iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52146

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Smart Video Surveillance Using YOLO Algorithm and OpenCV

Anubhav Sharma[1], Akash Kumar[2], Anubhav Shail[3], Aryan Kumar[4], Aparna Singh[5], Akash Verma[6]
Department of Computer Science & Engineering, IMS Engineering College Ghaziabad, U.P

Abstract: ***Due to object detection's close relationship with video analysis and image understanding, it has attracted much research attention in recent years. Traditional object detection methods are built on handcrafted features and shallow trainable architectures. Their performance easily stagnates by constructing complex ensembles that combine multiple low-level image features with highlevel context from object detectors and scene classifiers. With the rapid development in deep learning, more powerful tools, which can learn semantic, high-level, deeper features, are introduced to address the problems existing in traditional architectures.***

These models behave differently in network architecture, training strategy, and optimization function, etc. In this paper, we provide a review of deep learning-based object detection frameworks. Our review begins with a brief introduction to the history of deep learning and its representative tool, namely Convolutional Neural Network (CNN). Then we focus on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance further. As distinct specific detection tasks exhibit different characteristics, we also briefly survey several specific tasks, including salient object detection, face detection, and pedestrian detection. Experimental analyses are also provided to compare various methods and draw some meaningful conclusions. Finally, several promising directions and tasks are provided to serve as guidelines for future work in both object detection and relevant neural network-based learning systems.

Keywords: **Machine Learning, Neural networks, Convolutional Neural Network (CNN), YOLO Algorithm.**

I. INTRODUCTION

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform efficient tasks without using specific guidelines and relying on models and assumptions. It is considered a subset of intellectual intelligence. Machine learning algorithms create mathematical models of data models called "databases" to make predictions or decisions without performing the task explicitly. Machine learning algorithms are used in many applications, such as email filtering and computer vision, where it is not possible to create algorithms for certain instructions to work.

Machine Learning [15] is closely related to statistical analysis, which focuses on using computers to make predictions. Mathematical optimization work provides methods, theories, and applications for machine learning.

Data mining is a field of machine learning that focuses on data discovery through unsupervised learning. In its application to business problems, machine learning is also known as predictive analytics. Intelligent video surveillance can monitor activity, behavior or other changes in the environment. This includes remote monitoring via electronic devices.

Video monitoring can be done in real time or data can be collected and stored for evaluation as needed. An Intelligent Video Surveillance System (VSS) usually includes: a camera and an output device such as a monitor. video surveillance requirements include: storage, encoders, interfaces and control software. YOLO is a successful mindfulness time experiment described in a 2015 report by Joseph Redmon et al. "You Only Look Once: Unified Real-Time Object Detection". In this article, we introduced the concept of object detection, the algorithm itself, and one of its open implementations.

Image classification is one of many exciting applications of neural networks. Beyond simple image classification, there are many interesting problems in computer vision, of which object detection is one of the most interesting. It is often associated with self-driving cars, where the system includes computer vision, lidar, and other technologies to create a multidimensional representation of the road and all its participants. Object detection can also be used in video surveillance, especially for crowd monitoring to prevent terrorist attacks, people counting for statistics, or analyzing customers on a walk in the mall. It provides intelligent video surveillance, automatic perimeter monitoring and secure area protection. Regularly monitor user sites for human or vehicle access. The cross line function automatically detects moving objects crossing a defined line.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

II. APPLICATIONS OF SMART SURVEILLANCE

In this section, we introduce some applications for monitoring smart devices. In this section, we describe some applications. We've divided apps into three broad categories: real-time alerts, automated video calls, and situational awareness.

A. Real Time Alerts

Intelligent Surveillance System can generate two types of alerts, user alerts and automatic alerts for abnormal activity.

1) User-Defined Alerts

Here, the system should be able to detect various events occurring by the user in the monitoring area and notify the user in real time, thus giving the user time to evaluate the situation and take action if necessary. Below are some situations.

1.General alerts: These alerts are only dependent on the movement of objects in the monitoring area. A few examples are given below.

a. Motion Detection: This alarm detects the movement of any object in space.

b. Abandoned Object Alert: This detects[1] abandoned objects such as an empty suitcase at an airport or a car parked on a loading ramp.

c. Object removal: This detects the movement of an object specified by the user that they do not want to be moved, such as a painting in a museum.

2. Class-specific warnings: These warnings use object types in addition to dynamic objects. A few examples are given below.

a. Type Specific Movement Detection: Consider a camera monitoring an airport runway. In this case, the system can alert to the presence or movement of people on the pavement, but not to the presence or movement of the aircraft.

b. Statistics: Example application includes statistical[2] reports (eg. For example, having more than one person in a secure locker) or crowding (for example, a disco is more crowded than is acceptable).

3.Behavioral Alerts: These alerts are based on following or deviating from the pattern of physical activity. Such models are usually trained by analyzing long-term movement patterns. These alerts require special handling and seem to use too many data points like

a. Check the crowd at the merchandise store and alert the store manager when the queue on the shelf exceeds the specified number.

b.Check for suspicious behavior in the parking lot, such as someone pulling over and trying to open more than one car.

4. High Value Video Capture: This is an application that enhances real-time alerts by capturing selected videos based on predefined criteria. This becomes important in the context of smart cameras using wireless communication.

2) Automatic Abnormal Activity Alerts

Unlike user alerts, here the system generates an alert when it detects "activity deviating from the norm". Intelligent analytics does this as a normal working "learning" model. For example, a smart monitor monitors the road and knows "trucks walking on the road" and "people walking on the road". According to this mode, the system will give an audible warning while the vehicle is on the road. This type of discovery is important for cognitive monitoring because users cannot describe all the situations they are interested in.

B. *Automatic Forensic Video Access (AFVR)*

The ability to support video forensic access is based on a rich video directory generated by automatic search techniques. This is an important added value of using smart monitoring devices. Generally, this measurement includes metrics such as product shape, size, and data such as physical properties over time, product type data, and sometimes fabric information about specific products. In advanced systems, indexes will contain activity data. The Washington DC sniper incident is a prime example of how AFVR [10,11] can be a useful technology. At the time of the incident, investigators took hundreds of hours of security footage from multiple security cameras covering the area near the various events. However, if the video collection is evaluated by visual inspection, the video can be recovered in as follows:

1) *Spatial-Temporary Video Retrieval:* An example query in this class is retrieving a "blue car" that was driving in front of the "7/11 store on 23rd Street" between 2pm and July 27th.
2) *Surveillance video Mining:* In a sniper incident the application will try to watch all parts of the video, covering a few scenes from the camera setup to present the user with the car's mobility system. of the researcher "Is there a car that is everywhere?

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. *Situational Awareness*

Ensuring complete security requires systems that monitor the identity, location and movement of people and vehicles in the surveillance area. For example, current surveillance tools cannot answer questions such as: Does someone often pass by security buildings? This persistent monitoring can be the foundation of enhanced security. Generally, tracking systems focus on location and activity tracking, while biometric systems focus on identifying individuals. According to the analysis of smart technology, it will be possible to solve all three of these important problems in a unified way, resulting in a unified analysis system. The application forms the basis of state information.

III. YOLO ALGORITHM

There are many types of search algorithms that can be divided into two groups:

1) *Classification-based Algorithms:* They are used in two stages. First, they choose an area of interest for the picture. Second, they used a convolutional neural network to classify these regions. This solution can be slow because we need to make an estimate for each selected region. A good example of such a technique is the region-based convolutional neural network (RCNN) and its cousins Fast-RCNN and Faster-RCNN [12-14].
2) *Regression-based Algorithms:* Instead of selecting a part of the image of interest, they predict the class and bounding box of all images in one run of the algorithm. The most famous example of this algorithm is YOLO ("You Only See One"), which is often used to detect real objects.To understand YOLO algorithm, it is necessary to determine which one is correct. Finally, our goal is to guess a box indicating the class of an object and the position of the object. Each bounding box can be identified using four identifiers:

a) The width of the bounding box (bxby)
b) Width (bw)
c) Height (bh)
d) The value c corresponds to the category of a product.

IV. PROPOSED WORK

A. *Requirement Analysis*

1) *Scope*

The main purpose of this function is to determine the dropped load and send a notification after a certain data time limit.

2) *Feasibility Study*

Feasibility study is important as it helps analysis and development [6,7]. An analyst's decision whether to install a particular system depends on the feasibility study he or she does. Feasibility studies are done when it is possible to improve existing systems or install new systems. Feasibility studies help meet customer needs.

a. Financial Capability: The purpose of this product is to be able to carry out this project clearly and with all the necessary elements, and to ensure that the financial tightness of the project will be ensured large scale.

b. Technical Feasibility: High performance in terms of machine learning needs to be as high as possible. Our goal is to achieve maximum performance using for low power consumption.

c. Functionality: What is the main function of image processing and how it can assist in number recognition should be discussed in this report.

3) *Software and Hardware Requirements*

Hardware Specifications: -

- Processor: I7 9th Gen
- RAM: 16GB
- Hard Disk: 1 TB
- GPU: GT 1030 and above, intel HD Specifications 5444 & above
- Operation System : Windows, Linux
- Technologies Used: Machine Learning, Neural Networks, TensorFlow

B. Problem report

Video surveillance at current location only provides images of location. It can be difficult to keep track of suspicious things in a crowded place. It is necessary to manually monitor and monitor the area, which is a laborious task. Create a system to monitor and monitor the situation.

The system we recommend will look forward to the following resources:

• It will send a notification when it finds suspicious packages.

• In our project example, the suspect will be an abandoned bag/luggage.

• It will send an alert if a bag is tracked in the frame for a certain period of time.

The aim of this project is to create a model that can recognize and identify suspicious objects in live video streams using the principles of Convolutional Neural Networks. The main purpose of this system is to understand Convolutional Neural Networks and apply them to tasks in TensorFlow.

C. Project Design

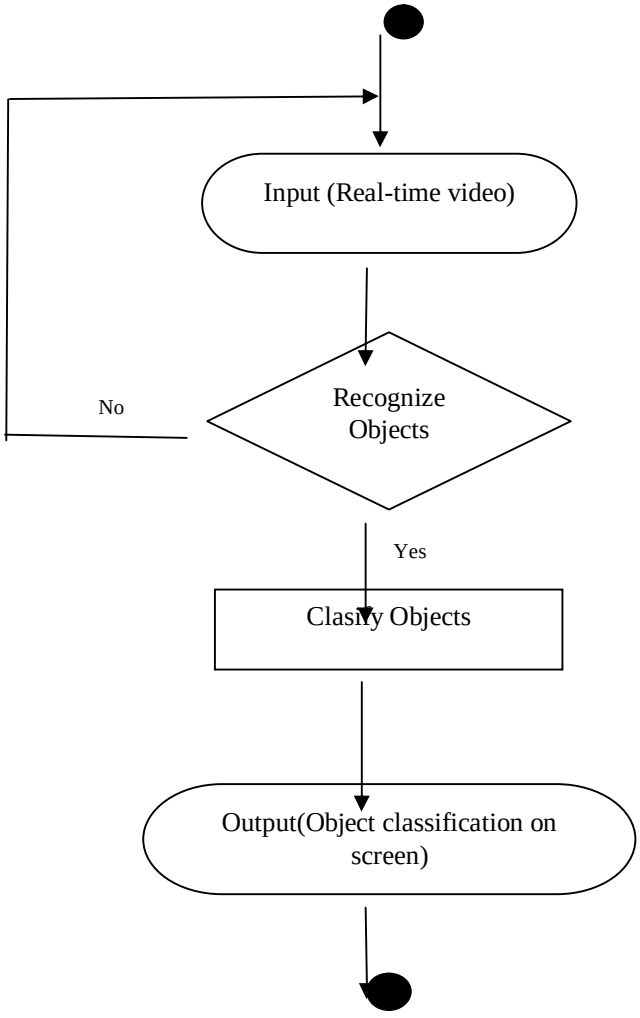


FIG 1. ACTIVITY DIAGRAM

D. Implementation Plan

1) Initial-Stage

In the initial stage, different Training models were used and comparison was made between the output of several models such as faster r-CNN [4, 16], YOLO, ssd300 and faster YOLO. Before switching to real hardware/software system tests were made on google Collab. The initial stage is divided into two sections image and video (not the live video streaming). In the first section, images were given as input and made to detect the objects in it. In the second section, a video was given as input and was made to detect the objects in the video given as input. The training was made to detect the still objects in the video.

2) Second-Stage

In the second stage live video was given as input to detect the objects in it. In the second stage live video streaming tool place and it detected the objects successfully by the webcam of the laptop. For this purpose, YOLO v3 [5] has been used.

Image Classification aims to assigning an image [8] to one of different number of categories (e.g. car, dog, human, etc.), essentially asking "what's is in this picture?". an image has only one category assigned.

Object localization then allows us to locate our object in the image, so our question changes to "what is it and where it is?". In a real instances, we need to go beyond locating just not one object but rather multiple
objects in one image.

Object detection provides the facility to finding all the objects in an image and drawing the bounding boxes around them. There are also some situations where we want to find the exact boundaries of our objects in the process called instance segmentation, but this is a subject for another article.

V. FUTURE SCOPE

Despite the rapid development and progress in object detection, there are still many open questions for future work. The first is the small object that appears in the COCO dataset and the face search function. To improve the localization accuracy of small objects in partial shutdown, it is necessary to change the network architecture from the following.

- Multitasking co-optimization and multimodal information aggregation. Because there is a lot of social discrimination.
- Customize. Objects are often found at different scales, which are more prominent in face detection and pedestrian detection.
- Spatial correlation and context modelling. Spatial distribution has important role in object detection. Therefore, the area concept generation and regression grid are used to obtain the product position. However, the relationship between various concepts and product categories is neglected. Additionally, the location-aware score map in the RFCN dispenses with the overall data model.
- Cascading networks. In a cascading network, cascading detectors are created at different levels or layers. Reject simple layer instances so that later features and products can use the decisions of previous stages to solve more complex models.
- Unsupervised and Weakly Supervised Learning. Manually drawing lots of bounding boxes is very time consuming. To reduce this overhead, feature prioritization, unsupervised object detection [9], multivariate learning, and deep neural network prediction are combined to achieve image-level processing.
- Network optimization. Given a particular application and platform, it's important to balance speed, memory, and accuracy by choosing the best detection architecture. However, despite the obvious reduction, it makes more sense to review compact models and this can be resolved by introducing better pre-training, knowledge dissemination and teaching.
- 3D visualization. With the application of 3D sensors [2,3] (eg. LIDAR and cameras) can use depth data to better understand 2D images and extend image-level information to the real world. However, few of these 3D sensing techniques focus on placing a 3D bounding box around the detected object.
- Video object detection. Temporal information in different frames plays an important role in understanding the behavior of different objects.
-

VI. CONCLUSIONS

Thus, we have learned:

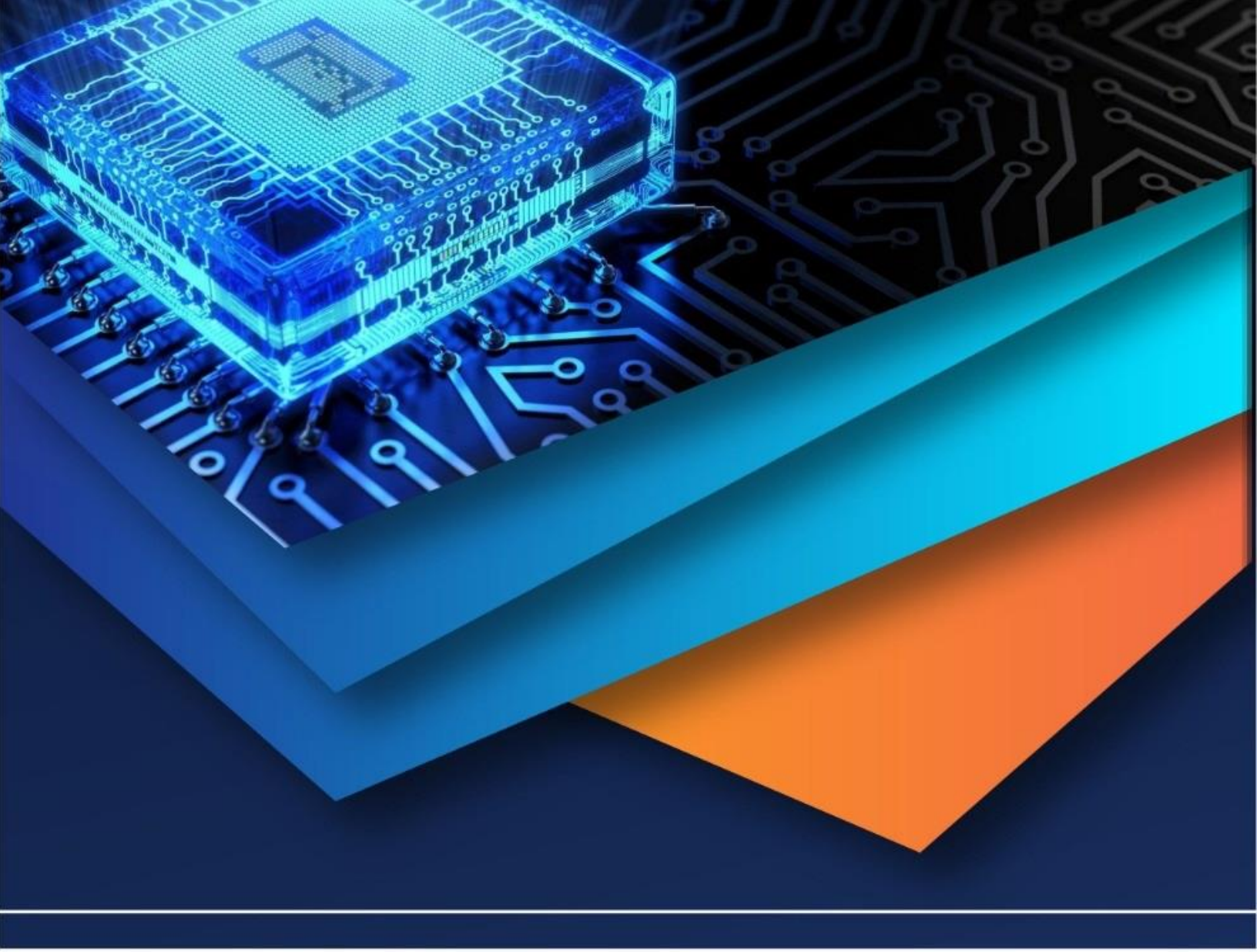
1) To establish a system that will track and monitor the scene.
2) Detection of objects using trained models.
3) Real-time detection of any object and things instantly.
4) Includes facial detection, movement detection, people counting and etc.

VII. ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my project supervisor Mr. Anubhav Sharma, Assistant Professor, IMS Engineering College, Ghaziabad, for his supervision, encouragement, suggestions, and trust throughout the development of this project. We are grateful to my project committee member for reading our project work and helpful comments. We would also like to thank our project coordinator Dr. Nitin Sharma for their valuable suggestions and advices in carrying out this work.

REFERENCES

[1] " Objects Talk - Object detection and Pattern Tracking using TensorFlow " University of Oulu, Degree Programme in Mathematical Sciences by P. Mustamo (2018).
[2] "Efficient Detection of Patterns in 2D Trajectories of Moving Points" by Joachim Gudmundsson, Marc J. van Kreveld, BettinaSpeckmann(2007).
[3] " The automatic detection of patterns in people's movements" by Gordon Forbes, GerharddeJager(2002).
[4] "Object recognition in images using convolutional neural network" by Duth P. Sudharshan ,SwathiRaj
[5] "Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices" by BERNARDO AUGUSTO GODINHO DE OLIVEIRA,FLÁVIA MAGALHÃES FREITAS FERREIRA, AND CARLOS AUGUSTO PAIVADA-SILVA-MARTINS(2017).
[6] ‘‘Region lets for generic object detection,’’ by Wang, M. Yang, S. Zhu, and Y. Lin, IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 10, pp. 2071–2084, Oct. 2015.
[7] "Improvements of object detection using boosted histograms," by Laptev, in Procedure BMVC, vol.3., pp.949–958. (2006).
[8] ‘‘Text classification using WordNet hypernyms,’’ by S. Scott and S. Matwin, in Proc. Conf. Use WordNet Natural Lang. Process. Syst., 1998, pp. 38–44.
[9] ‘‘Object recognition from local scale-invariant features,’’ by D. G. Lowe, in Proc. IEEE Int. Conf. Compute. Vis., vol. 2. Sep. 1999, pp. 1150–1157.
[10] ‘‘Learning algorithm for non-linear support vector machines suited for digital VLSI,’’ by D. Anguita, A. Boni, and S. Ridella, Electron. Lett., vol. 35, no. 16, pp. 1349–1350, Aug. 1999.
[11] ‘‘Architecting the next generation of service-based SCADA/DCS system of systems,’’ by S. Karnouskos and A. W. Colombo, in Proc. 37th Annu. Conf. IEEE Ind. Electron. Soc. (IECON), Nov.2011, pp.359–364.
[12] ‘‘A performance study of general-purpose applications on graphics processors using Cuda,’’ by S. Che, M. Boyer, J. Meng, D. Tarjan, J. W. Sheaffer, and K. Skadron, J. Parallel Distribution Compute., vol.68, no.10, pp.1370–1380,2008.
[13] ‘‘cu DNN: Efficient primitives for deep learning.’’ by S. Chetluret al. (2014). [Online]. Available: https://arxiv.org/abs/1410.0759
[14] ‘‘Image Net: A large-scale hierarchical image database,’’ by J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, in Procedure IEEE Conf. Compute Vis. Pattern Recognition. (CVPR), Jun.2009, pp.248–255.
[15] "Machine Learning: Parallel and Distributed Approaches." by R. Bekkerman, M. Bilenko, and J. Langford, Eds., Scaling up, Cambridge, U.K.: Cambridge Univ. Press, 2011.
[16] “Object Detection with Deep Learning” A Review Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shoutao Xu, and Xindong Wu, Fellow, IEEE



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)