definitions, protocols, and tools for building application software. In general terms, it is a set of clearly defined methods of communication between various software components. APIs are basically used for information hiding. It can also be used by other systems to access the underlying system. Web APIs are the defined interfaces through which interactions happen between an enterprise and applications that use its assets. API are mostly linked with software as service in cloud computing and can implemented through JSon, xml or other methods. API is typically defined as a set of Hypertext Transfer Protocol (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language (XML) or JavaScript Object Notation (JSON) format. APIs are provided by the underlying vendor sites. Developer requests for data using API key, which will be provided by the vendor site. API key is a code through which API is called.

II.LITERATURE REVIEW

Fast R-CNN [3]
Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. Fast Region-based Convolutional Network method (Fast R-CNN) efficiently classifies object proposals using deep convolutional networks. Fast R-CNN uses the softmax classifier learnt during fine-tuning instead of training one-vs-rest linear SVM's post-hoc, as was done in R-CNN Rich feature hierarchies for accurate object detection and semantic segmentation: [2] one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects. When labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. The R-CNN method trains

CNNs end-to-end to classify the proposal regions into object categories or background.

You Only Look Once: Unified, Real-Time Object Detection: [4]

Base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is far less likely to predict false detections where nothing exists. It outperforms all other detection methods, including DPM and R-CNN, by a wide margin when generalizing from natural images to artwork on both the Picasso Dataset and the People-Art Dataset.

III. EXISTING SYSTEM

Current object detection systems can only detect objects and name it. There are a number of such applications. These systems can make approximations using various machine learning and computer vision techniques. For the e-commerce product finding, the buyer has to find the needed product by entering the name of the product in the input area and search it. If the user searches a particular site then he will be getting results from the same site and not a comparison from other vendors.

IV. PROPOSED SYSTEM

Proposed system for object detection has a sequence of operations like reading input image, extracting features of that image; identify the object in the image (as per the training); using the API from vendor sites(Flipkart Affiliate Program for Developers) to provide price range and direct purchase links. Several algorithms will be used for extracting features from images, training the network and testing. Large datasets will be required for every object that will be in the set. Efficiency and improving accuracy for identifying objects will be done by merging several algorithms. C, Python and OpenCV programming environment will be used for implementation.

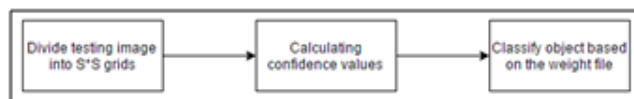## V. WORKFLOW OF THE SYSTEM

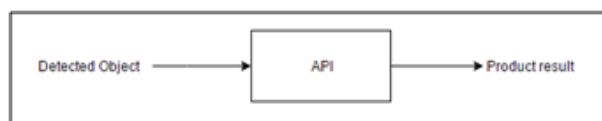**Figure.1. Phase 1**

Figure. 2. Phase 2

Figure. 3. Phase 3

In the above diagram it shows the workflow of the proposed system in phases.

Phase 1: An image is taken from the data set. The dataset used in this project is a modified version of the COCO dataset. The image is sent to the first convolution neural network layer where it is divided into S*S grids.

Each grid is processed to calculate 5 values:
- h, Predicted height of the entire object
- w, Predicted width of the entire object
- x, Predicted x coordinate of the centre of the object
- y, Predicted y coordinate of the centre of the object
- c, Predicted confidence of the object; confidence shows the probability of the object belonging to the stated class.

The probabilities are multiplied together to calculate which object exists in the image. Since the class of the object is mentioned in the training set, these values are stored in the weight file. This weight file is used for testing and classification of object.

Phase 2:
The testing image is taken by the webcam of the machine and the algorithm is applied. The weights of the bound are compared by the weights stored in the weight file.

Phase 3:
Features are selected and object is identified. After identification, an API[7] call is performed which send the name of the detected object along with the required keys. The API results are obtained in JSON format which is appropriately formatted and viewed to the user.

VI. IMPLEMENTED SYSTEM

The system uses a web interface to capture images through the camera and then detects the object by running Darknet framework and the YOLO algorithm. The detected object's class is stored in a file and then a query regarding the object is sent to Flipkart. Flipkart returns response in JSON format which is formatted so that it displays the results in a list. If multiple objects are detected it opens the other objects in a new tab.
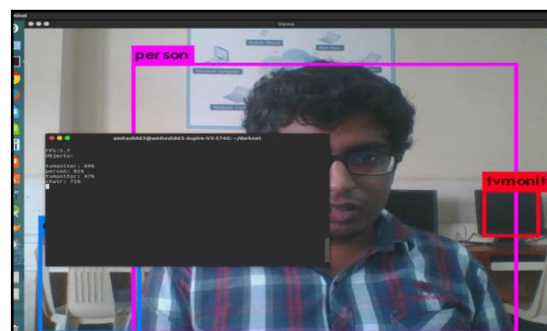

Figure. 4. Object Detection

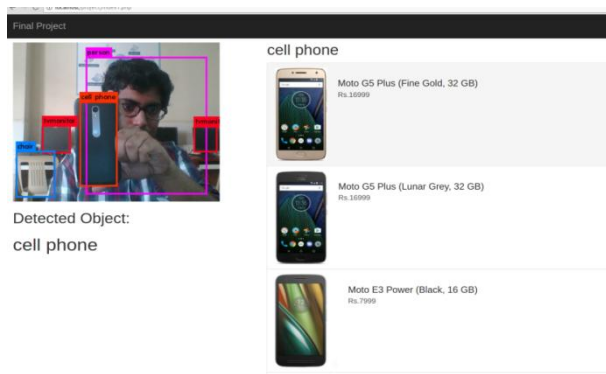**Figure. 5. Displaying the result from the api call**

Figure. 6. Interface with the detected object image and the results

VII. DRAWBACKS

- There is a noticeable delay between clicking the image of the object to providing the list of products.
- Changes in lighting conditions affect the accuracy and the ability to detect the objects well.
- Huge database is needed to be maintained to detect the object accurately.
- There is a limit to the number of products the API can return for every call

VIII. CONCLUSION

The proposed system is mainly designed in order to help the online buyers find the products more easily and efficiently. Although there are some drawbacks to the system, it still performs the required action. Several algorithms were tried and tested for the detection. We are using the Darknet framework and the YOLO algorithm. The API that we are using is the one made available by Flipkart Affiliate program [7].

IX. REFERENCES

[1]. Real Time Object Detection and Recognition and Estimation of the Price of the Object -International Journal of Research in Science &Engineering Special Issue 7-ICEMTE March 2017e-ISSN: 2394-8299p-ISSN: 2394-8280

[2]. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik and UC Berkeley. Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v3), 2014

[3]. Ross Girshick Microsoft Research. Fast R-CNN, 2015

[4]. Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection, 2016

[5]. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016

[6]. Joseph Redmon and Ali Farhadi. YOLO9000: Better, Faster, Stronge, 2016

[7]. https://affiliate.flipkart.com/api-docs/affiliate_index.html