iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49210

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 II Feb 2023- Available at www.ijraset.com

A Literature Review of Object Detectionusing YOLOv4 Detector

S Bhupinder Singh Saini[1], Shamsher Tiwari[2], Naman Gupta[3], S Jaishree[4]
Department of Computer Science, Dayananda Sagar college of Engineering, Bangalore, India

Abstract: ***Object detection is an advanced form of image classification where a neural network predicts objects in an image and points them out in the form of bounding boxes. Compared to the approach taken by object detection algorithms before YOLO, which repurpose classifiers to perform detection, YOLO proposes the use of an end-to-end neural network that makes predictions of bounding boxes and class probabilities all at once. Object detection not solely includes classifying and recognizing objects in an image but also includes localizing those objects and attracts bounding boxes around them. Its application includes field like face detection, detecting vehicles, autonomous vehicles and pedestrians on streets.***
Keywords: YOLOv4, Object detection, classificationTensorFlow, computer vision

I. INTRODUCTION

Eyes are one of the most important necessary tools that any creature has. laptop vision additionally deals with object finding, image recognition, image super- resolution, and so on. In the past, long-term object recognition has been used in many areas, such as in voice-controlled buses, roving tours, facial recognition and different security systems. With progress in software and hardware, object detection has exploded among large advanced systems. Once an object is found, the second task to be done is to accurately draw boundaries around all detected objects. If it's possible to research techniques that rely on finding objects as part of a mission, accurate bounding boxes make it easy to scale the value of that gear. tone control rails and the robot system use many objects detection software to distribute accessories to the exact location of the object. When one's life is in the hands of technology, that's why millimetre perfection matters. Many experimenters have many methods that count subjects from video. vary from them to use several ways, that where and there the single area mixes and angles between completely different methods.

Background Deduction is a system that extracts a moving object of interest from the video. Background subtraction is full of significantly non-static background and lighting changes. In the application, this charge is removed using an optical flow rule algorithm, but it provides a warning for irregular conditions. In most cases, the regional commodity tracking unit will tell the background information in cases of background deduction, but this will cause an error in the classification. A big part of choosing robust classifiers is the challenge of increasing the sensitivity of the algorithmic rule. To deal with this problem CNN based algorithm came into light.

II. LITERATURE SURVEY

1) S. Mane et. el. in [1] The proposed approach can recognize objects in different ways. The proposed approach achieves that 90.88% accuracy on selfgenerated image sequences. A frame usually contains a background and a foreground This foreground object is Main procedures for motion detection, movement, etc. Background Subtraction is a method of extracting the motion of an Object from video frame. the object Trackersare affected by background information, but we have to increase the accuracy of the claalgorithm. this Method uses a TensorFlow-based object detection and CNNA based object tracking algorithm Approaches are reliably detected and tracks objects inside Complex scenes and complex background conditions.

2) B N Krishna sai et. al. in [2] This paper is regarding the object detection for threatening objects in a given input image. The object detection is just not regarding detecting objects for any image but also includes localizing those objects and defining bounding boxes around them. For training the model, this paper uses the Tensorflow object detection API, and later they use faster R-CNN(Region-based convolutional neural network) algorithm for implementation. Here, they have simplified the classification by just taking two objects for classifying which are knife and guns. By using Fast R-CNN, you can train an entire image with just one CNN rather than training each image with multiple CNNs. In this paper, they have taken image dataset(78) from google and then split the data in test and train set. From the test cases, they noticed the more we train the model better the performance. They have taken 11 images for test among which 7 are knife and 4 are guns, and they are manage to achieve 81% accuracy. They have train model around 4500 steps to get loss under 0.1 which took them almost 12 twelve hours. The future work involves training the model with more dataset and also training it with more steps to achieve better results.

3) Huy Hoang Nguyen et. al. in [3] Here we talk about the video surveillance using image processing to detect humans in edge devices. Since edges devices have limited resources we need to take proper consideration before implementation of any algorithm since they could be very costly(computational).This paper is particularly motivated by you only look once(YOLO),Spatial pyramid pooling(SPP) and residual learning. They have focuses on designing a network structure so that their model could have good trade-off between accuracy and processing time. Their results showed that their model could process at least 2FPS on raspberry PI with accuracies of around 95% and 96% with INRIA and PENN FUDAN datasets. It is able to outperform the Tiny-YOLO versions and SSD based L-CNN method. In order to assess the detection performance, average precision (AP) was used which summarizes the shape of precision/recall curve. The residual blocks and SPP modules though makes the size of the network model a little bigger and it also requires more floating point operations, but it definitely improves the accuracy of human detection.

4) Shaji Thorn Blue et. el. In [4] The following method takes YOLOv3 as a root model and improves the accuracy of bounding boxes around objects. The given method uses a pre-trained COCO dataset. As for the detection method, ConvNet-based region proposal YOLO turns MultiBox into a local proposition Solution for object detection method by adding SoftMax Predict object classes directly. Additionally, YOLO use a grid A bounding box detection approach. This method takes an image as input YOLOv3 algorithm. YOLOv3 consists of 106 layers not including that layer. YOLOv3 Convolutional Neural Network Algorithm takes an image and returns a tensor after processing represents the coordinates points of the box and the probability Each box contains an object and object class. Bounding box is excluded with nonmaximum suppression used for detection. used for detection. Used for detection. Used for detection of 80 different classes of object recognition.

5) A. Raghunandan et. el. In [5] This paper talks about various known face, skin, color and shape detection algorithm and then simulate them using MATLAB2017b to locate and identify various objects for video surveillance. Object detection involves identification of real-world objects like people, objects, animals or threatening objects. This technology is commonly used in fields of security, defense, medical field etc. Fundamentally, object detection is extracting out the foreground object from a given background frame.A foreground item is interesting because of its appearance or local motion.. It tends to change from frame to frame. Whereas a background object is just stationary object in a frame. At first, a foreground item is identified by performing a background subtraction operation on the current video frame. All face features—eyes, mouth, and nose—were detected using the Viola Jones algorithm. For skin detection, We initially convert the input image from RGB to YCbCr and they detect the skin pixels which are represented either by '1' or '0'.Now for target selection the most widely accepted method for this Background subtraction which follows Bayes rules. Atlast target detection is done with detecting the desired object from the frame. The results were obtained for various Euclidean Threshold values (T).The foreground, the cleaned-up foreground, shadows and object of interest was detected from the original code with different values of T to get the optimal accuracy. MATLAB 2017b is used to simulate different object identification techniques such as skin detection, color detection, face detection, and target detection with accuracy above 95% at least. The detection accuracy, RGB Euclidian Threshold 'T' in Target Detection, and Y, Cb, and Cr in Skin Detection parameters have all been modeled and implemented to improve the efficiency of algorithms for video surveillance applications.

6) Htet Aung et. el. In [6] This research paper is using a deep learning-based hybrid method for face detection. Firstly, they are trying to perform feature extraction with help of VGG16 pre-trained convolutional network. The output of feature extraction is combined with the object detection algorithm YOLO for face detection. They have managed to achieve high accuracy of 93%, a lot greater than the google network and Alex network model working with a combination of the same YOLOv2. The proposed method mainly consists of 3 different stages: building a database of labeled ground truth, features extraction stage and face detection stage. To create the ground truth database we could utility like image Labeler Toolbox where labels could be added either manually or automatically by image analysis. A pre-trained VGG16 convolutional neural network with 41 distinct layers was employed for the feature extraction, but since we are using it only for feature extraction so all the unnecessary layers of VGG16 model are being removed and only the first 32 layers remain in the model. The output of the VGG16 network is mixed with the YOLOv2 algorithm in the last part. The biggest advantage of YOLO network is that it needs to pass through the image only once to find all the desired object and that's why YOLO is very fast model. The ground truth dataset contains 5000 images which are people in different poses, illuminations, skin and colors.In the training process, they split the dataset in 80 and 20 ratio for training and testing set. The model was trained with different optimizer like Adam solver and SDGM solver with 86% and 98% average precision respectively on test set. The only disadvantage of this method is the necessity of GPU graphic cards for high computational need. One possible future work could be using this method as a pre-processing step of face recognition systems.

7) Chen Zhaoyang et. el. In [7] This paper mainly aims in increasing the performance of motion and object detection in security monitor.Here object detection is found based on its motion.firstly they tried to improve the accuracy of motion detection and then added the result of it with object detection. The advantages of this method are: improved the detection of tiny target, reduced the negative detection rate ,false detection, and decreased the computing cost of detection. Here they mainly focus on detecting the changes in pixel, they created background and foreground image segments using background modeling algorithms like inter-frame difference method, Gaussian mixture model, background subtraction method. object detection network is divided into two categories like 1)regional recommendation called two-step method using convolution neural network 2) end-to-end-one-stage method because it directly regress the border of the target. Dataset containing 5000 images were used.The model is trained and were labeled and detection network was trained until the detection result is achieved. Finally, the object category and boundingbox are obtained.

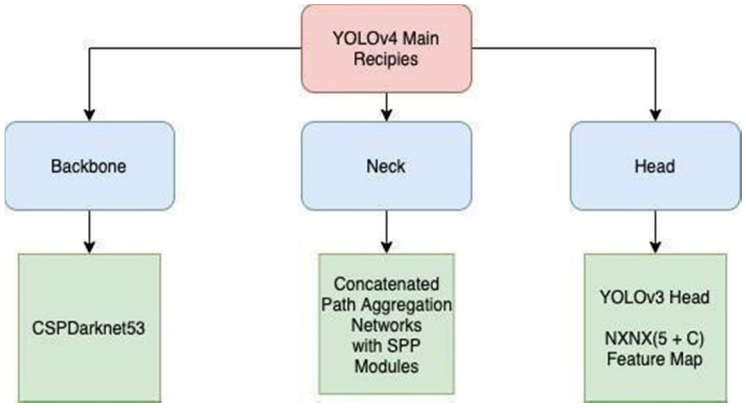
8) T. Hui Teo et. al. In [8] Autonomous vehicles using AI will require object detection on road for visual perception input. sensors produce huge quantity of data that should be calculated with the cloud computing wireless technologies but real time object detection with latency less than 0.2ms needed is achieved using deep learning. Here primarily they would handle the distortion and then use object detection algorithms like Deformable Part Models, Region-based CNN (RCNN), You- OnlyLook-Once(YOLO), Single Shot Detector, RetinaNet. They Created three training Datasets each datasets with different range of distorted and normal images are formed and tested.In future determining the extent to what the input images are distorted can be achieved.

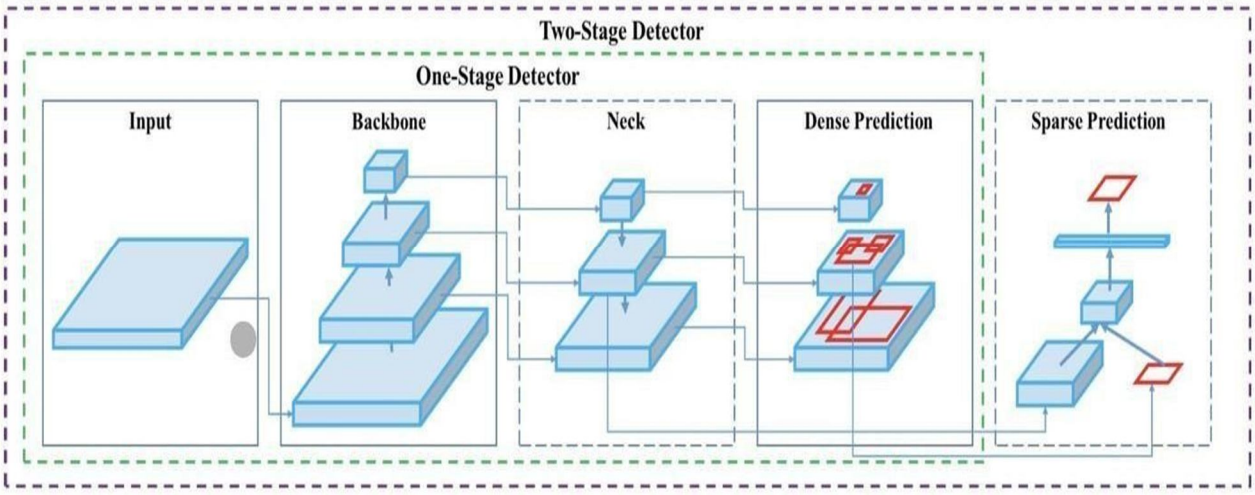
9) Gong Cheng et. al. in [9] In the field of satellite and aerial image observation, object detection in optical remote sensing images is a basic. It is important for a variety of applications . They analyzed 270 papers and looked at two types of object detection techniques: knowledge-basedand machine learning based templatematching, object-based image analysis. In template matching they have rigid and deformable template, check for similarity measure .For Knowledge-based object detection rules establishment and then hypothesis testing is done. In OBIA image is segmented and classification of object, Machine learning based object detection feature is extracted and classifiers like SVM, Adaboost, KNN , conditional random field are used. Datasets NWPU VHR dataset is used and evaluation metrics using precision ,recall.

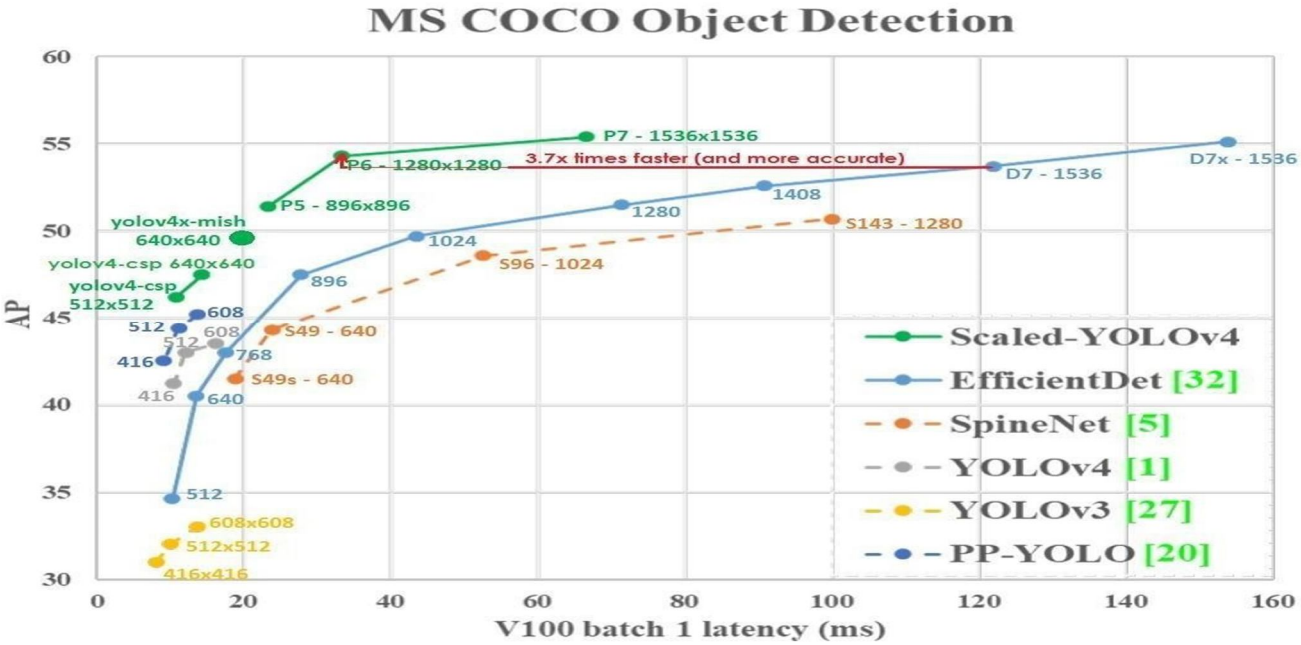
10) Yang Liu et. el. In [10] With the quick improvement of deep convolutional neural networks in computer vision, significant improvements in object detection have been made (CNN). In-depth analyses of recently created deep learning techniques for tiny object detection are provided in this work. We cover key deep learning approaches, such as fusing feature maps, providing context information, balancing ground segment instances, and producing enough positive examples, and we review problems and solutions related to tiny itemrecognition. We present relevant methods that have been developed in four study fields, such as segmentation, face ,object detection in aerial photography, and generic object detection. Additionally, based on three sizable benchmark datasets of tiny objects, this article evaluates the results of many top deep learning techniques for small object recognition, including SSD, Faster R-CNN, and YOLOv3.

11) Jung Uk Kim et .el in [11] Image and Video Systems Lab, School of Electrical Engineering, KAIST, South Korea .One of the main areas of computer vision is object detection. The duties of object categorization and localisation are carried out during object detection. Performance using feature maps of previous deep learning-based object detection networks generated by networks all of which are entirely shared. The object classification process, on the other hand, concentrates on the feature map's most important object region. Directly in opposition, object localisation initiates a feature map that is concentrated on the full object's region.In this study, they evaluated the difference between the two objectives and present an unique object detection network. The two primary components of the proposed deep learning-based network are 1) the attention network portion, which generates task-specific attention maps, and 2) the layer separation part, which divides the layers for estimating two tasks. PASCAL VOC dataset and MS COCO datasets are used.using datasets they introduced attention network to generate the two task-specific attention maps. Features were fed into layer separation and task was performed. Rakshitha Gopal et. el. In [12] The identification of objects, especially tiny ones from a distance, is extremely difficult because of the movement of the vehicle and the variety of item dimensions. Here they have tried to detect using single stage CNN detectors. Three versions, including SSD300 and its variations, Tiny-YOLO V3, and RetinaNet, were compared. The networks were evaluated after being trained using 7082 example pictures.in 6500 pictures. At a distance of three to four meters, detection was tested using pictures of various scale and rotation and a single class object with a diameter of between two and three inches (20*20 pixels to 40*50 pixels).With 60% accuracy, a speed of 0.09 SPF, and a model size that satisfies the memory restriction, Tiny-YOLO V3 performs significantly better.

III. DATA FLOW DIAGRAM





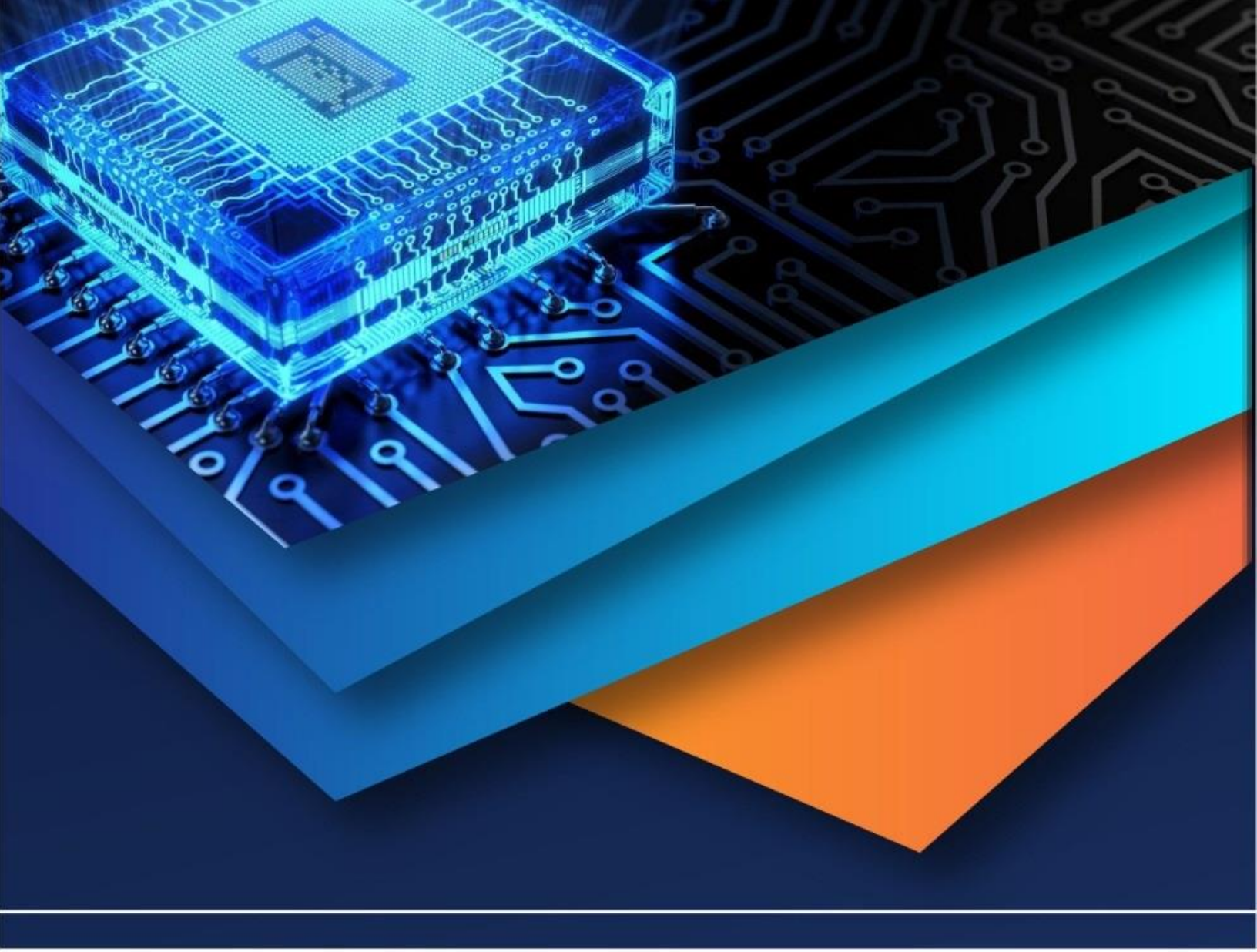


IV. CONCLUSION

Object detection methods are getting lot of attention in multiple research fields. Multiple object detection in a video surveillance is a difficult procedure that depends on timings and the density of the items in the monitoring area or on the road. Multiple object detection is a crucial capability for most computer and AI (Ai - based) vision systems. This paper implements a multiple object detection algorithm, which is suited for traffic and numerous surveillance applications. In this paper we have tried to implement the object detection using YOLOv4 detector, reason being is its ability to pass through the image only once to find all the desired object and that's why YOLO is very fast model. It may also be trained and used on a standard GPU with 8–16 GB of VRAM, enabling a wide range of applications. We have manage to achieve accuracy of 95-98% using this method. There is still potential for improvement when it comes to multiple and single item recognition with greater accuracy, despite the fact that we were able to improve the algorithms connected to object detection jobs. The future work for this could be around the detection of license plate recognition for the live traffic feed.

REFERENCES

[1] Shraddha Mane, Prof.Supriya Mangale "Moving object detection and tracking Using Convolutional Neural Networks" Proceedings of the Second International Conference on IntelligentComputing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3
[2] B N Krishna Sai, Sasikala T., "Object Detection and Count of Objects in Image using theTensor Flow Object Detection API" Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019) IEEE Xplore Part Number: CFP19P17-ART; ISBN:978-1-7281-2119-2
[3] Huy Hoang Nguyen, Thi Nhung Ta, NgocCuong Nguyen, Van Truong Bui, Hung Manh Pham, Duc Minh Nguyen "YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge" 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE) |978-1-7281-5471-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCE48956.2021.9352144
[4] Shaji Thorn Blue, M. Brindha "Edge detection based boundary box construction algorithm for improving the precision of object detection in YOLOv3" 10th ICCCNT - 2019 July 6-8, 2019,IIT - Kanpur Kanpur, India
[5] Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya "Object Detection Algorithms for Video Surveillance Applications" International Conference on Communication and Signal Processing, April 3-5,2018, India
[6] Htet Aung, Alexander V. Bobkov, Nyan Lin Tun "Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network" 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)
[7] Chen Zhaoyang, Gao Haolin, Wang Kun "A motion based object detection method" 2020 2nd International Conference on Information Technology and Computer Application (ITCA)
[8] T. Hui Teo, Yi Shu Tan "Fast Object Detectionon the Road" The 2020 IEEE Asia Pacific Conference on Circuits and Systems(APCCAS).
[9] Gong Cheng, Junwei Han "A survey on objectdetection in optical remote sensing images" ISPRS Journal of Photogrammetry andRemote Sensing 117 (2016) 11–28
[10] Yang Liu, Peng Sun , Nickolas Wergeles, Yi Shang "A survey and performance evaluationof deep learning methods for small object detection" Expert Systems With Applications172 (2021) 114602
[11] Jung Uk Kim, Yong Man Ro "ATTENTIVE LAYER SEPARATION FOR OBJECT CLASSIFICATIONAND OBJECT LOCALIZATION IN OBJECT DETECTION"978-1-5386-6249-6/19/$31.00 ©2019 IEEE
[12] Rakshitha Gopal, Sandeep Kuinthodu, Muthukumar Balamurugan, Mallabadkar Atique "Tiny Object Detection:Comparative Study using Single Stage CNN ObjectDetectors" 2019 5th IEEE International WIE Conference on Electrical and Computer Engineering(WIECON-ECE)
[13] Xia, Z. (2019). An Overview of deep Learning.Deep Learning in Object Detection and Recognition, 1-18.
[14] Leung, H., & Haykins, S. (1991). The complexbackpropagation algorithm. IEEE Transactionon Signal Processing, 39(9), 2101-2104.
[15] Real-Time Object Detection for Aiding Visually Impaired using Deep Learning. (2020). International Journal of Engineering and Advanced Technology Regular Issue, 9(4), 1600-1605.
[16] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[17] Zhang, D. (2018). Vehicle target detection methods based on color fusion deformable part model. EURASIP Journal on Wireless Communication and Networking.
[18] Zhao, L., & Wan, Y. (2019). A New Deep Learning Architecture for Person Detection. 2019 IEEE 5th International Conference on Computer and Communications (ICCC).
[19] Hung, P. D., & Kien, N. N. (2019) SSD-Mobilenet Implementation for Classifying Fish Species. Advance in Intelligent System and Computing Intelligent Computing andOptimization, 399-408.
[20] Andre, T., Neuhold, D., & Bettstetter, C. (2014). Coordinated multi-robot exploration:Out of the box packages for ROS. 2014 IEEE Globecom Workshops(GC Wkshps).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY