iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** V **Month of publication:** May 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.61802

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 12 Issue V May 2024- Available at www.ijraset.com*

# Object Detection Model: An Enhanced SSD

Damini[1], Aman Kumar Sharma[2]
*Department of Computer Science, Himachal Pradesh University, India*

***Abstract:*** ***Object Detection plays a significant role in today's era. Object detection plays a crucial role in various applications such as surveillance, medical images, and autonomous vehicles. The primary goal of object detection is to accurately detect the boundaries of objects of interest and classify those objects into predefined categories. In this research paper, we enhanced an object detection technique i.e. Single shot multibox detector (SSD) which is one of the top object detection technique in both aspect accuracy and speed. SSD is an object detection model that predicts object bounding boxes and class probabilities in a single forward pass. It uses multiple feature maps at different scales to cater to various object sizes and aspect ratios.The main contribution of this research is to enhance the accuracy, recall time, precision time and mean average time (mAP). The performance of this object detection technique is improved as the number of feature maps is increases, improved backbone, and enhanced activation functions, additional layers for enhanced detection and loss function modifications. An accuracy of 0.75 indicates that the model correctly predicts the class label for approximately 75% of all objects in the dataset. The enhanced model showed these results: a precision score of 0.75 implies that 75% of the objects identified by the model as positive (detected objects) are indeed true positives. A recall of 0.73 indicates that the model successfully identifies and detects 73% of all positive instances in the dataset. A mAP of 75.25 signifies the average precision across all classes, computed at different IOU thresholds (e.g., 0.5, 0.75).***

***Keywords: Object Detection, Single Shot Multibox Detector (SSD), Recall Time, Precision Time, Mean Average Precision (mAP), and Intersection over Union (IOU).***

## I. INTRODUCTION

During the last years, there has been a rapid and successful expansion on computer vision research [1] . In recent years, convolutional neural networks (CNNs) have been applied to object detection algorithms in various ways, improving the accuracy and speed of object detection [2]**.** Among various object detection methods, SSD [3] is relatively fast and robust to scale variations because it makes use of multiple convolution layers for object detection. Although the conventional SSD performs well in both the speed and detection accuracy, it has a couple of points to be supplemented. The single shot multibox detector (SSD) is based on a feed-forward CNN that generates bounding box sets and scores of different types on the boxes, followed by non-maximum value suppression to complete the final detection process. This explains the incorporation of both the regression concept in YOLO and the anchor mechanism in Faster RCNN in single shot multibox detector (SSD), as regression is performed on the multi-dimensional region features of every location in the entire picture, which retains YOLO's characteristics of being fast while ensuring the window prediction is as accurate as Faster RCNN [4]. SSD uses VGG1636 as the backbone feature extraction network, and makes certain modifications to VGG16 and adds extended convolution or perforated convolution to expand the field of view of convolution. And, SSD uses convolution with stride = 2 to reduce the size and resolution of the feature map, thereby obtaining six feature maps of different levels and scales, and convolving each layer of feature maps to obtain category confidence and bounding box positions.37,38 During the training of SSD, the prediction frame is obtained by using the method of non-maximum value suppression based on the prior frame.39–41 Non-maximum suppression is the combination of category confidence and intersection ratio and elimination of redundant frames to generate the final prediction frame [5]. Normally, the two-stage detector has higher accuracy and the one-stage detector is faster. Considering that the speed and accuracy of the detector are equally important in some scenarios, such as autonomous driving, how to achieve high accuracy at low-computational cost remains a challenging task. Another difficulty in object detection is the detection accuracy of small objects [6]. To solve these problems we enhanced the SSD model to eliminate these problems. Our Contribution is to enhance the SSD object detection model: Key Architectural Features to Integrate, Loss Function Modifications, Data Augmentation Techniques and Inference Optimizations.

The research paper is divided into sections wherein the first section is about the introduction of model. Section two is related work. Third section is taxonomy of models in which it highlights the object detection models i.e. SSD, YOLOv7 and YOLOv8. In section 4 The Enhanced Single Shot Multibox Detector (ESSD) is elaborated with its description. Section 5 onwards is the comparison, verification & validation of the enhanced single shot multibox detector (ESSD) with YOLOv7 and YOLOv8. Section 6 i.e.

environment where the experimental specification is highlighted. Model is performed with the aid of experimental whose details are in section 7. Section 8 has contained the results and analysis. In section 9 it conclude the conclusion of this research.

## II. RELATED WORK

The paper [7] "An Approach on Image Processing of Deep Learning Based on Improved SSD" by Liang Jin et al., proposes an efficient ship detection algorithm for SAR remote sensing images based on an improved version of the Single Shot Detector (SSD) deep learning model The author adding a feature pyramid network to SSD, which introduces context information to the large-scale feature maps responsible for detecting small objects. The author also adding a Squeeze-and-Excitation (SE) module to the model, which enables dynamic channel-wise feature recalibration to improve feature extraction and detection accuracy. The authors state that these improvements to the standard SSD model result in better detection accuracy, especially for smaller objects, without significantly impacting the inference speed. Experiments on benchmark datasets show the improved SSD algorithm outperforms the original SSD as well as other popular object detection models like Faster R-CNN and YOLO.

The paper [8]"SSD with multi-scale feature fusion and attention mechanism" by Qiang Liu proposes an improved Single Shot MultiBox Detector (SSD) algorithm that combines multi-scale feature fusion and an attention mechanism to enhance the detection performance, especially for small objects. This paper introduces an enhanced SSD algorithm (SSD+MFA), which employs diverse fusion methods for feature extraction across various scale layers. The author integrates a channel attention mechanism to redistribute the channel weights of the fused feature map. The experimental results: the mean Average Precision (mAP) on the PASCAL VOC2007 dataset achieves 90.57%, marking a notable improvement of 3.27% over the traditional SSD algorithm and 2.00% over YOLOv4.

This paper [9]“Object detection in real time based on improved single shot multi-box detector algorithm” by Ashwani Kumar et al proposed an algorithm to enable real-time object detection with high accuracy, especially for small objects. The author combines the multilayer convolutional network to achieve precise and swift object detection. This algorithm performs well on both still images and videos. The accuracy of the proposed model is more than 79.8%, with a training duration of around 5 to 6 hours. These convolutional neural networks analyze image features and then map them to classify objects. The main aim of our approach is to optimize the selection of default boxes by using the most suitable aspect ratio values, thereby enhancing the SSD algorithm's capability to detect objects effectively.

The paper [10] "Object Detection Based on the Improved Single Shot MultiBox Detector" by Songmin Jia proposes a deep learning framework that enhances the Single Shot MultiBox Detector (SSD) for object detection tasks. The study introduces improvements to the SSD algorithm to achieve more accurate object detection results. By enhancing the training model and optimizing the detection process, the proposed framework aims to provide better performance in object classification tasks. The experimental results: mAP of the proposed SSD object detection algorithm based on feature fusion i.e. 78.0%, which is higher than the SSD and Faster RCNN algorithms.

This paper [11] “Fast single shot multibox detector and its application on vehicle counting system", by Lili Chen et al, presents a vehicle counting system that utilizes the Single Shot MultiBox Detector (SSD) algorithm for real-time object detection The authors utilized the depth-wise separable convolution and spatial separable convolutions to the convolutional layers, which helps maintain computational speed while improving classification accuracy. The author used a larger number of default boxes compared to the original SSD, resulting in more accurate object detection. Experiment results: experiment is implemented on PASCEL dataset which showed that the improved SSD algorithm achieves high accuracy in terms of metrics like loss function, mean average precision (mAP), and frames per second (FPS), which is making it suitable for real-time vehicle counting applications.

The research paper [6] titled "Accurate and fast single shot multibox detector" by Lie Guo and Dongxing Wang focuses on developing an efficient single shot multibox detector for accurate and fast object detection. The proposed method aims to improve the efficiency and precision of detecting objects in various scenarios. By utilized the advancements in single shot detection (SSD) techniques, the research presents a method that combines accuracy with speed in detecting objects.

The paper [5] "Improved single shot multibox detector target detection method based on deep feature fusion" by Dongxu Bai that used the deep feature fusion to enhance the target detection performance. The author adopted a deep feature fusion approach, where features from multiple layers of the convolutional neural network backbone are combined to improve the feature representation capability. The author introduced a feature pyramid structure to capture multi-scale information, which helps address the challenges of detecting objects of varying sizes. The paper also highlights that the author optimized the anchor box design and loss function to further improve the detection accuracy. Experimental results: the proposed improved SSD algorithm with deep feature fusion outperforms the original SSD in terms of both detection accuracy, speed and suitable for real-time applications.

The research paper [2] titled " Enhancement of SSD by concatenating feature maps for object detection" by Jisoo Jeong, Hyojin Park, and Nojun Kwak proposes a method to enhance the accuracy of the Single Shot Multibox Detector (SSD) in object detection. The study focuses on improving the performance of SSD by effectively utilizing feature maps to enhance the network's accuracy without simply increasing the number of feature maps. In this paper, the authors proposed and analyzed how to use feature maps effectively to improve the performance of the conventional SSD. The enhanced performance was obtained by changing the structure close to the classifier network, rather than growing layers close to the input data, e.g. by replacing VGGNet with ResNet. Regarding the Pascal VOC 2007 test set that was trained using VOC2007, using the VOC 2012 training sets, the suggested network with a 300×300 input size attained 78.5% mAP at 35.0 frames per second. Using an Nvidia Titan X GPU, the network with a 512×512 input size was able to achieve 80.8% mAP at 16.6 FPS. Modern mAP is demonstrated by the suggested network, outperforming those of YOLO, Faster-RCNN, RFCN, and traditional SSD. It is also quicker than RFCN and Faster-RCNN.

The research paper [12] titled "A SINGLE-SHOT OBJECT DETECTOR WITH FEATURE AGGREGATION AND ENHANCEMENT" by Weiqiang Li focuses on improving the accuracy of the Single Shot Multibox Detector (SSD) by effectively utilizing feature maps. In this research, the author proposed a feature aggregation and enhancement (FAENet) based one shot object detector that is both accurate and efficient. The authors integrated two new feature enhancement blocks and two unique feature aggregation modules into the SSD's original architecture. Prolonged tests on the MS COCO and PASCAL VOC datasets show that the suggested approach outperforms SSD in terms of accuracy. The author also highlights that proposed method performs better than the state-of-the-art one-stage detector RefineDet on small objects and can run at a faster speed.

The research paper [13] titled "ASSD: Attentive Single Shot Multibox Detector" by Jingru Yi, Pengxiang Wu, and Dimitri's N. Metaxas proposes a novel deep neural network architecture for object detection. The proposed network, termed ASSD, focuses on building feature relations in the spatial space of the feature map to enhance object detection. This approach involves learning to highlight useful regions on the feature maps while suppressing irrelevant information, providing reliable guidance for object detection. Experimental results show that ASSD competes favorably with the state-of-the-arts, including SSD, DSSD, FSSD and RetinaNet. This paper also highlights that the ASSD utilizes a fast and light-weight attention unit to help discover feature dependencies and focus the model on useful and relevant regions. ASSD improves the accuracy of SSD by a large margin at a small extra cost of computation. In particular, it achieves better performance than the one-stage detector RetinaNet, while being easier to train without the need to heavily tune the loss parameters.

The research paper [14] titled "SS3D: Single Shot 3D Object Detector" a single-stage 3D object detection algorithm that utilizes a Single Shot Detector in combination with a simple input representation. With its single-stage approach, the proposed approach seeks to streamline the 3D object detection process, which is highly helpful for embedded and real-time applications. In comparison to state-of-the-art stereo-based detectors, the algorithm achieves competitive accuracy by combining stereo input and additional semantic segmentation information. The research reveals that the SS3D algorithm can achieve accuracy that is on able to compete with two-stage detectors and has the potential to be used in a variety of practical applications, such as those that require LiDAR and stereo input for 3D object detection.

The research paper [15] "Detecting Small Objects in Thermal Images Using Single-Shot Detector" by Hao Zhang and Xiang-gong Hong aims to enhance the Single Shot Multibox Detector's (SSD) ability to identify small objects in thermal images. One of the best object detectors is SSD, but because the shallow feature layers don't contain any semantic information, it has trouble identifying small objects. In order to improve small object detection, the paper suggests an improved version of SSD called DDSSD (Dilation and Deconvolution Single Shot Multibox Detector). DDSSD makes use of a unique feature fusion module that uses deconvolution to expand the size of high-level feature maps and dilation convolution to help expand the region of interest of shallow features. The suggested DDSSD method outperforms state-of-the-art object detectors, particularly for small objects where it achieves 79.7% mAP on PASCAL VOC2007 and 28.3% mAP on MS COCO.

The research paper [16] titled "Vehicle Target Detection Method Based on Improved SSD Model" by Guanghui Yu, Honghui Fan, Hongyan Zhou, Tao Wu, and Hongjin Zhu suggests improving the Single Shot Detector (SSD) model to detect targets in vehicles. The goal of the study is to enhance the SSD model in order to increase the accuracy of vehicle detection. The author's most likely focus on trying to resolve the shortcomings of conventional computer vision inspection technology in terms of vehicle detection. This paper highlights that the method uses the ResNet101 neural network layer instead of the VGG16 feature extraction neural network layer, because the ResNet101 neural network layer has stronger feature extraction capabilities.

This research paper [17] titled "Enhanced Single Shot Multibox Detector for Pedestrian Detection by Yongren. This paper proposes an improved SSD pedestrian detection algorithm, which solves the problems of SSD detectors' poor detections on small-scale targets and inability to handle dense pedestrians.

The author utilized dense connections to the first three feature maps and an additional shallow feature map to add more information, allowing the algorithm to identify more small-scale pedestrian targets. The study also shows that the model guarantees detection efficiency in addition to achieving high detection accuracy on PRW, Caltech, and VOC datasets.

The research paper [5] titled "Improved Single Shot Multibox Detector Target Detection Method Based on Deep Feature Fusion" by Dongxu Bai and Ying Sun focuses on enhancing the Single Shot Multibox Detector (SSD) for target detection through deep feature fusion. In order to increase target detection efficiency and accuracy, this study proposes a way that integrates deep feature fusion techniques into the SSD framework. The proposed method seeks to improve the model's detection capability of targets with different scales and complexities by combining features at different levels of abstraction. The experimental findings demonstrate that the enhanced SSD algorithm increases the target's detection accuracy and rate, with the effect being more visible for relatively small-scale targets.

The research paper [6] titled "Accurate and Fast Single Shot Multibox Detector" by Lie Guo and Dongxing Wang which includes context comprehensive enhancement (CCE) module and feature enhancement module (FEM). In this study, to solve the difficulties of low detection accuracy of small targets and difficult balance between speed and accuracy, the paper introduced CCE module and FEM on the basis of SSD. Experiments on Pascal VOC and MS COCO data sets and comparison with other methods, the results show that this method can achieve almost state-of-the-art performance while maintaining real-time speed. The author's network can achieve 81.2 mean average precision on the PASCAL VOC 2007 test with an input size of $320 \times 320$ on a single Nvidia 2080Ti GPU.

The research paper [11] "Fast single shot multibox detector and its application on vehicle counting system" by Zhengdao Zhang and Li Peng focuses on the application of a Fast Single Shot MultiBox Detector (SSD) for vehicle detection and counting systems. The study presents an efficient method based on SSD to address the challenge of constructing a system for vehicle detection and counting. The proposed strategy involves using convolutional neural networks (CNNs) and a KLT tracker for vehicle detection, tracking, and counting in fixed camera scenes. The research highlights the advantages of SSD in achieving real-time vehicle detection, improved accuracy, low computational complexity, and effectiveness in handling various traffic scenarios. Fast-SSD achieves the accuracy of 76.7 mAP on the PASCAL visual object class's 2007 test set. The network can be implemented at the speed of 20.8 FPS based on the GTX650Ti. The number of vehicles is calculated when the center of the vehicle passes the virtual loop detector. Results show that the vehicle detection accuracy achieves 99.3% and the classification accuracy is 98.9%.

The research paper [10] "Object Detection Based on the Improved Single Shot MultiBox Detector" by Songmin Jia et al aims to address the issue of poor performance of deep learning algorithms in detecting small objects. To achieve this, the authors propose an improved Single Shot MultiBox Detector (SSD) method based on feature fusion. This approach enhances the detection capabilities of the SSD by combining features from different layers of the network, which helps in detecting objects of varying sizes. The method is designed to improve the accuracy of object detection, particularly for small objects, by leveraging the strengths of different features from different layers. This research contributes to the development of more robust and accurate object detection systems, especially in scenarios where small objects are present. Experimental result indicates: when the confidence is set to 0.5, the mAP of the SSD method based on feature fusion is 78.04%, which is 0.8% higher than the classical SSD algorithm and 4.8% higher than the Faster RCNN algorithm. The proposed algorithm improves the ability of small objects, and verifies the effectiveness of the proposed algorithm.

The research paper specifically [18] titled "Underwater Object Detection Based on Improved Single Shot MultiBox Detector" by Zhongyun Jiang. This paper aims to enhance the network's small object detection capability by incorporating a shallow object detection layer into the original SSD model. Enhance the feature information of the objects in the underwater images by processing the original images using the Multi-Scale Retinex with Color Restoration (MSRCR) algorithm. Results of the experiments show that the algorithm suggested in this paper performs better in terms of detection than other algorithms, such as YOLO v3 and the original SSD. This is a significant advancement towards the realization of underwater object detection. While the improved SSD algorithm has a faster detection speed than the original SSD algorithm, it still improves the mean average precision by almost 5%. It also has detection accuracy comparable to that of Faster R-CNN.

The research paper [9] "Object Detection in Real Time Based on Improved Single Shot MultiBox Detector Algorithm" by Ashwani Kumar, Zuopeng Justin Zhang, and Hongbo Lyu focuses on the development of a real-time object detection system using an improved version of the Single Shot MultiBox Detector (SSD) algorithm. The authors aim to enhance the performance of the SSD by improving its ability to detect objects in real-time on any device. The study proposes an improved SSD method that addresses the limitations of the original SSD in terms of detection accuracy and speed. The authors introduce a new approach that combines the strengths of different layers in the network to improve the detection of objects of varying sizes.

This method is designed to enhance the accuracy of object detection, particularly for small objects, by utilizing the strengths of different features from different layers. The improved SSD method can be used in scenarios where real-time object detection is crucial, such as in traffic monitoring systems. The accuracy in detecting the objects is checked by different parameters such as loss function, frames per second (FPS), mean average precision (mAP), and aspect ratio. Experimental result: proposed improved SSD algorithm has high accuracy. The accuracy of the proposed model is more than 79.8%. The training time for this model is about 5–6 h.

The research paper [19] "Classification and Localization of Defects using Single-Shot MultiBox Detector" by Manjeet Kaur, Krishan Kumar Chauhan, Isibor Kennedy Ihianle, Kayode Owa, Naveen Aggarwal, Renu Vig, and Garima Joshi focuses on utilizing the Single-Shot MultiBox Detector (SSD) for defect classification and localization. The proposed methodology involves the integration of the SSD algorithm with feature fusion techniques to enhance the detection of defects, ensuring both classification and precise localization. By combining the strengths of the SSD framework with feature fusion, the research aims to improve the accuracy and efficiency of defect detection systems. The proposed method greatly improves the accuracy for fault detection to 98.04%.

The research paper [7] titled "An Approach on Image Processing of Deep Learning Based on Improved SSD" by Liang Jin and Guodong Liu. The average detection speed is 31FPS, and the mean average precision (mAP) is 94.41%, according to the experimental results based on the Synthetic Aperture Radar ship detection dataset (SSDD). The addition of a feature pyramid network to SSD allows Conv4_3 and Conv7 two large-scale feature maps in charge of small object detection to receive context information. The paper also emphasizes that a SE module is added, allowing the model to perform dynamic channel-wise feature recalibration, in order to further enhance the ability of feature extraction, raise the significant channel-wise feature, as well as reduce insignificant channel-wise feature.

### *A. Summarized*

According the literature the model performance can be improved. This performance can be improved by combining the effective features like feature pyramid network (FPN), channel-wise feature, feature fusion. The performance of model also can be improved by adding the additional layers like CONV4_3, CONV_7, and CONV_6 in the model.

## III. TAXONOMY

### *A. Single Shot Multibox Detector (SSD)*

The original SSD model is based on a feed-forward convolutional network, which produces a fixed-size collection of bounding boxes and classifies the bounding box, followed by a non-maximum suppression step to produce the final detection [11].The structure of the SSD, which is divided into three parts: the main layer based on VGG16(very deep convolutional networks for large-scale image recognition), the feature extraction layer, and the classification layer. The VGG16 network structure in the main layer was optimized. First, the sixth and seventh convolution layers, Conv6 and Conv7, are used to replace FC6 and FC7 (fully connected layers) in the original structure of VGG16 to avoid the interference of the full connection layer with the detection object features and position information. Second, the feature mapping relationships of conv4_3, conv7_2, conv8_2, conv9_2, conv10_2, and conv11 are combined to form a multiscale feature extraction layer in the SSD. Finally, a 3×3 convolution is used to calculate the output feature graphs of the detection layer one by one to obtain the confidence [20] . Fig. 1, which shows the overall structure of conventional SSD, each layer in the feature pyramid1 is used independently as an input to the classifier network [2].

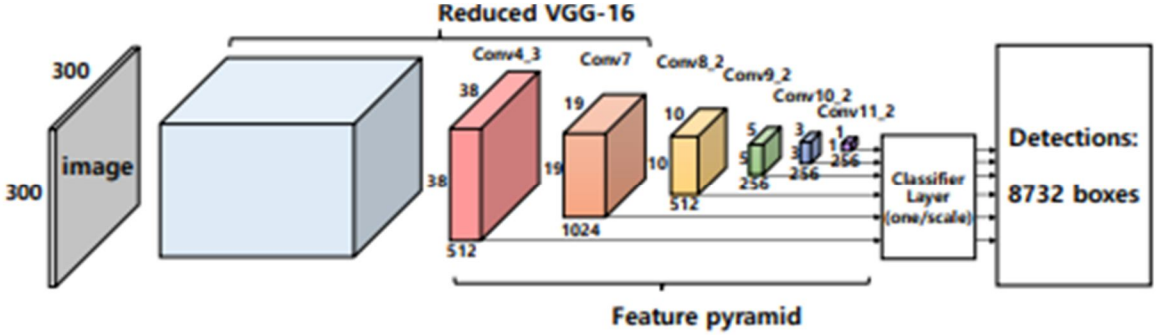



Figure 1: SSD Architecture [2]

The original SSD algorithm uses multi-scale feature maps for prediction, which improves the detection accuracy to a certain extent, but it does not make full use of the information of the shallow feature maps, and small objects in the image may be missed [18]. Therefore, in order to improve the capability of detecting small object of the original SSD network, we enhances the SSD model.

### *B. You Only Look Once (YOLOv7)*

The authors of YOLOv7 are Chien-Yao Wang et al. One of the improvements of YOLOv7 is that the activation function is changed from LeakyReLU to Swish. Other basic modules are optimized by using the residual design idea for reference, but the basic architecture of the network has not changed much and still includes three parts: backbone, neck, and head. Backbone: Darknet, the basic backbone network of the YOLO algorithm, was built by Joseph Redmon. Other versions of the YOLO algorithm are optimized on its architecture. The backbone network of YOLOv7 includes the CBS, E-ELAN, MP, and SPPCSPC modules. CBS, as the most basic module, is integrated into other modules. Feature Fusion Zone The feature fusion layer of the network is to enable the network to better learn the features extracted from the backbone network. The features of different granularities are learned separately and merged in a centralized way so as to learn as many image features as possible. Detection Head: The YOLOv7 algorithm follows the advantages of previous algorithms and retains three detection heads, which are used to detect, confidence, and predicted frame coordinates of the target object. The detection heads output three feature scales: 20 × 20, 40 × 40, and 80 × 80. The target scales detected by the three scales respectively correspond to a large target, a medium target, and a small target [21].

### *C. You Only Look Once (YOLOv8)*

Yolov8 is the latest iteration of the YOLO series detection models, which is a high-performance object detection model. The Yolov8 model architecture is similar to Yolov5 and consists of four parts: Input, Backbone, Neck, and Head. The Input part serves as the input stage for the object detection model. It includes an image preprocessing phase, where the input images are adaptively scaled to match the network's input size and normalized. Additionally, Yolov8 utilizes various image augmentations such as Mosaic, Mix-up, random perspective, and HSV augmentations to enhance the model's adaptability. The Backbone part is the main network architecture responsible for feature extraction from the target images. Yolov8 employs an improved CSPDarknet53 structure, comprising convolution modules, C2F modules, and SPPF modules. The Neck part acts as the connector in the network and is primarily used to obtain and fuse feature information. Yolov8 uses the FPN-PAN structure, where the FPN structure is a top down approach that obtains prediction feature maps through up sampling and fusing bottom-level feature information. The PAN part employs a bottom-up structure to complement the FPN feature maps, forming a feature pyramid network. The Head part has undergone significant changes compared to Yolov5. It adopts the popular Decoupled-Head structure, which separates the classification and detection heads. Moreover, Yolov8 shifts from the traditional Anchor Based to Anchor-Free approach [22].

## IV. THE ENHANCED SSD MODEL

SSD is an object detection model that predicts object bounding boxes and class probabilities in a single forward pass. It uses multiple feature maps at different scales to cater to various object sizes and aspect ratios.

### *A. Key Architectural Features to Integrate*

#### *1) Backbone Network*

- *SSD Original Backbone:* Typically MobileNetV2.
- *Improved Backbone:* Adopt the backbone networks used in YOLOv7 and YOLOv8, such as a modified CSP (Cross Stage Partial) network, which is known for its balance between speed and accuracy.

#### *2) Feature Pyramid Network (FPN) SSD Mechanism*

SSD uses multiple feature maps for detecting objects at various scales like CNN and RCNN.

- *Enhanced FPN:* Utilize the more sophisticated FPN designs from YOLOv7 and YOLOv8 that integrate better context and fine details across scales, potentially using bi-directional feature pyramids.

#### *3) Activation Functions SSD Activations*

Typically ReLU, LeakyReLU or a similar function.

- *Advanced Activations:* Implement Mish or Swish activation functions as used in YOLOv7 and YOLOv8, which have shown to improve gradient flow and model performance.

4) *Additional Layers for Enhanced Detection*

- *Context and Attention Layers:* Adding context-aware and attention-driven layers can further improve the model's ability to focus on relevant features and suppress background noise.
- *Action Steps:* Integrate attention mechanisms such as Squeeze-and-Excitation (SE) blocks or CBAM (Convolutional Block Attention Module) within the detection layers to enhance the model's focus on informative features. Add context modules that aggregate information across a broader range of inputs, similar to the mechanisms used in the PANet architecture seen in newer YOLO versions.

5) *Loss Function Modifications*

- *SSD Loss:* Combination of localization loss (smooth L1) and confidence loss (cross-entropy).
- *YOLO Loss Enhancements:* Integrate the CIOU or DIOU loss functions from YOLOv7 and YOLOv8 for better bounding box regression and to penalize poorly formed predictions more effectively.

6) *Data Augmentation Techniques*

- *SSD Techniques:* Basic augmentations like flips, color changes, and random crops.
- *Advanced Techniques:* Incorporate mosaic augmentation and mix-up strategies from YOLOv7 and YOLOv8, which create more varied and challenging training examples that help in generalizing the model better.
- *Initial Training:* Use a similar strategy to YOLOv7 or YOLOv8 by starting with pre-trained weights on a large dataset like ImageNet and then fine-tuning on COCO object detection datasets.
- *Optimizers:* Experiment with advanced optimizers such as AdamW or SGDP which are often used in the training of YOLO models to stabilize and speed up convergence.

7) *Inference Optimizations*

Post-processing: Integrate NMS (Non-Max Suppression) techniques from YOLOv7 and YOLOv8, which might include using a class-specific NMS or integrating a more efficient version of the DIoU-NMS.

8) *Evaluation and Testing*

- *Metrics:* Use standard object detection metrics such as mAP (mean Average Precision) for various IOU (Intersection over Union) thresholds to evaluate the enhanced SSD model.

9) *Coordination of Detection Boxes and Ground Truths*

- *Detection Boxes in SSD: Box Representation:* SSD predicts bounding boxes using the (x, y) coordinates of the box's centre, width, and height, typically encoded as offsets from default boxes at various aspect ratios and scales.

*B. Output Format*

The model predicts these bounding box parameters along with class probabilities for each default box across multiple feature maps at different scales.

1) *Handling Ground Truths:* Ground Truths in Training: During training, the SSD model is trained to match predicted boxes with ground truths. Ground truths consist of annotated bounding boxes (x1, y1, x2, y2) for each object instance in the training dataset.
2) *Matching Strategy:* SSD matches predicted boxes with ground truths based on IOU (Intersection over Union) criteria, where a predicted box with high IOU with a ground truth box is assigned the corresponding object label and optimized for localization accuracy.
3) *Training Objectives: Localization Loss:* SSD minimizes localization error by optimizing the predicted box parameters (centre coordinates, width, height) using smooth L1 loss.
4) *Confidence Loss:* The model also optimizes class probabilities associated with each predicted box using cross-entropy loss, ensuring accurate object classification.

## V. COMPARSION

The performance of YOLOv8 is better than the performance of YOLOv7 model. The performance of YOLOv7 model is better than the performance of SSD according to the this paper [1].The results of YOLOv8 model > The results of YOLOv7 model >The results of SSD model. To improve the performance of SSD we have proposed the Enhanced Single Shot Multibox Detector (ESSD).By implemented the ESSD model we achieved the experimented results: The results of Enhanced Single Shot Multi Box Detector (ESSD) > the results of Single Shot Multibox Detector.

## VI. ENVIRONMENT

The analysis is written on Google Colab using python language. Python language version used in the project is python 3.12.13. Colab notebook is an application which make us to access the run time GPU. Experiment is conducted using a 64 bit windows 10 pro system, the processor is Intel(R) Core(TM) i3-6006U CPU @ 2.00GHz 2.00 GHz. The dataset is used in our project implementation is COCO dataset from website: https://www.cocodataset.org/ accessed 16/04/2024 at 01:12 PM. The COCO data set containing 100 images and containing 90 classes. The size of the dataset is 7MB.For implementing object detection .Tenser flow is must and some other packages such as Numpy, Mat lab. Keras, pathlib etc.The theoretical approach used to carry out the study included analysis of many research papers based upon object detection algorithms mainly single shot multibox detector (SSD). The second approach used is the empirical approach. This approach evaluates and validates the results given by theoretical approach. For empirical approach experiment was carried out. The experiment is carried out on Google Colab by COCO dataset.

1) *Dataset:* COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features: Object segmentation, Recognition in context and 80 object categories [23]. COCO dataset is accessed by this website https://cocodataset.org/ accessed 16/04/2024 at 1:12 PM.
2) *Google Colab:* Google Colaboratory, or Colab, is an as-a-service version of Jupyter Notebook that enables you to write and execute Python code through your browser. Jupyter Notebook is a free, open source creation from the Jupyter Project. A Jupyter notebook is like an interactive laboratory notebook that includes not just notes and data, but also code that can manipulate the data. The code can be executed within the notebook, which, in turn, can capture the code output. Applications such as Matlab and Mathematica pioneered this model, but unlike those applications, Jupyter is a browser-based web application [24].
3) *Metrics Recall Time:* Recall measures the model's ability to capture all relevant instances of a class, indicating its sensitivity to detecting objects, especially those that are less prominent or challenging.
4) *Precision Time:* Precision reflects the model's ability to minimize false positives, ensuring that the detected objects are relevant and accurately classified.
5) *Accuracy:* This metric showcases the overall correctness of the model's predictions, providing a high-level assessment of its performance.
6) *Mean Average Precision:* MAP is a crucial metric in object detection tasks, quantifying the model's ability to precisely localize and classify objects across various conditions and scales.

## VII. RESULT ANALYSIS

These 100 images were validated on Google Colab by Enhanced SSD model. Recall Time, Precision Time Average Accuracy and Mean Average Precision, these are some metrics which is used to test the performance for Enhanced SSD model. To show 100 images as depicted in fig. (A), fig. (B) and fig. (C).

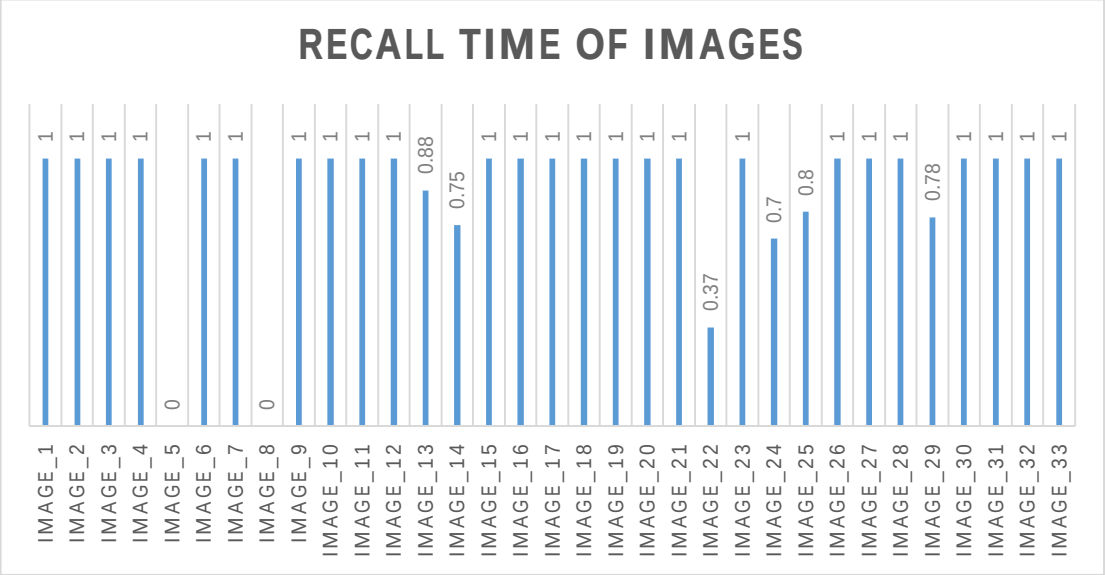



Fig 2 (a) i.e. “Recall Time of Images” by enhanced SSD model” shows the recall time of images. Recall measures how well our model can find true positives (TP) out of all predictions (TP+FN) [25].

The figure shows the images from 1 to 33. Fig 2. (a) Shows the recall time of image_13 by Enhanced SSD Model. The recall time of image_13 by Enhanced SSD Model is 0.88. Fig 2. (a) Shows the recall time of image_15 by Enhanced SSD Model. The recall time of image_15 by Enhanced SSD Model is 1. Fig 2. (a) Shows the recall time of image_20 by Enhanced SSD Model. The recall time of image_20 by Enhanced SSD Model is 1. Fig 2. (a) Shows the recall time of image_28 by Enhanced SSD Model. The recall time of image_28 by Enhanced SSD Model is 1.

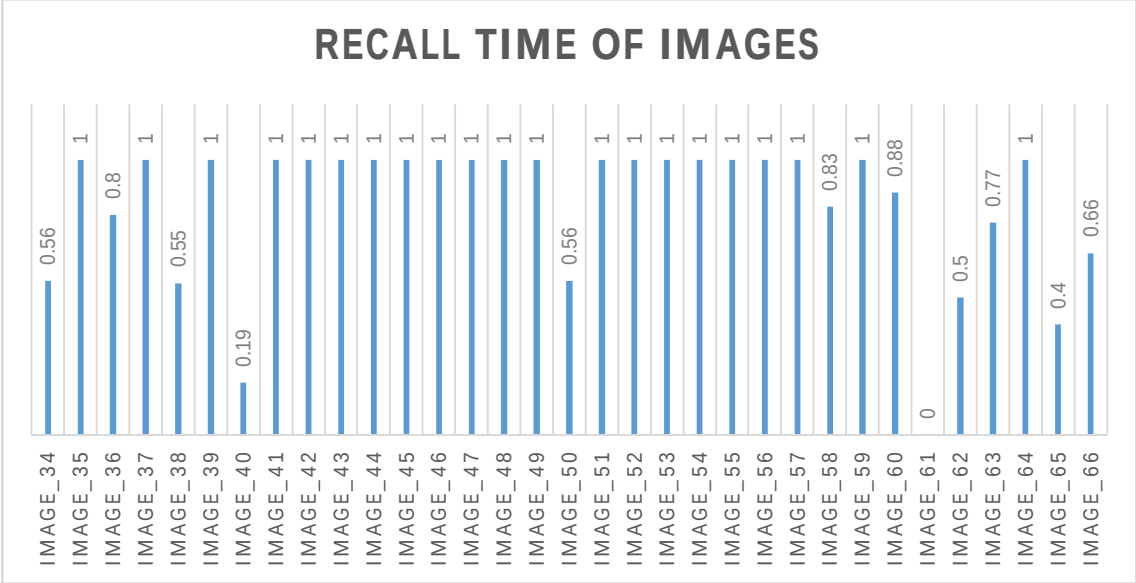



Fig 2 (b) i.e. "Recall Time of Images" by enhanced SSD model" shows the recall time of images. Recall measures how well our model can find true positives (TP) out of all predictions (TP+FN) [25].

The figure shows the images from 34 to 66.Fig 2. (b) Shows the recall time of image_37 by Enhanced SSD Model. The recall time of image_37 by Enhanced SSD Model is 1. Fig 2. (b) Shows the recall time of image_47 by Enhanced SSD Model. The recall time of image_47 by Enhanced SSD Model is 1.Fig 2. (b) Shows the recall time of image_60 by Enhanced SSD Model. The recall time of image_60 by Enhanced SSD Model is 0.88. Fig 2. (b) Shows the recall time of image_66 by Enhanced SSD Model. The recall time of image_66 by Enhanced SSD Model is 0.66.

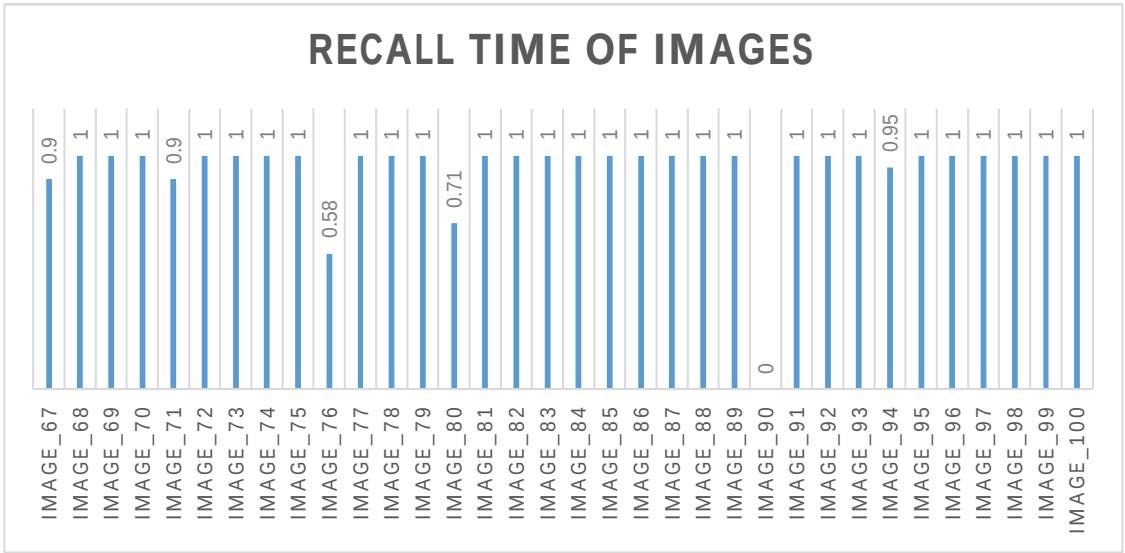



Fig 2(c): i.e. "Image's recall time based on enhanced SSD model" shows the recall time of images Enhanced SSD Model. Recall measures how well our model can find true positives (TP) out of all predictions (TP+FN) [14].

The figure shows the images from 67 to 100. Fig 2. (c) Shows the recall time of image_70 by Enhanced SSD Model. The recall time of image_70 by Enhanced SSD Model is 1.Fig 2. (c) Shows the recall time of image_80 by Enhanced SSD Model. The recall time of image_80 by Enhanced SSD Model is 0.71.Fig 2. (c) Shows the recall time of image_86 by Enhanced SSD Model. The recall time of image_86 by Enhanced SSD Model is 1.Fig 2. (c) Shows the recall time of image_96 by Enhanced SSD Model. The recall time of image_96 by Enhanced SSD Model is 1.

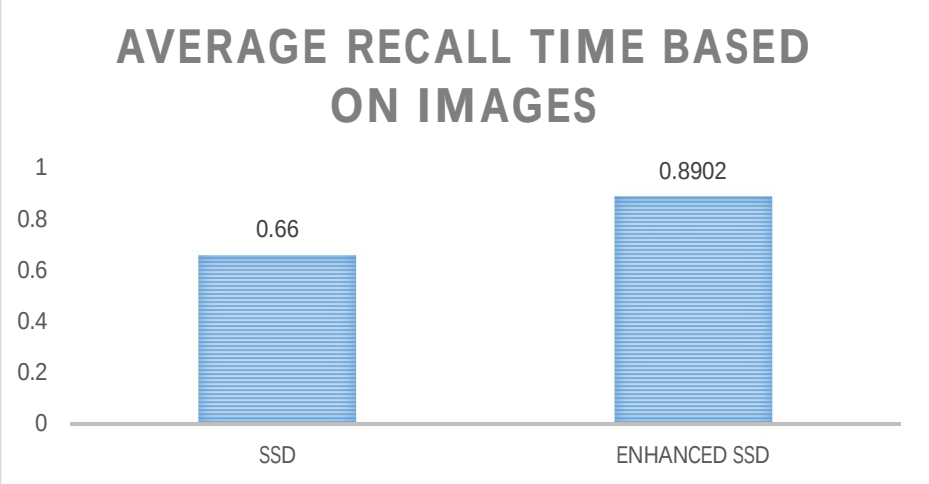



Fig 3. i.e. “Average Recall Time based on Images”.

Results of Average Recall Time = Average Recall Time of Enhanced SSD Model > Average Recall Time of SSD.

These 100 images were validated on Google Colab by Enhanced SSD model. Recall Time, Precision Time Average Accuracy and Mean Average Precision, these are some metrics which is used to test the performance for Enhanced SSD model. To show 100 images as depicted in fig. (A), fig. (B) and fig. (C).

Precision is the ratio of the number of true positives to the total number of positive predictions [26] .

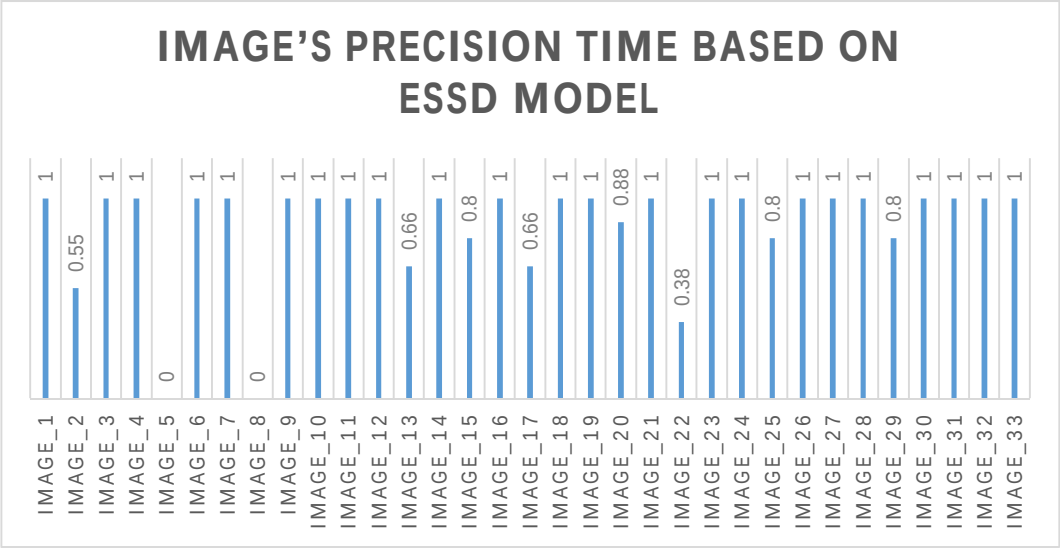



Fig 4 (a): i.e. “Image’s Precision time based on enhanced SSD model” shows the precision time of images by Enhanced SSD Model. The figure shows the images from 1 to 33.

Fig 4. (a) Shows the precision time of image_2 by Enhanced SSD Model. The precision time of image_2 by Enhanced SSD Model is 0.58.Fig 4. (b) Shows the precision time of image_13 by Enhanced SSD Model. The precision time of image_13 by Enhanced SSD Model is 0.66.Fig 4. (b) Shows the precision time of image_30 by Enhanced SSD Model. The precision time of image_30 by Enhanced SSD Model is 1.Fig 4. (b) Shows the precision time of image_33 by Enhanced SSD Model. The precision time of image_33 by Enhanced SSD Model is 1.

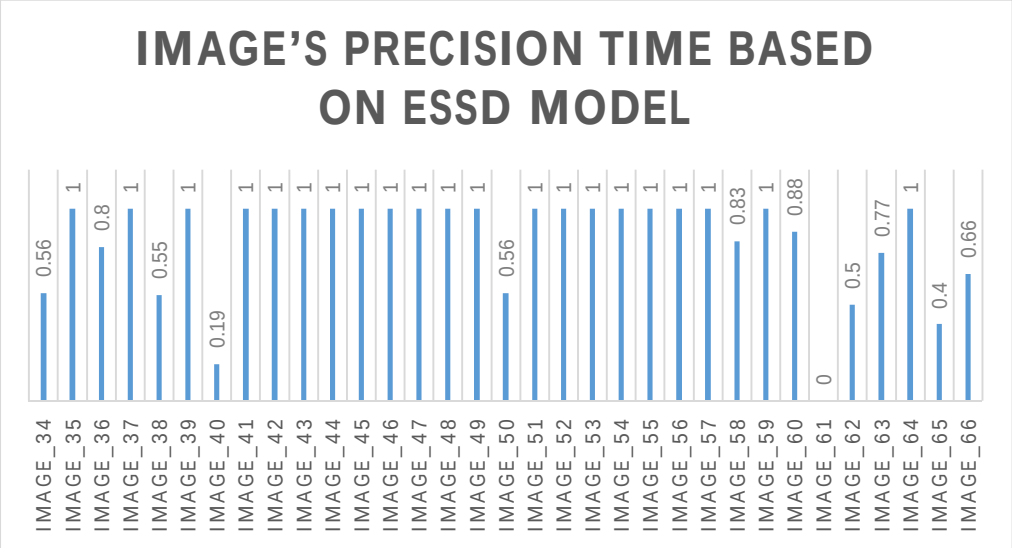



Fig 4 (b): i.e. "Image's Precision time based on enhanced SSD model" shows the precision time of images by Enhanced SSD Model. The figure shows the images from 34 to 66.

Fig 4. (b) Shows the precision time of image_34 by Enhanced SSD Model. The precision time of image_34 by Enhanced SSD Model is 0.56.Fig 4. (b) Shows the precision time of image_39 by Enhanced SSD Model. The precision time of image_39 by Enhanced SSD Model is 1.Fig 4. (b) Shows the precision time of image_48 by Enhanced SSD Model. The precision time of image_48 by Enhanced SSD Model is 1.Fig 4. (b) Shows the precision time of image_51 by Enhanced SSD Model. The precision time of image_51 by Enhanced SSD Model is 1.

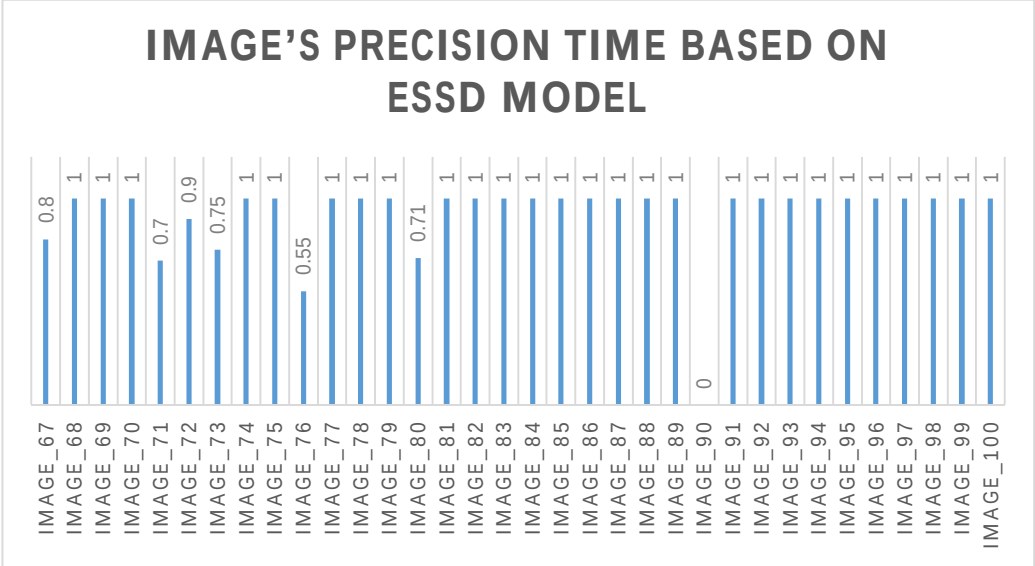



Fig 4 (c): i.e. "Image's Precision time based on enhanced SSD model" shows the precision time of images by Enhanced SSD Model. The figure shows the images from 67 to 100.

Fig 4. (c) Shows the precision time of image_67 by Enhanced SSD Model. The precision time of image_67 by Enhanced SSD Model is 0.8.Fig 4. (c) Shows the precision time of image_79 by Enhanced SSD Model. The precision time of image_79 by Enhanced SSD Model is 1.Fig 4. (c) Shows the precision time of image_88 by Enhanced SSD Model. The precision time of image_88 by Enhanced SSD Model is 1.Fig 4. (c) Shows the precision time of image_91 by Enhanced SSD Model. The precision time of image_91 by Enhanced SSD Model is 1.

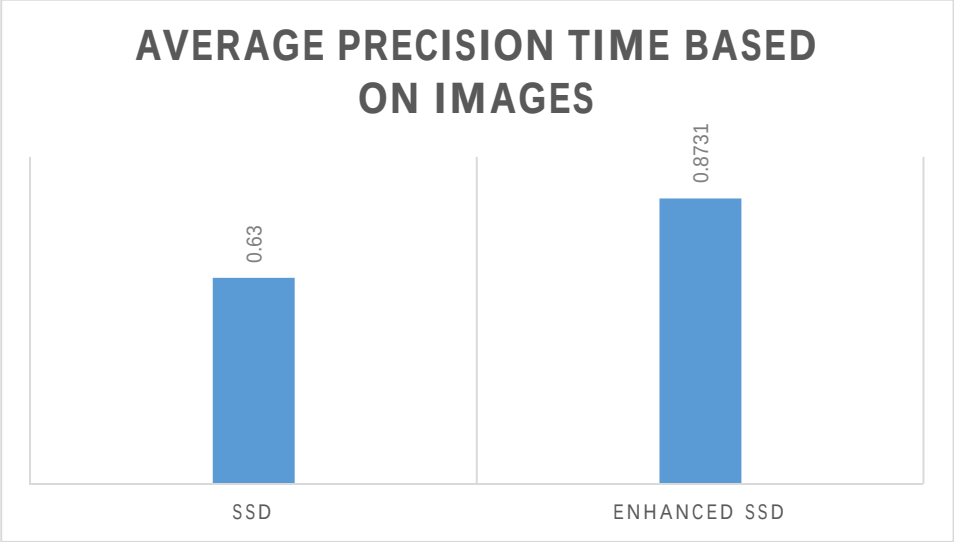



Fig 5. i.e. "Average Precision Time based on Images".

Results of Average Recall Time = Average Precision Time of Enhanced SSD Model (ESSD) >Average Precision Time of SSD.

Fig 6 (a): i.e. "Image's accuracy based on enhanced SSD model" shows the accuracy of images by Enhanced SSD Model. The figure shows the images from 1 to 33.

Fig 6. (a) Shows the accuracy of image_3 by Enhanced SSD Model. The accuracy of image_3 by Enhanced SSD Model is 100%.Fig 6. (a) Shows the accuracy of image_6 by Enhanced SSD Model. The accuracy of image_6 by Enhanced SSD Model is 100%.Fig 6. (a) Shows the accuracy of image_31 by Enhanced SSD Model. The accuracy s of image_31 by Enhanced SSD Model is 100%.Fig 6. (a) Shows the accuracy of image_33 by Enhanced SSD Model. The accuracy of image_33 by Enhanced SSD Model is 100%.

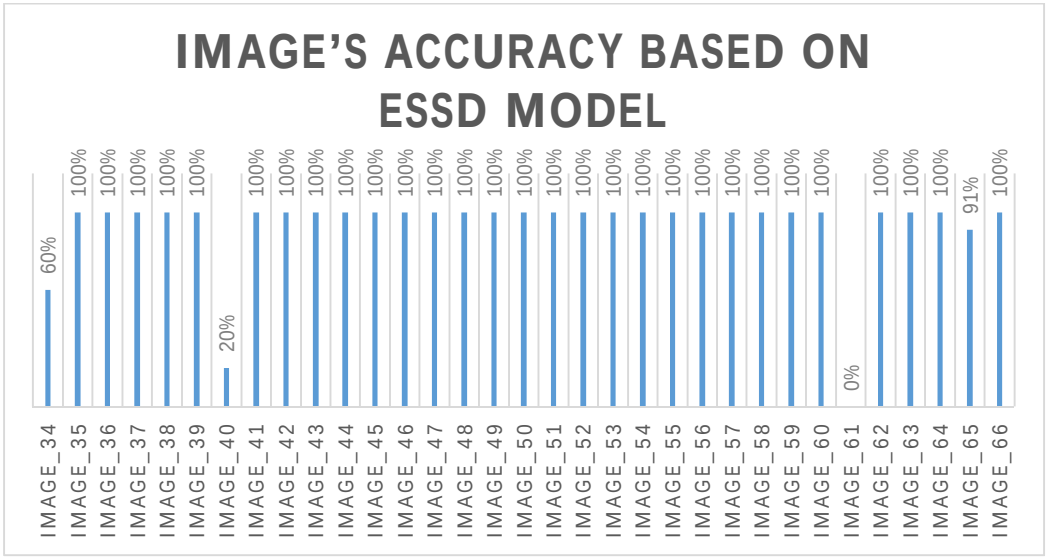



Fig 6 (b): i.e. "Image's accuracy based on enhanced SSD model" shows the accuracy of images by Enhanced SSD Model. The figure shows the images from 34 to 66.

Fig 6. (b) Shows the accuracy of image_35 by Enhanced SSD Model. The accuracy of image_35 by Enhanced SSD Model is 100%.Fig 6. (b) Shows the accuracy of image_40 by Enhanced SSD Model. The accuracy of image_40 by Enhanced SSD Model is 20%.Fig 6. (b) Shows the accuracy of image_58 by Enhanced SSD Model. The accuracy of image_58 by Enhanced SSD Model is 100%.Fig 6. (b) Shows the accuracy of image_66 by Enhanced SSD Model. The accuracy of image_66 by Enhanced SSD Model is 100%.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue V May 2024- Available at www.ijraset.com*

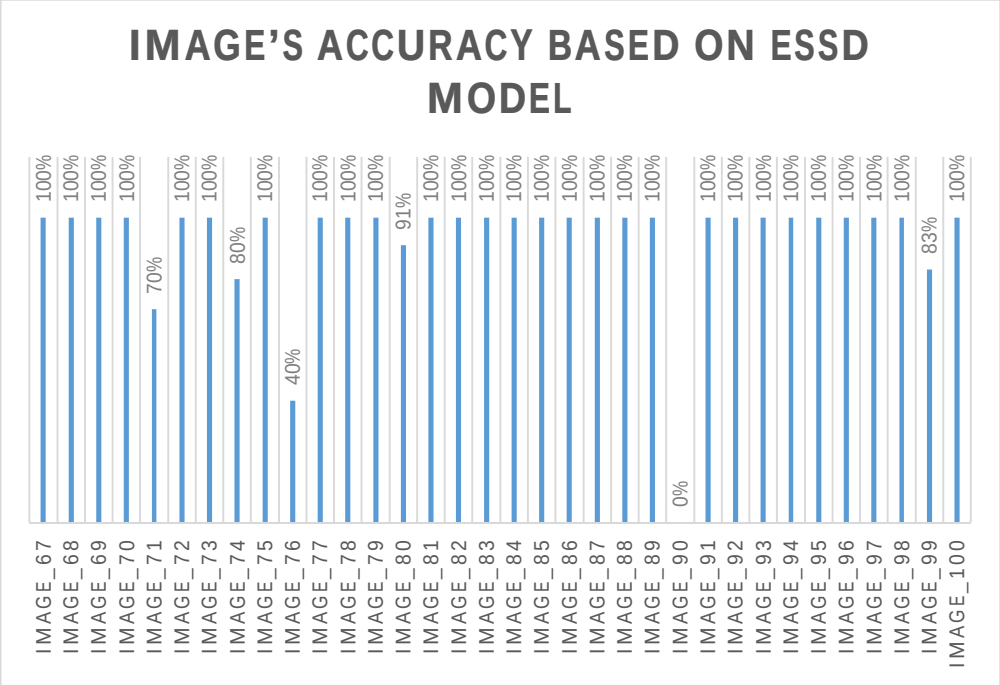



Fig 6 (c): i.e. "Image's accuracy based on enhanced SSD model" shows the accuracy of images by Enhanced SSD Model. The figure shows the images from 67 to 100.

Fig 6. (c) Shows the accuracy of image_68 by Enhanced SSD Model. The accuracy of image_68 by Enhanced SSD Model is 100%.Fig 6. (c) Shows the accuracy of image_70 by Enhanced SSD Model. The accuracy of image_70 by Enhanced SSD Model is 100%.Fig 6. (c) Shows the accuracy of image_78 by Enhanced SSD Model. The accuracy of image_78 by Enhanced SSD Model is 100%.Fig 6. (c) Shows the accuracy of image_100 by Enhanced SSD Model. The accuracy of image_100 by Enhanced SSD Model is 100%.

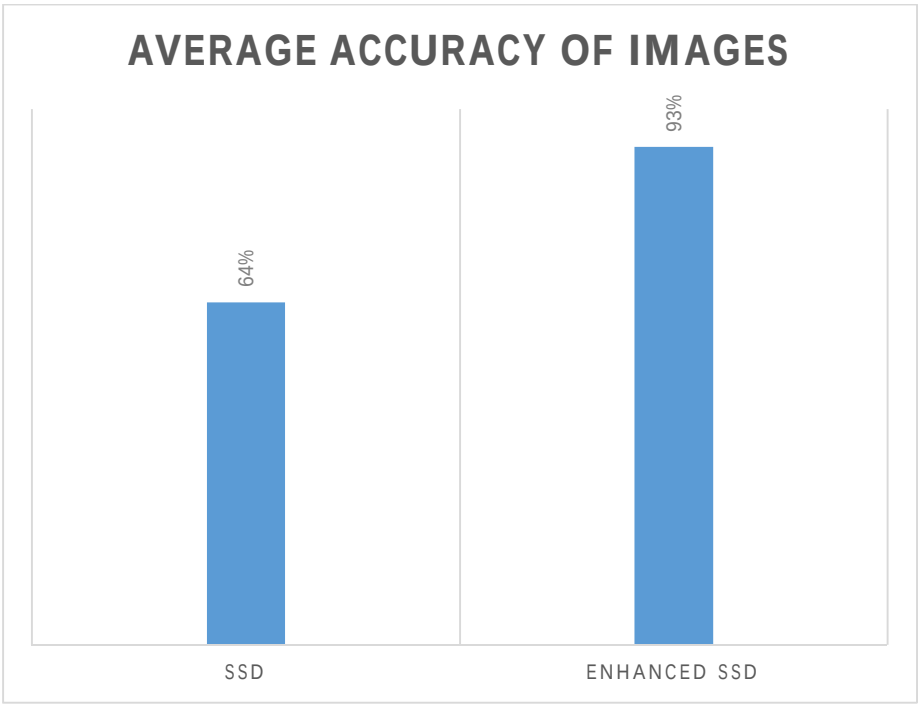



Fig 7: i.e. " Average Accuracy of Images

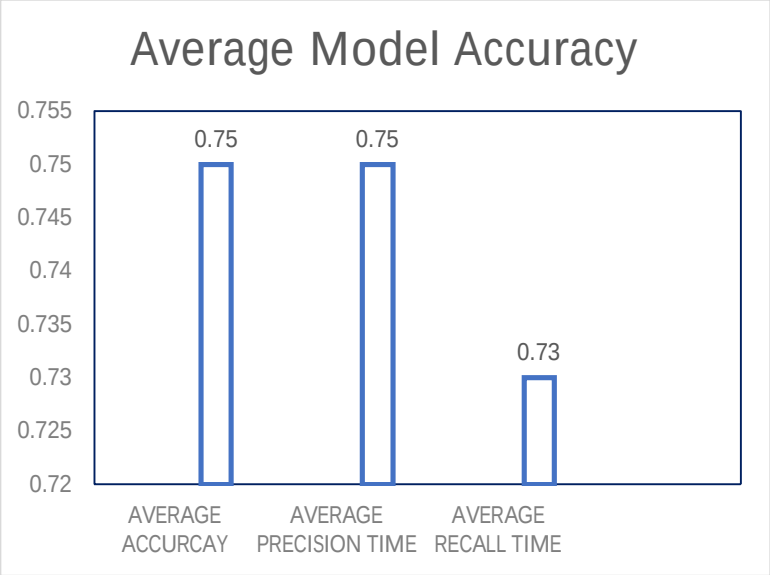



Fig 8 i.e. "Average Model Accuracy" which demonstrates the final accuracy, average precision time and average recall time.

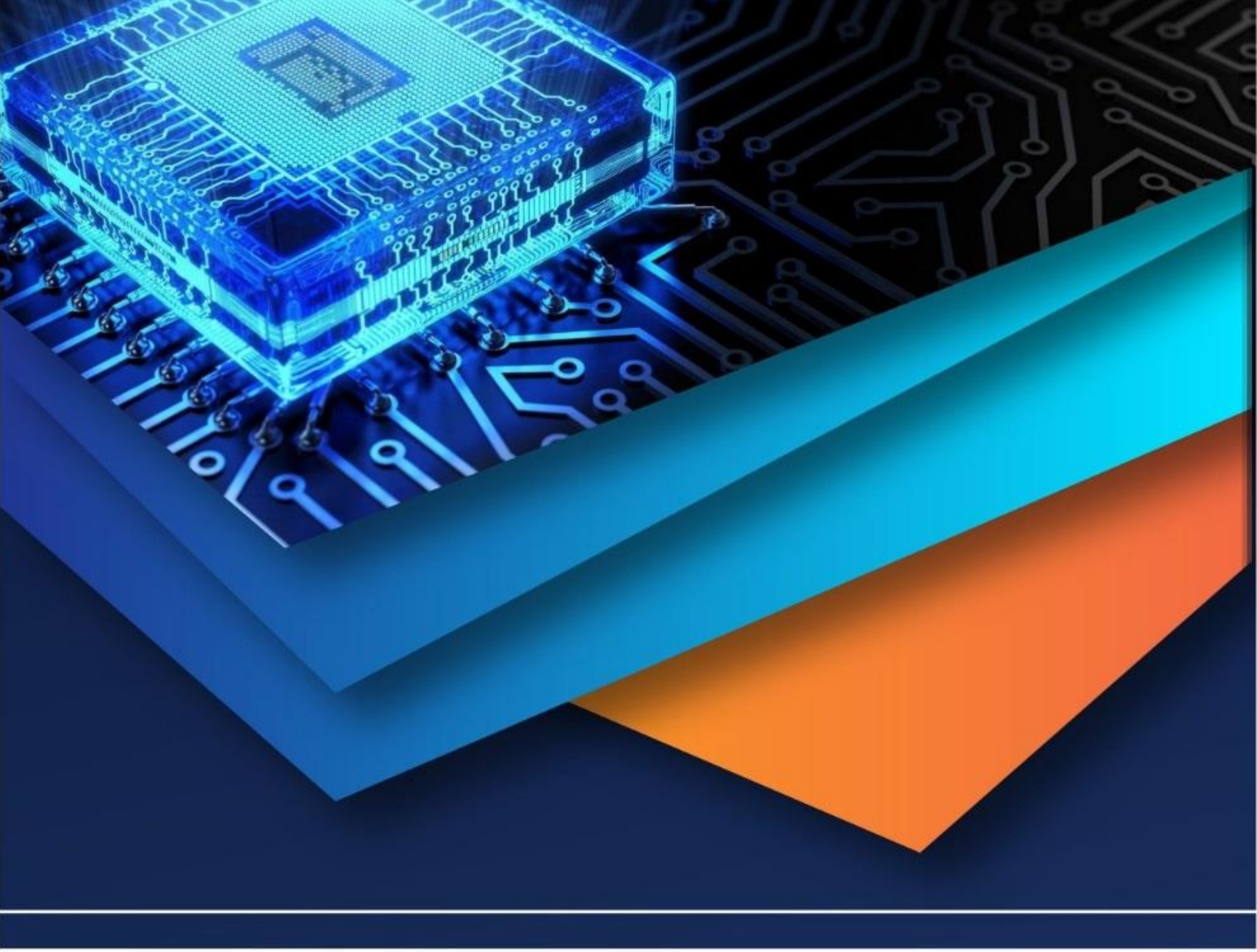
## VIII. CONCLUSION & FUTURE SCOPE

This research paper briefly highlights the significance of object detection and how it is important for today's era. In this paper we propose an enhanced SSD model for more effective object detection. Moreover the Enhanced SSD model competes favourably with the other state of the art techniques. Accuracy is improved by adding number of features, improved backbone. In particular, it achieves better performance than the one stage detector YOLOv7, YOLOv8 and being easier to train and there is no need for heavily tune the loss parameters. Average recall time of enhanced SSD model is 0.73. Average precision time of enhanced SSD model is 0.75. Average accuracy of enhanced SSD model is 0.75. Mean average precision of enhanced SSD model is 75.25. A number of things could lead to a very positive future for ESSD: The improvements made to SSD over its predecessor SSD may result in increased accuracy in object detecting jobs. More advancements and breakthroughs in computer vision research should be anticipated, resulting in even greater accuracy. ESSD is ideally suited for implementation on edge devices due to its capacity to attain high performance with constrained computational resources. Intelligent edge applications will be made possible by optimized object detection algorithms such as ESSD, which will become increasingly important as edge computing spreads throughout many industries. The average model accuracy is approx... 75% written studies may be conducted to further enhance it.

## REFERENCES

[1] A. K. Sharma, "Analysis of Object Detection Models," Eng. Technol., vol. 12.
[2] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection." arXiv, May 26, 2017. Accessed: May 02, 2024. [Online]. Available: http://arxiv.org/abs/1705.09587
[3] W. Liu et al., "SSD: Single Shot MultiBox Detector," vol. 9905, 2016, pp. 21–37. Doi: 10.1007/978-3-319-46448-0_2.
[4] W. Lung and Y.-R. Wang, "Applying Deep Learning and Single Shot Detection in Construction Site Image Recognition," Buildings, vol. 13, no. 4, p. 1074, Apr. 2023, doi: 10.3390/buildings13041074.
[5] D. Bai et al., "Improved single shot multibox detector target detection method based on deep feature fusion," Concurr. Comput. Pract. Exp., vol. 34, no. 4, p. e6614, Feb. 2022, doi: 10.1002/cpe.6614.
[6] L. Guo, D. Wang, L. Li, and J. Feng, "Accurate and fast single shot multibox detector," IET Comput. Vis., vol. 14, no. 6, pp. 391–398, Sep. 2020, doi: 10.1049/iet-cvi.2019.0711.
[7] L. Jin and G. Liu, "An Approach on Image Processing of Deep Learning Based on Improved SSD," Symmetry, vol. 13, no. 3, p. 495, Mar. 2021, doi: 10.3390/sym13030495.
[8] Q. Liu, L. Dong, Z. Zeng, W. Zhu, Y. Zhu, and C. Meng, "SSD with multi-scale feature fusion and attention mechanism," Sci. Rep., vol. 13, no. 1, p. 21387, Dec. 2023, doi: 10.1038/s41598-023-41373-1.
[9] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," EURASIP J. Wirel. Commun. Netw., vol. 2020, no. 1, p. 204, Dec. 2020, doi: 10.1186/s13638-020-01826-x.
[10] S. Jia et al., "Object Detection Based on the Improved Single Shot MultiBox Detector," J. Phys. Conf. Ser., vol. 1187, no. 4, p. 042041, Apr. 2019, doi: 10.1088/1742-6596/1187/4/042041.
[11] L. Chen, Z. Zhang, and L. Peng, "Fast single shot multibox detector and its application on vehicle counting system," IET Intell. Transp. Syst., vol. 12, no. 10, pp. 1406–1413, Dec. 2018, doi: 10.1049/iet-its.2018.5005.

[12] W. Li and G. Liu, "A Single-Shot Object Detector with Feature Aggregation and Enhancement," in 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan: IEEE, Sep. 2019, pp. 3910–3914. doi: 10.1109/ICIP.2019.8803543.
[13] J. Yi, P. Wu, and D. N. Metaxas, "ASSD: Attentive Single Shot Multibox Detector." arXiv, Sep. 26, 2019. Accessed: May 02, 2024. [Online]. Available: http://arxiv.org/abs/1909.12456
[14] A. Limaye, M. Mathew, S. Nagori, P. K. Swami, D. Maji, and K. Desappan, "SS3D: Single Shot 3D Object Detector".
[15] Hao Zhang, X. Hong, and L. Zhu, "Detecting Small Objects in Thermal Images Using Single-Shot Detector," Autom. Control Comput. Sci., vol. 55, no. 2, pp. 202–211, Mar. 2021, doi: 10.3103/S0146411621020097.
[16] G. Yu, H. Fan, H. Zhou, T. Wu, and H. Zhu, "Vehicle Target Detection Method Based on Improved SSD Model," J. Artif. Intell. vol. 2, no. 3, pp. 125–135, 2020, doi: 10.32604/jai.2020.010501.
[17] Y. Cheng, C. Chen, and Z. Gan, "Enhanced Single Shot MultiBox Detector for Pedestrian Detection," in Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya China: ACM, Oct. 2019, pp. 1–7. Doi: 10.1145/3331453.3361665.
[18] Z. Jiang and R. Wang, "Underwater Object Detection Based on Improved Single Shot MultiBox Detector," in 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya China: ACM, Dec. 2020, pp. 1–7. Doi: 10.1145/3446132.3446170.
[19] D. I. K. Ihianle et al., "Classification and Localization of Defects Using Single-Shot Multibox Detector." 2023. Doi: 10.2139/ssrn.4673716.
[20] G. Lu, X. He, Q. Wang, F. Shao, J. Wang, and Q. Jiang, "Bridge crack detection based on improved single shot multi-box detector," PLOS ONE, vol. 17, no. 10, p. e0275538, Oct. 2022, doi: 10.1371/journal.pone.0275538.
[21] Y. Zhang, Y. Sun, Z. Wang, and Y. Jiang, "YOLOv7-RAR for Urban Vehicle Detection," Sensors, vol. 23, no. 4, p. 1801, Feb. 2023, doi: 10.3390/s23041801.
[22] S. Li, F. Chen, Z. Sun, Z. Zhu, L. Zhou, and K. Tang, "Research on YOLOv8 Object Detection Algorithm in UAV Scenarios." Mar. 01, 2024. Doi: 10.21203/rs.3.rs-3995816/v1.
[23] "COCO - Common Objects in Context." Accessed: May 02, 2024. [Online]. Available: https://cocodataset.org/#home
[24] "Why and how to use Google Colab | TechTarget," Enterprise AI. Accessed: May 02, 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/tutorial/Why-and-how-to-use-Google-Colab
[25] "What is Precision, Recall, and F1 Score in Object Detection? How are They Calculated?," VisoByte. Accessed: Dec. 29, 2023. [Online]. Available: https://www.visobyte.com/2023/05/precision-recall-and-f1-score-in-object-detection-how-are-they-calculated.html
[26] "How Compute Accuracy for Object Detection works—ArcGIS Pro | Documentation." Accessed: Mar. 13, 2024. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/how-compute-accuracy-for-object-detection-works.htm



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY