Section A-Research paper
ISSN 2063-5346

ANALYSIS ON DEEP LEARNING FOR REAL-TIME OBJECT DETECTION SYSTEM BASED ON SINGLE-SHOOT DETECTOR AND OPENCV

A. Nagarjuna Reddy
Associate Professor, Department of IT
Sridevi Women's Engineering College, Telangana
Anr304@gmail.com

Gurram Pavani
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
pavanigurram001@gmail.com

Balne Manaswini
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
manaswinibalne3@gmail.com

Pabbathi Hasini
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
pabbathihasini7@gmail.com

ABSTRACT

Computer vision systems have applications in surveillance, self-driving automobiles, and robots, and one of the most important components of these systems is the ability to recognize and track objects in real time. The computer vision tasks of object identification and tracking are extremely important and have a wide range of applications in the real world. Some examples of these applications include surveillance, robots, and autonomous driving. In this paper, we provide a proposal for a system that can detect and track objects in real time by utilizing deep learning and OpenCV. Using frame differencing, optical flow, background separation, single-shot detection (SSD), and MobileNets, we propose a system that can recognize and track objects in real time. On many different datasets, the accuracy of the proposed system is very good, and it performs well in real time.

Keywords: Deep Learning, Image Processing, Motion Detection, Feature Extraction.

1. INTRODUCTION

Detecting and following the movement of objects has become an increasingly crucial part of computer vision. The process of identifying objects inside photos or video streams is referred to as object detection, while the process of continuously tracking an object through time is referred to as object tracking. The ability to detect and track real-time objects has a variety of potential uses, including but not limited to surveillance, robots, and autonomous vehicles.

ISSN 2063-5346

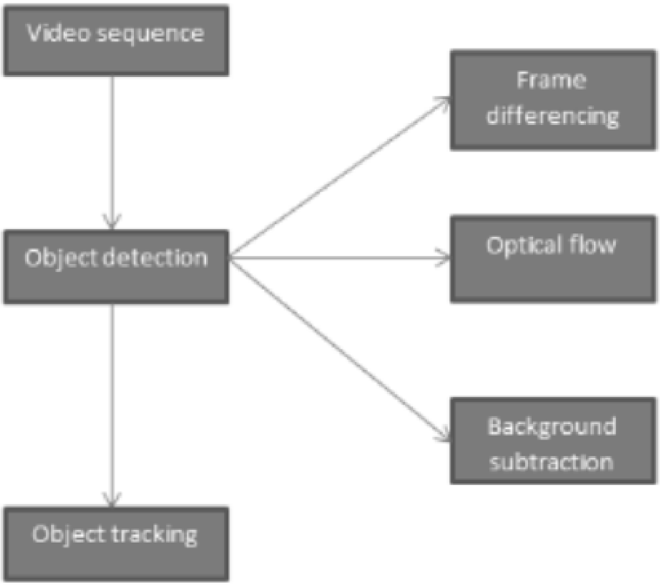

Figure 1: Real-time object detection and tracking system.

Object detection and tracking are two essential tasks in the field of computer vision that have attracted a large amount of research and development focus in recent years. Traditional methods for detecting and tracking objects rely on manually constructed features and heuristics, which can make the process more time consuming and result in less accurate results. In this paper, we provide a proposal for a system that can detect and track objects in real time by utilizing deep learning and OpenCV [1]. The suggested method makes use of frame differencing, optical flow, background separation, single-shot detection (SSD), and MobileNets to create a real-time item detection and tracking system.

2. LITERATURE REVIEW

According to Mishra and Saroha (2016), applying CV approaches to a traffic monitoring system can be beneficial. According to Biswas et al. (2019), one of the most important topics in CV is that of traffic surveillance systems that can identify and recognize moving objects. By analyzing the series of still images that were cut out of the film, it is possible to gain a deeper comprehension of the behavior of moving objects. It solves the problems that are inherent in conventional approaches that are dependent on the intervention of human operators. These kinds of systems are categorized as completely automatic, semi-automatic, or manual according to the amount of participation of the human hand in their operation. Moving object recognition is widely regarded as the most significant and essential aspect of computer vision applications (Luo et al., 2018; Runz et al., 2018). In today's world, the significance of a well-written curriculum vitae (CV) simply cannot be stressed. Applications for curriculum vitae are given a high level of consideration by many of our essential organizations, including the Security Organization.

As a result of our analysis and research, we are able to claim unequivocally that

ISSN 2063-5346

multiple frameworks are being utilized to recognize items. Due to the fact that neither our system nor the entire CV industry is at the bleeding edge of innovation, it has been utilized in the development of a number of similar frameworks. Additionally, with CV, a portion of the data is obtained from large and tiny chunks of information sets and then processed into vital data for the objectives of promoting utilization or making preparations. CNN, R-CNN, Faster R-CNN, Fast R-CNN, and YOLO are all examples of contemporary methods for detecting objects (Du, 2018; Wei and Kehtarnavaz, 2019). Other methods include YOLO. According to Asadi et al. (2018) and Chandan et al. (2018), the following are the many deep learning algorithms that are currently being employed for the detection of real-time objects. These methods can also be applied to a variety of other domains, including health-related issues, the detection of actions, and so on.

The primary contribution of this study is the design and implementation of real-time object detection and recognition systems. These systems make use of the SSD algorithm and deep learning techniques using a pre-trained model to do the detection and recognition. The system that we have suggested is able to recognize both stationary and moving objects in real time and categorize them. This study's key objectives were to examine and construct a real-time object identification system that makes use of deep learning and neural systems to detect and recognize items in real time. These objectives were the primary foci of this research. In addition, we put the free, pre-trained models that are available online through the SSD algorithm's paces on a variety of datasets to identify which models offer the best balance of accuracy and speed when it comes to locating objects. In addition to this, the system needs to be able to run on equipment that is not unreasonable. During the process of coding, we investigated a number of different deep learning structures and methods, designed and suggested a highly accurate and efficient object identification system, and then coded it.

The problems that CV is able to shine light on are the source of its significance. It is a cutting-edge technology that makes it possible for rich countries and underdeveloped countries to communicate with one another. CV enables autonomous vehicles to comprehend their surroundings and respond appropriately. Cameras are placed at various spots all around the car, and the video from those cameras is fed into a computer software. The program then forms the images in real time in order to determine street limits, examine activity signs, and differentiate. The detection and recognition of non-stationary or moving objects is an important problem that may be addressed with CV approaches, which might be beneficial for a traffic surveillance system that uses CV. By analyzing the frame sequence that was retrieved from the live video, we are able to gain additional insight into the behavior of moving objects. It solves the problems that are inherent in conventional approaches that are dependent on the intervention of human operators. These kinds of systems are categorized as

ISSN 2063-5346

completely automatic, semi-automatic, or manual according to the amount of participation of the human hand in their operation. According to Runz et al. (2018), the detection of moving objects is the component of CV applications that is the most important and critical.

3. SYSTEM MODEL

The basic objective of this system is to determine the class of an object by identifying whether the object is static or moving based on whether it is seen in an image or a video. Additionally, the functional requirements explain what the system is capable of doing. According to Yang et al. (2018), one of the primary prerequisites for the functionality of the system that we have developed is the ability to recognize both stationary and moving objects. The data processing module, the deep learning module, the static object detection module, the moving object tracking module (Shilpa, 2016), the pre-defined object module, and the object identification module make up these functional requirements. The suggested system receives a picture from the camera, compares it to the dataset, compares it to the dataset classes, executes the pre-trained models, and, ultimately, boxes the object and shows the instance of the object along with its accuracy level. The functional needs are represented by the system modules. In total, our system is comprised of six different parts.

The system modules are illustrated and explained in Figure 2, together with the different operations that are performed by each module. Before integrating these individual modules into the overall system, we will provide an in-depth analysis of each component of the system, beginning with its diagram and on to its standard operating procedure.

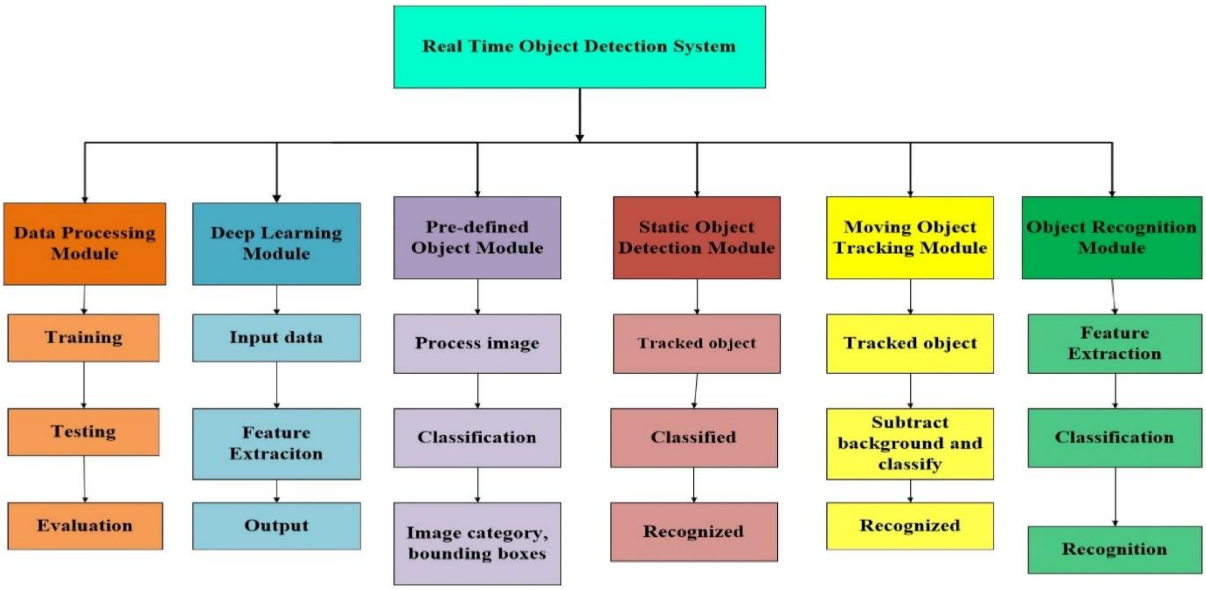

Figure 2. The functional requirements system modules.

ISSN 2063-5346

Data processing module analysis

In this section, we will talk about image datasets, which are used to train and benchmark various object detection algorithms. These processes include facial recognition, text recognition, and object recognition. In the following paragraphs, we will discuss the datasets that will be incorporated into our real-time object identification system. Deep learning strategies are able to make use of a wide variety of free datasets that are available on the internet and may be accessed by those approaches. ImageNet, PASCAL VOC, and MS Common Objects in Context (COCO) are the three datasets that are used the most frequently for object detection at the moment. DNN requires a significant amount of labeled data, also known as structured data, in order to train the model.

Kitti dataset

Kitti is a dataset that was collected using stereo cameras and lidar scanners in a variety of environments, including urban, rural, and highway driving conditions. According to Zhao et al. (2016), it is broken down into a total of ten groups, which are as follows: vans, small vehicles, trucks, sitting people, walkers, cyclists, miscellaneous, trams, and do not care. According to Salvador et al. (2016), the dimensions of the images are 1,382 by 512, and 7,500 of them yield a total of 40,000 object labels. These labels are categorized as easy, modest, or tough based on the degree to which the images are occluded and truncated.

MS common objects in context dataset

A common object in context is denoted by the abbreviation MS COCO. Microsoft is the organization that underwrites COCO, and the annotation includes categorizations, information about positions, and a textual semantic description of the image. The open source nature of the COCO dataset makes a contribution to the progression of object detection as well. The COCO dataset is a new one for picture segmentation, recognition, and captioning, and it is funded by Microsoft. This dataset's open-source has made substantial advancements in semantic segmentation in recent years (Girshick et al., 2015; Wang et al., 2016), and it has become a "standard" dataset for image semantic understanding performance, with COCO offering a distinct challenge (Chen et al., 2018). In addition, this dataset has become a benchmark for measuring the performance of image semantic understanding algorithms.

4. OBJECT DETECTION DEEP LEARNING MODULE

Deep learning is a subfield of machine learning and artificial intelligence that concentrates on the process of teaching computer models that are made up of artificial neural networks that have multiple layers. Deep neural networks (DNNs) are simply artificial neural networks (ANNs) that include numerous layers. The depth of a DNN is determined by the number of layers of hidden neurons that are present between its input and output layers. Typically, this number is greater than two. DNNs have been shown to dramatically improve the accuracy of a

ISSN 2063-5346

variety of modern applications, including speech recognition, object detection, image classification, language translation, and others. Instead of doing task-specific calculations, deep learning algorithms focus on learning representations (highlights) from information such as content, photographs, or recordings. These representations can be learned from the data. One can choose to learn independently or in a controlled environment. However, a substantial percentage of the pragmatic frameworks conveys managed learning to leverage the benefits of deep learning (Saqib et al., 2017). This is an important distinction. A crucial component of managed learning is the acquisition of knowledge through observation of the market.

Moving object tracking module

In the next paragraphs, we shall talk about moving items such as cars and people on the go. Moving object detection is a term that refers to the process of determining whether or not an object is physically moving in each area or region. By acting as a division between moving items and stationary regions or localities, the movement of moving things can be tracked and evaluated. Moving objects can be distinguished from stationary regions or places. In our framework, the camera will identify and recognize the moving protest, such as a person on the road or a car on the street. Cameras distinguish moving items surrounding the vehicle while it is stopped or gradually maneuvering; in addition, cameras distinguish moving objects around the vehicle when it is traveling gradually. After that, in many intelligent frameworks, the framework will notify the driver both visually and audibly. There are two different frameworks: one makes use of the all-encompassing See Screen and four cameras positioned on the front, back, and sides of the car, while the other makes use of a single camera mounted on the raise. Both frameworks operate together to provide the driver with a 360-degree view of their surroundings. The system that utilizes four cameras can sound an alarm for drivers when they are engaging in three distinct motions: coming to a complete stop or changing into neutral, going forward, or reversing. Whether the car is traveling forward or in backward, both the front and rear cameras are able to identify moving objects independently.

Section A-Research paper
ISSN 2063-5346

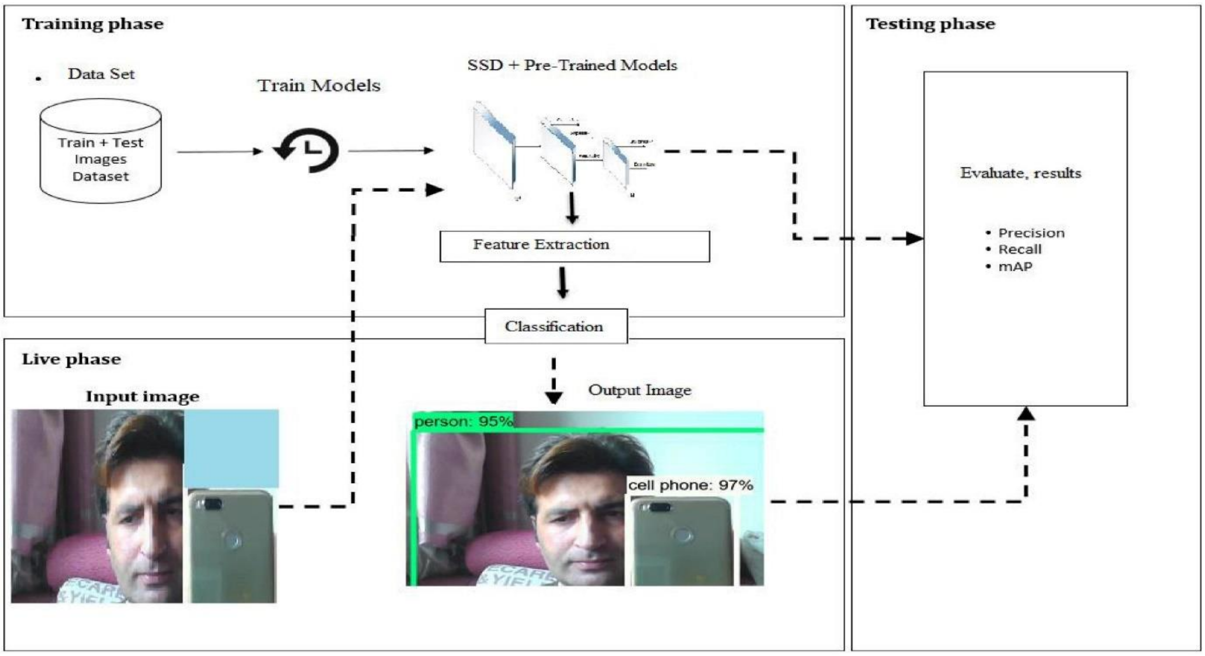

Figure 3. The flow chart of the deep learning module.

The framework generates video symbols with the assistance of the cameras, which it then uses to locate things that are moving about. The architecture that makes use of the Around See Screen has been changed to enable the analysis of video signals from the four cameras that are now mounted on the front, back, and both side-view mirrors of the vehicle. After that, it will be able to rapidly differentiate between the many moving things that are near the vehicle. It is possible for the transmission to select which of the three options—going forward, stopping, or backing up—applies at any given time depending on where the vehicle is moving. Figure 4 displays the flow chart that is associated with the deep learning module.

CONCLUSION

The fundamental objectives of this study were to analyze deep learning and its many different methodologies and structures, and then to construct a real-time object detection system that makes use of deep learning and neural systems for the purpose of object identification and recognition. In a similar manner, the system needed to be maintained on equipment that was adequate. During the process of coding, several different deep learning structures were experimented with and assessed. The primary objective of this paper is to establish which of the pre-trained models can more accurately detect and recognize an item, as well as which model performs the best on whatever dataset, by applying SSD to test them on a variety of different kinds of datasets. The results of these tests are the core contribution of this paper. As a consequence of this, we came to the realization that the pre-trained model SSD_MobileNet_v1_coco performed better than the other options on the MS COCO dataset.

REFERENCES

[1] R. S. Shankar, L. V. Srinivas, P. Neelima and G. Mahesh, "A Framework to Enhance Object Detection Performance by using YOLO Algorithm," 2022

International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 1591-1600, doi: 10.1109/ICSCDS53736.2022.9760859.

[2] Shailesh Tiwari · Erma Suryani · Andrew Keong Ng · K. K. Mishra · Nitin Singh Editors Proceedings of International Conference on Big Data, Machine Learning and their Applications ICBMA 2019.

[3] Sharma, Manish and Deshmukh, Rutuja, "Intelligent Recommendation System to Evaluate Teaching Faculty Performance Using Adaptive Collaborative Filtering" John Wiley & Sons, Ltd, isbn 9781119842286, Object Detection by Stereo Vision Images,chapter 9,pages159-169.

[4] Chandan G, Ayush Jain, Harsh Jain, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV", presented at International Conference on Inventive Research in Computing Applications (ICCSP), pp. 0676-0680, 2018- 19.

[5] Keerthana T , Kala L , "YOLO based Detection and Classification of Objects in video records", 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,IEEE Sensors Journal, vol. 15, no. 5, pp. 2679- 2691.(RTEICT) 2018, India.

[6] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, "Object Tracking Algorithms for video surveillance applications", International conference on communication and signal processing (ICCSP),Science,pp.7-30, India, 2018.

[7] Mukesh Tiwari, Dr. Rakesh Singhai, "A Review of Detection and Tracking of Object from Image and Video Sequences" ,International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, pp. 745-765,2017.

[8] Manjunath Jogin, Mohana, "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning", 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT)pp.593– 605,2018.

[9] Wei Liu and Alexander C. Berg, "SSD: Single Shot MultiBox Detector", Google Inc.,(ICCV), Venice, pp. 2980-2988. Dec 2016.

[10] Leonardo Bombonato ; Guillermo Camara-Chavez ; Pedro Silva , "Real-time single-shot brand logo recognition", published in: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) pp. 0676-0680. 2017.

[11] Amanjeet Kaur; Lakhwinder Kaur ,"Concealed weapon detection from images", published in: 2016 Online International Conference on Green Engineering and Technologies (IC-GET) VLSI, pp. 289- 295.2016.

[12] Ce Gao ; Liru Guo ; Miao Sun ; Xingfang Yuan ; Tony X. Han ,"Age Group and Gender Estimation in the Wild With Deep RoR Architecture", published in IEEE Access(ICCSP) pp.0570-0575. 2017.

[13] Dinh Viet Sang ; Le Tran Bao Cuong ; Do Phan Thuan," Facial smile detection using convolutional neural networks", published in 2017 9th International Conference on Knowledge and Systems Engineering (KSE)pp. 1606-1611. 2017.

Section A-Research paper
ISSN 2063-5346