e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

OBJECT DETECTION IN REAL TIME

Sughosha KS*1, Madhu Nagaraj*2

*1PG Student, The National Institute Of Engineering, Mysore, Karnataka, India.

*2Assistant Professor, The National Institute Of Engineering, Mysore, Karnataka, India.

ABSTRACT

Object Detection in real time leads us to find the objects and balance with speed and accuracy. We will be able to run it in real time detecting multiple common objects. And the best part in this project we don't require any third party libraries to detect the objects to run other than OpenCV. So this means that we will able to run in lines of code. And we will face the duplicates like it will overlap the objects. So we are going to resolve that overlapping and give the single object output for that particular thing. So that's give us more efficient detection of those objects in real time and more smoother output for the one who sees the result of the project. It will also display the name of that object by creating bounding boxes around that object.

Keywords: Object Detection, Yolo.

I. INTRODUCTION

This project's objective was to research the relatively recent technology known as OpenCV for real-time. One of the object detection libraries for computer vision is this one. We created an application that leverages methods from the aforementioned computer vision libraries in order to do this experiment. This programmer takes camera video from every classroom to perform object detection, which allows us to determine whether or not things are detectable.

This application's entire concept was developed on the campus of our college. Many of the students began studying deep learning in order to improve their understanding of fundamental ideas. We therefore considered developing an application to examine whether or not things are detected in that specific area in real time. Consequently, both students and teachers will gain from that.

The primary goal of this project is:

- To be aware of the items in their immediate environment.
- To pinpoint the targets.
- Using video data to pinpoint the targets.
- All the physical tasks will likewise be performed excessively.
- It will help us comprehend and evaluate video scene details.

II. LITERATURE REVIEW

Right now, as a result of security concerns, the number of surveillance cameras is steadily rising. However, because it requires powerful computational power, building an intelligent detection system is challenging. With limited resources, the goal of this project is to build a practical video surveillance system that can quickly identify people who are moving. To do this, we combine background subtraction with convolutional neural networks to create a straightforward. framework for detecting and identifying moving objects in outdoor CCTV camera footages (CNNs). In order to identify the regions of interest, each video frame is first subjected to a background subtraction method (ROIs). To classify the collected ROIs into one of the predetermined classes, a CNN classification is next used. When compared to previous object detection techniques, our method significantly reduces the computational complexity. We introduce YOLO, a novel method for object detection. Classifiers have been used in the past to detect objects. Instead, we conceptualise object detection as a regression issue to spatially separated bounding boxes and associated class probabilities. Bounding boxes and class probabilities are directly predicted by a single neural network from entire images in a single assessment. One network makes up the whole detection pipeline, allowing for end-to-end optimization of detection performance. Our integrated architecture is incredibly quick. Our basic YOLO model does real-time picture processing at a frame rate of 45 frames per second. Fast YOLO, a condensed form of the network, processes around 155 frames per second while still achieving double the mAP of other real-time detectors. Modern detection techniques produce fewer incorrect predictions than YOLO, but it makes more localization errors

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

than those systems. Finally, YOLO picks up very broad representations of objects. On the Picasso Dataset and the PeopleArt Dataset, it performs far better than all other detection techniques, including DPM and RCNN. By putting the issue of object recognition within the larger framework of scene understanding, we want to advance the state-of-the- art in object recognition. This is accomplished by assembling photographs of intricate everyday scenarios with typical objects in their natural settings. Per-instance segmentations are used to label objects in order to provide accurate item localization. Our dataset includes images of 91 different object kinds that a 4-year-old could recognise with ease.

III. PROPOSED SYSTEM

The combination of three algorithms is the foundation of our method for object identification, classification, and localization. The scene in the field of view of the IP web camera is shown as an IP, a depth map, and a point cloud. The three algorithms for item detection and classification are then applied to the image and the depth map concurrently. These three techniques are the convolutional neural network (CNN), "You Only Look Once," and "Background Subtraction Algorithm" (further also abbreviated BS).

IV. IMPLEMENTATION INTRODUCTION

Python is a generally accepted, dynamic, measured, utilitarian, and item-focused programming language. Python is preferred by 90% of people above other advances because it is straightforward and enduring. It combines proficient programming with a practical strategy for enhancing programming in a variety of sectors. Python defines programming as the construction of several open-source tools that enable the building of programmes with outstanding levels of efficacy and security. Python employs practical and object-centered design principles, which results in the development of clear and useful code in all of the Python-based applications, making them simpler to maintain. The programming language Python is being used to carry out this project. Python has dynamic typing and garbage collection. Programming paradigms such as procedural, object-oriented, and functional are all supported. Python's extensive standard library has earned it the moniker "batteries included" language. Machine learning methods are used in this research.

Python can be interfaced immediately with C/ObjC/Java/Fortran. Some of Python's key features include its traditional explanation of procedural code, good care limits, accurate, understandable language structure, instinctive thing bearing, powerful data types, modules and increases helpfully written in C, C++, broad standard libraries and complete estimated quality, exceptional case-based error handling, and embeddable inside programming as a prearranging interface. Python is also compatible with the Internet Communications Engine (ICE) and a number of other innovative compromises.

The OpenCV (Open Source Computer Vision Library) programming function library was created by Intel with real-time computer vision as its primary focus. It works on various platforms. Real- time image processing is its core area of focus. OpenCV is adaptable to some particular systems, such as digital signal processors, thanks to the C interface that was originally designed for the library. To promote adoption by a larger audience, wrappers have been created for languages including C#, Python, Ruby, and Java (using JavaCV). OpenCV now offers a new C++ interface in addition to its classic C interface as of version 2.0. With the use of automatic data allocation and deallocation, this new interface aims to cut down on the amount of lines of code required to implement vision capabilities as well as on common programming issues like memory leaks that can happen when using OpenCV in C. In OpenCV, the C++ interface is now where the majority of new innovations and algorithms are produced. Unfortunately, providing wrappers in other languages for C++ code is significantly more challenging than it is for C code, therefore they typically lack some of the more recent OpenCV 2.0 features.

For Python scripting, PyCharm is the most often used IDE. The features of PyCharm are explained here, along with an introduction to it. Code completion and inspection, advanced debugging, and support for web programming languages and frameworks like Django and Flask are just a few of the best things that PyCharm has to offer its users and developers.

As an IDE for Python, JetBrains created the hybrid platform known as PyCharm. For the creation of Python applications, it is frequently employed. PyCharm is a popular Python IDE that is used by some unicorn companies like Twitter, Facebook, Amazon, and Pinterest.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

Versions 2.x and 3.x are supported.

On Windows, Linux, or Mac OS, PyCharm can be used. It also includes modules and packages that facilitate the quick and easy development of Python-based software by programmers. Additionally, it can be modified in accordance with the demands of developers.

Data flow diagrams should be drawn in multiple stacked layers. On a high level diagram, a single process node may be extended to display a more in-depth data flow diagram. After you have drawn the context diagram, add several layers of data flow diagrams. In graphical or visual form, data flow diagrams show the logical progression of information through a system.

The communication between analysts and users is facilitated by the data flow diagram's limited symbol set of four. Data flow diagrams (DFDs) depict how processes in a system utilise and supply data.

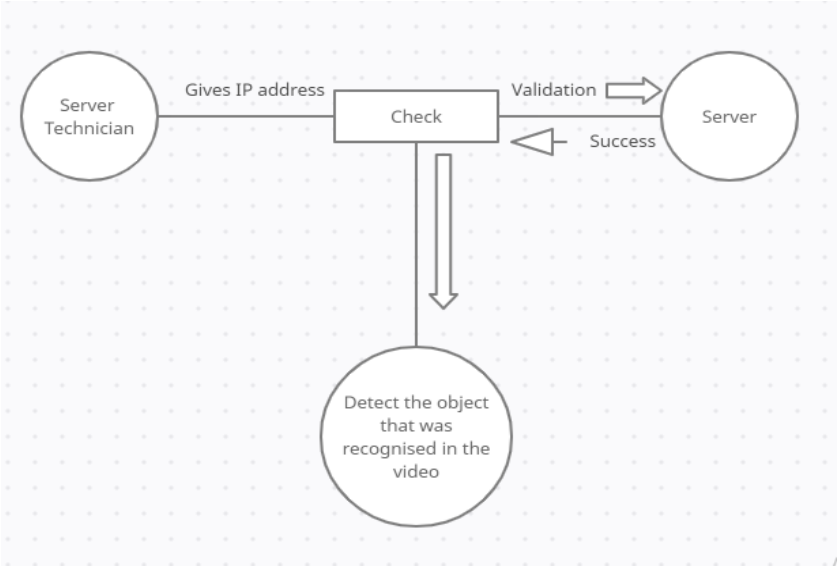


Fig 1: Data flow Diagram

The Project is implemented in different phases as follows

- The first stage of the project entails installing the software needed to install the deep learning algorithms that will be used to recognise objects.
- The project's yolo must be loaded during the second phase.
- The third phase involves attaching a mobile camera to OpenCV and utilising IPwebcam to activate the camera and begin recording a video for detecting purposes.
- Detecting the objects and placing them in a box shape are included in the fourth phase.

1. First Phase:

Installing the software needed to install the deep learning algorithms that will be used to recognise objects

2. Second phase:

Loading the yolo to the project.

It consists of yolo weight file and yolo configuration file. .Yolo weight file consists of the core of the algorithm and yolo configuration file consists all the setting of the algorithm that admin can do in the project.

YOLO – Algorithm Used

YOLO examine an image to determine the presence and location of any objects. YOLO is refreshingly straightforward: Multiple bounding boxes and class probabilities for those boxes can be predicted simultaneously by a single convolutional network. YOLO immediately optimises the performance of detection while training on full photos. Compared to conventional object detection techniques, this unified model has a number of advantages. First of all, YOLO moves quickly. We can avoid using a complicated pipeline because we frame detection as a regression problem. When predicting detections, we merely run our neural network on a fresh image. Our base network performs at 45 frames per second on a Titan X GPU with minimal batch processing, while a fast version works at more than 150 frames per second.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com

YOLO also achieves mean average precision that is more than twice as high as that of other real-time systems. Visit our (anonymous) YouTube channel at https://goo.gl/bEs6Cj to witness a demonstration of our technology in action on a camera. Second, while making predictions about an image, YOLO thinks broadly about it. Because YOLO sees the full image during training and testing, unlike sliding window and region proposal-based systems, it can encode contextual information about classes in addition to just their appearance. A popular detection technique called Fast R-CNN misinterprets background patches in an image for objects since it can't see the bigger context. When compared to Fast R-CNN, YOLO makes less background mistakes than half as many. Third, YOLO discovers universal representations of items.

YOLO performs noticeably better than top detection techniques like DPM and R-CNN when trained on real photos and tested on artwork. YOLO is highly generalizable, which reduces the likelihood that it would fail when used in new domains or with unexpected input. Online at [removed for review], you can access all of our teaching and testing code, which is open source. You can download a number of relevant models as well.

3. Third phase:

Involves attaching a mobile camera to OpenCV and utilising IPwebcam to activate the camera and begin recording a video for detecting purposes.

4. Fourth phase:

Detecting the objects and placing them in a box shape are included in the fourth phase.

Object Detection

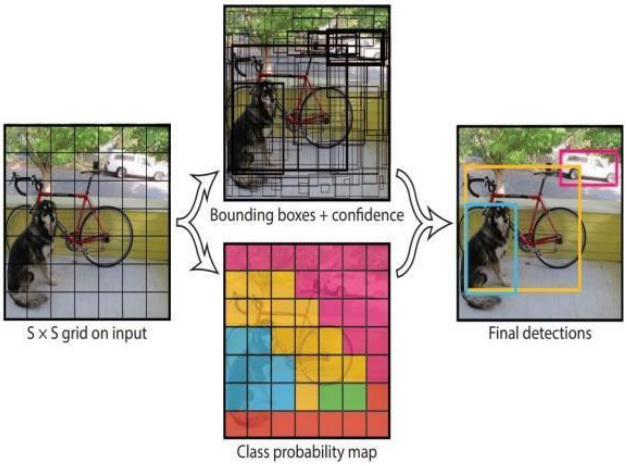
Our system can construct an exhaustive overview of the world in its range of view thanks to the employment of many complementary object detection algorithms. For the purpose of identifying moving objects in photos, background subtraction methods are thought of. An technique for background subtraction based on Gaussian Mixture Models is used, and it is implemented in OpenCV4. For noise reduction, morphological closing, and contour extraction, Gaussian blur is applied post-processing to the algorithm's findings. The detections are created, and then bounding boxes are made around them. Applying a Non- Maximum-Suppression (NMS) technique results in receiving just one bounding box for each object. However, we do not use the Intersection over Union (IoU) as a criterion to determine if a box is suppressed. If two bounding boxes intersect, our method looks for it. If they do, the area of the larger box divided by the area of the smaller box is calculated. The smaller bounding box is suppressed if this ratio is above a certain level. Since our fusion process is based on similar ideas, Section 5 goes into greater depth about why we modified the NMS. In colour photos, the CNN yolo recognises and categorises things. This ne network provides details regarding the class of the observed object as opposed to the other two approaches. But it can only recognise object classes for which it has already received training. Since the Jetson TX2's computation time per image is substantially less than that of the full version, the little version of YOLOv2 (Tiny YOLO) 5 is employed in our system. On the basis of the COCO6 dataset, we employ pertained weights.

Unified Detection

Our algorithm creates a S S grid from the input image. An object will be detected by a grid cell if its centre lies within that grid cell. Bounding boxes are seen at every grid cell, together with confidence ratings for those boxes.

The model's level of confidence that the box contains an object is shown by these confidence scores, as well as how accurately the model believes the box's prediction was made. The official definition of confidence is Pr(Object) IOUtruth prod. If the cell is empty of objects, the confidence ratings should be zero. We want the confidence score to be equal to the intersection over union (IOU) between the projected box and the actual data in the absence of it. The five predictions that each bounding box is composed of are X, Y, W, H, and confidence. In reference to the bounds of the grid cell, the coordinates (x, y) represent the location of the box's centre. The expected width and height of the image are displayed in relation to its whole. The final representation of the confidence prediction is the IOU between the predicted box and any ground truth box. These probabilities depend on whether an object is present in the grid cell in question. No matter how many boxes B there are, we can only forecast one set of class probabilities per grid cell.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com



The conditional class probabilities and the individual box confidence forecasts are multiplied at test time, Pr(Classi|Object) Pr(Object) IOUtruth pred = Pr(Classi) IOUtruth pred (1), giving us class-specific confidence ratings for each box. These ratings represent the likelihood of that class being in that box and the degree to which predicted box fits object.

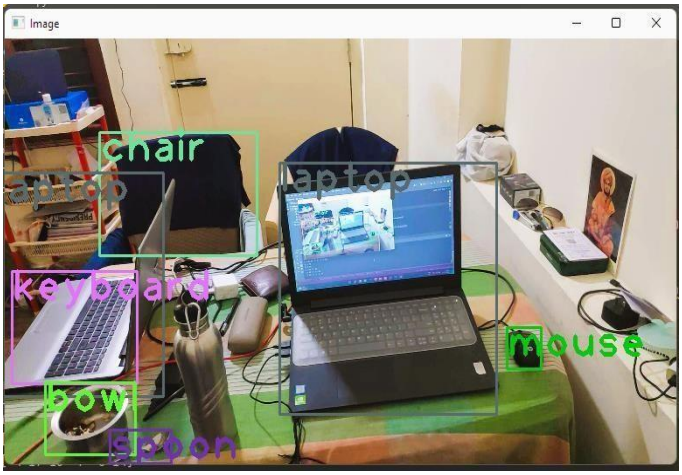
We can detect items using the deep learning algorithm Yolo. Three files are required to execute the algorithm:

- Weight file: This is the trained model, the primary component of the method used to identify objects.
- CFG file: The algorithm's configuration file contains all of the settings.
- Name file: Contains the names of the things the algorithm is able to identify.

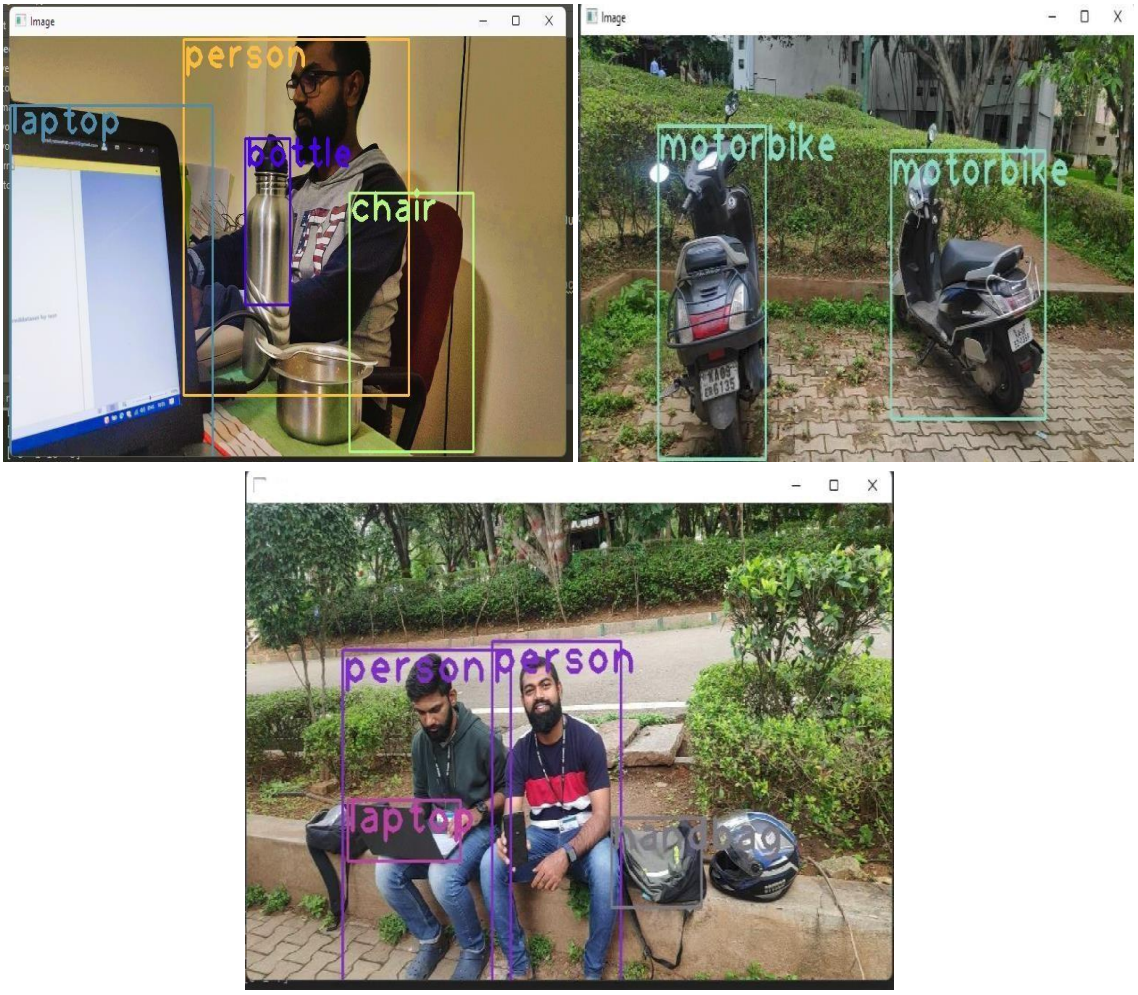
We read the weight file and configuration files using OpenCV DNN.. Fig establishes significant spatial constraints on bounding box predictions by allowing just two boxes and one class to be predicted in every grid cell. Our model's ability to anticipate the number of nearby items is constrained by the available space. Small items that occur in groups, like flocks, present a challenge for our approach. Since our model learns to anticipate bounding boxes from data, it finds it difficult to generalise to objects with novel or odd aspect ratios or configurations. Our architecture has multiple down sampling layers from the input image, hence our model also uses somewhat coarse attributes to forecast bounding boxes. Finally, even though we train on a loss function that generally corresponds to detection performance, our loss function treats errors in both small and big bounding boxes equally. However, a tiny mistake from that small kind box has a considerably bigger impact on IOU than a small mistake from a large box. Inaccurate localizations are the primary cause for the errors.

The phases are seen to cascade into one another, with the progression appearing to flow steadily downward (like a waterfall) through each phase. The succeeding phase doesn't start until the prior one's predetermined set of objectives have been completed and it has been accepted, hence the term "Waterfall Model," which refers to this process. In this paradigm, phases don't overlap..

V. RESULTS



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:07/July-2022 Impact Factor- 6.752 www.irjmets.com



VI. CONCLUSION

The application offers reliability, time savings and easy control. It can be used as a base for viewing whether the objects are detected or not for colleges or any workplace. Teachers and college staffs will also view and curriculum details of objects using this application. Also students can view details, anywhere and anytime as it is real time object detection. If we implement this technology in our colleges it will be helpful for all the stakeholders in the college.

VII. FUTURE ENHANCEMENTS

This object detection technology will motivate the students to learn this kind of technology for the betterment of the future. In future we want to elaborate this technology in every college and if in future we can elaborate this technology in every college and can also make separate login and registrations for the teachers, students and staffs of the college to get to know about this technology.

VIII. REFERENCES

[1] A hybrid framework combining background subtraction and deep neural networks for rapid person detection Chulyeon Kim1 , Jiyoung Lee2, Taekjin Han1 and Young Min Kim1*

[2] Object Detection, Classification and Localization by Infrastructural Stereo Cameras Christian Hofmann, Florian Particke, Markus Hiller and Jorn Thielecke ¨ Department of Electrical, Electronic and Communication Engineering, Information Technology, Friedrich- Alexander-Universitat Erlangen-Nurnberg, Am Wolfsmantel 33, Erlangen, Germany.

[3] You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon University of Washington pjreddie@cs.washington.edu Santosh Divvala Allen Institute for Artificial Intelligence santoshd@allenai.org Ross Girshick Facebook AI Research rbg@fb.com Ali Farhadi University of Washington ali@cs.washington.edu