Object Detection Using YOLO and EfficientDET Model

Ayush Raj, Computer Science and Engineering, Graphic Era Hill University; Student, Graphic Era Hill University, Dehradun,248002, India.
ayushraj.210111941@gehu.ac.in

Mayank Chouhan, Computer Science and Engineering, Graphic Era Hill University; Student,Graphic Era Hill University, Dehradun,248002, India.
mayankchouhan.210111874@gehu.ac.in

Satvik Vats, SMIEEE Computer Science and Engineering, Graphic Era Hill University; Adjunctprofessor, Graphic Era Deemed to beUniversity, Dehradun,248002, India.
svats@gehu.ac.in

Satvik Vats, SMIEEE Computer Science and Engineering, Graphic Era Hill University; Adjunctprofessor, Graphic Era Deemed to beUniversity, Dehradun,248002, India.
svats@gehu.ac.in

Satvik Vats, SMIEEE Computer Science and Engineering, Graphic Era Hill University; Adjunctprofessor, Graphic Era Deemed to beUniversity, Dehradun,248002, India.
svats@gehu.ac.in

Vikrant Sharma, SMIEEE Computer Science and Engineering, Graphic Era Hill University; Adjunctprofessor, Graphic Era Deemed to beUniversity, Dehradun,248002, India.
vsharma@gehu.ac.in

***Abstract*- In this research paper we presented you the implementation of the YOLO(You Look Only Once) algorithm for the static object i.e images and EfficientDet model for the moving Object. As we know in this developing world object detection has been one of the crucial feature in the field of computer vision for identification and localization of the objects whether it is moving or static. We used YOLOv8 as it provides very high accuracy anda very efficient real time object detectionmaking it very suitable for various applications such as auto driving, surveillance , etc.**

This research paper mainly focuses on achieving object identification and localization efficiently along with detecting moving object with higher accuracy. In addition it also aims to identify the objects in dull images and challenging condition which helps in improving the efficiency of this model as not always capabilities matters but efficiency is also a concern. Overall, in this research paper demonstration the practical implementation of YOLOv8 and EfficientDet for object detection tasks is done ,which providing valuable insights into their performance and capabilities in real-world scenarios.

***Keywords**: YOLOv8, EfficientDET model, Object detection, Computer vision, Accuracy, Python, OpenCV, NumPy, Pandas, Ultralytics, Dataset, Annotated images, Training, Validation, Test sets, Localization, Confidence, Classification loss function, Non-Maximum Suppression (NMS).*

I. INTRODUCTION

The development and advancement in the computer vision and Machine learning (ML) have introduced the development of innovative ideas for human-computer interaction . Object detection is a crucial tsk as computer vision is gaining its popularity throughout world, finds its usage across various industrial applications , as Deep learning algorithms like YOLO (You Only Look Once) and EfficientDet have developed object detection, offering remarkable and a high accuracy and speed. In this project, we have explored the implementation of YOLOv8 and EfficientDet for object detection, focusing on detecting both static objects, such as number plates, and moving objects in video streams.Our project aims to identify and localize the static images with the help of modern object detection techniques, including model training, evaluation, and practical applications.Specially this project aims to perform with high accuracy and capabilities on a large dataset . This algoritms could be effiviently applicable to various scenarios like traffic monitoring ,surveillance, and automated industrial applications. With the advancement in the Deep learning [1], multiple object detection algorithms has been introduced like Faster R-CNN, R-CNN, Mask R-CNN , Fast R-CNN which is also known as Two Stage networks. Talking about One Stage networks like YOLO algorithms , SSD, RetinaNet which efficiently predicts ,localize and provide the probability ans position coordinates.

The objective behind this research is to make efficient model which accurately identifies the object with a high precision.YOLO is also suitable to identify multiple object having different classification .These wide range abilities make it suitable for various advanced applications.The models using YOLO are very easy to deploy using Embedded systems, GPUs, and cloud servers etc,.

This project focuses on Object detection and tracking which can be helpful in various in-need applications like most of the Fast lane is Camera based now which captures the picture of moving object with more than the max limit and a online challan delivered to their registered Number.Like this there are various applications which can be helpful in a

efficient way.

Having several object detection algorithm YOLO being single-stage network algorithm helped I performing detection in single pass improving on several problems faced by two-stage algoritms like CNN , R-CNN etc, through the neural network. This algorithm eliminates the method of having separate region ,feature extraction process stuffs which at the end resulted in the faster detection and localization.

In its core, YOLOv8 consists of several convolutional neural network (CNN) that takes a image as input and produces a set of bounding boxes along with class probabilities for each object detected in the image.As seen in the figure The architecture constitutes of multiple convolutional layers synced in a sequential manner having the final layer responsible for generating predictions.

II. LITERATURE SURVEY

Object detection is a fundamental task in computer vision with wide-ranging applications across various domains, including autonomous driving, surveillance, healthcare, and industrial automation. Over the years, significant advancements in deep learning techniques have led to the development of highly efficient and accurate object detection models. In this literature review, we explore the evolution of object detection algorithms, focusing on recent advancements such as the YOLOv8 and EfficientDet models.

Earlier object detection used to rely on handcraft features and sliding window methods for object localization and classification. At that time these methods used to suffer from scalability issues and not able to perform and solve the complex problems.The birth of deep learning algorithms helped in solving those issues faced earlier by providing end-to-end learning of feature representations from raw pixel data.

One of the influencing works in deep learning-based object detection is the R-CNN family of models introduced by Girshick et al. R-CNN's family of models like Faster R-CNN and Fast R-CNN, introduced the concept of region-based convolutional networks (RCNNs), which helped in significant improvement in detection accuracy and efficiency.In 2014, Girshick et al. introduced the R-CNN models which helped drastically in the object detection with the help of region-based convolutional networks R-CNNs which helped in enhancing the detection accuracy and efficiency ,these works paved the way for advancement in the object detection algorithms.

In 2018, S. Mane et al. proposed an approach for moving object detection and recognition ,and also achieved a high accuracy of 90.88% and self-generated sequences.This paper subsequently focused on moving object detection and tracking on bad background conditions and complex scenerios which reliably helped in detecting fast moving objects .The main problem is that object tracker mainly got distracted by background noise which effect accuracy but this approach helped to detect complex images and tracked it asd well.

In 2019 by B N Krishna sai al. in [2] has proposed the object detection for threatening objects.In this object detection was done with the help of Tensorflow Object detection API and after period of time they used R-CNN algorithm for the implementation of it. The accuracy the got while this project was approx 81% for test set of the data. They classified in only 2 objects which were Knife(11) and Guns(7) only total of 11 images.It doesn't means they stopped at 11 images only but for future research they trained their model at large datset.

In 2020 by Huy Hoang Nguyen et. al. in [3] proposed Computer vision technique for surveillance of video feeds on the devices present on the edge of the network . As we know the edge devices has the limited features a proper consideration need to be taken befor its implementation of algorithm as it could be very costly in terms of computation .The complete research was done on the basis of You Look Only Once(YOLO).The main focus throughout the research was to have a design of network structured in such a way to have good interrelation between accuracy and processing time of around 95% to 96% on INRIA and PENNFUDAN dataset . The L-CNN model was used for the detection of the objects and the performance was analyzed on the basis of precision and recall (average).

In 2019 by Shaji Thorn Blue in [4], proposed about the improvement of the bouding boxes and its accuracy which formed during implementation of YOLOv3 .The dataset which was followed during the complete research was COCO dataset which is ideally used by the beginners in this field. A eRegion-based algorithms like CovNet was used where YOLO has the capability to convert into multibox by usinf softmax class.YOLO use a grid based approach to for the detection process. As we know YOLOv3 takes sequence of image as a input.

.

In 2018 A. Raghunandan in [5] , This paper mainly focused on the face detection considering some of the features like skin's color and shape of the face using a variably used algorithm MATLAB-2017 to localize and detect the object over video streams. It's not neede to tell what object detection means but identifying the objects like people ,trees, animals kingdom or any unwanted object for the society. Several parameters were used in his research like threshold has been set for the detection of the Target object on the video streams. Like Rgb Euclidean parameter were used for detection of different skin colors to increase over the efficiency as accuracy was accurate but not efficient variable skin colors .

In 2020, As ultralytics released the YOLOv5 model there were many concerns regarding the naming as it was not released by the Authors i.e YOLO but once it was released it came up with many drastic changes like smaller size of the model which makes it computable on the smaller devices and maki esy to deploy and helps beginners to train their model with higher accuracy .

In 2020 YOLOv4, developed by Alexey Bochkovskiy [7], In this model multiple new features were integrated to the model like self-adversarial training, cross mini-batch normalization, and mish activation. All of these enhancements along with GPU optimization helped in increasing accuracy , efficiency and speed of the model.

In 2018 L. Yang, L., Wang, L., & Wu, S [8] .The proposed research paper speaks about the object detection within the image on classic principle of detection process that is selection of region , extracting appropriate features and classifying the object to the

correct class.By applying single-stage Network i.e Convolutional Neural Network on the image. We know that the YOLOv2 has the capability of to work as a single regression problem that helps to draw bounding boxes over the image pixes along with classification and probability at a time. The detailed texture and Semantic info. is done by lowlayers filter and high layers filters respectively. This progress in the deep CNN helps to improves the security concers in various needed fields.

As we have described earlier YOLOv8 and EfficientDet model have been applied in different field of domains like

In Autonomous driving, object detection is the most important feature to be done by any organization for enhancing the security concerns. While Yolov8 also helps in real-time object detection of traffic signs, vehicles etc,.

Now Secondly in Surveillance YOLOv8 and EfficientDet are Very useful in increasing the security concerns and

widely used in surveillance systems for detecting and tracking objects of interest in real-time video streams, enhancing security and situational awareness.

Thirdly in Healthcare object detection has a very huge impact as various disease detection like tumor detection ,organ segmentation as yolo helps to analyse medical images efficiently. At last in the Industrial automation YOLOv8 and EfficientDet Helps in object tracking, ,quality checking, maniputaing tasks through Robotics which heps in improving productivity.

III.DATASET

We have Collected a large dataset from multiple open sources to train our model, the different data into dataset is always the first step after confirming the aim of the project so we have collected a large dataset dividing into three Test , Train and validation set of the data.One thing which ned to be taken care of is object size, background details and quality of the image should be proper for better performance.Here annoted format is also kept for the betterment of the algorithm performance. Annotate the collected images with bounding boxes to indicate the location and extent of objects in the images. Use annotation tools such as LabelImg or CVAT to create XML or JSON files containing object annotations.After that augmentation of the data is done whick helps in traiuning over correct data features like scaling,rotation, flipping,etc.,

The YOLOv8 algorithm is operated on an S × S grid, where each grid cell will be responsible for predicting a fixed number of bounding boxes (typically B) and corresponding confidence scores is also recorded. The formula for generating bounding box predictions needs to bedividided in grids of the given dimension , predicting bounding box offsets, scales, and probabilities and applying non-maximum suppression (NMS) to filter out redundant detections as it may reduce and affect the efficiency of the model while working on redundant images and dull images.

2. Loss Function: The loss function measures the difference between the model's predictions and the ground truth (actual labels). It's used to guide the training process by minimizing the loss..

3. Localization Loss **:** Localization loss measures how far off the predicted bounding box is from the ground truth (actual) bounding box. It penalizes the model for inaccurate positioning of the bounding .

4. Classification Loss:

FORMULA:Σ denotes summation over all predicted bounding boxes.`class_pred` represents the model's predicted probability distribution for different object classes.`class_gt` is a one-hot encoded vector representing the ground truth class label for the object in the bounding box.

5. **Intersection over Union (IoU):**

Formula:

IoU = Area of Intersection / (Area of Predicted Box + Area of Ground Truth Box - Area of Intersection)

6. Non-Maximum Suppression (NMS)

NMS is a post-processing technique used in object detection to eliminate redundant bounding boxes for the same object. It ensures the model outputs only the most confident and non-overlapping detections.

IV. METHODOLOGY

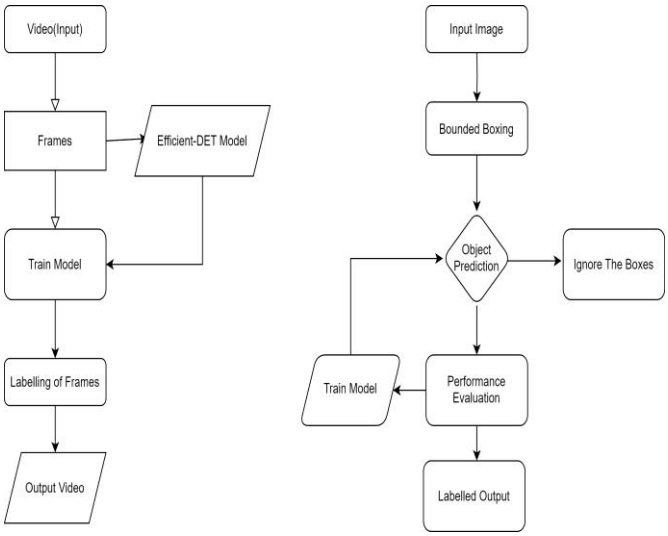

Figure 1: Methodology Followed

A. Data Collection and Preprocessing: Collecting a large And different data into dataset is always the first step after confirming the aim of the project so we have collected a large dataset dividing into three Test , Train and validation set of the data.One thing which ned to be taken care of is object size, background details and quality of the image should be proper for better performance.Here annoted format is also kept for the betterment of the algorithm performance. Annotate the collected images with bounding boxes to indicate the location and extent of objects in the images. Use annotation tools such as LabelImg or CVAT to create XML or JSON files containing object annotations.After that augmentation of the data is done whick helps in traiuning over correct data features like scaling,rotation, flipping,etc.,

B. Training:In the training of the YOLO model ARQ having pretrained weights form the beginning. At first we have distributed a dataset into three sub sets named Train , Test, Valid set of data for better implementation. Selection of right hyperparameters such as batch size, no. of epochs were accurately done for getting accurate configuration.For EfficientDet model video streams were divided into frames whicgh when passed to EfficientDet algo trains the model over it .Now Training of the model is done over the Train set of data having a combination of localization ,confidence and classification and loss function.Monitorin the model's performance is the last step on the validation set and fine-tune hyperparameters as needed to done..

Figure 3: Confusion Matrix

C. Libraries and Modules Used:.

1) OpenCV (Open Source Computer Vision Library): *Opencv is an open-source library which provides various utilities related to computer vision task.(Computer vision means interpreting and understanding visual information like videos and images). It helps in processing videos and images to identify handwriting, faces or any other object.*
2) Pandas and Numpy: *Pandas is the library used mainly for data manipulation as it provides different useful datastructures to handel labeled data such as tabular data and CSV files.*
   Where as , Numpy is mainly used for manipulating Arrays and metrices having multidimensions.Numpy is used for Numerical calculations as it provides useful functions for it.
3) Shutil: *It is a acronym for shell utilities . It is a module which provides multiple predefined functions for files and directories operational works. It is used for operations like copying, renaming or deletion.*
4) Tqdm: *Tqdm is derived from Arabic word 'taqaddum' which means 'Progress' It is a famous Python library which provides function for adding progress bars to loops and iterable objects. This is mainly done for the developers for having the track over the progress of work.*

Model Evaluation is one of the important step as we evaluate the trained models using over the test set using different techniques like map,f1 score, precision, recall etc,. The visual data will help us to determine its evaluation in better manner .Now verification is also done to check whether the model is correctly able to localize and identify the object in the real world scenario.

D. Performance*: Model optimization need to be done in any project whether by compression I.e reducing the size of the model or pruning can also be done .*

V. RESULTS

In this project using YOLOv8 and EfficientDet model has shown tremendous results, demonstrating the efficiency of the models to detect the objects in the excellent manner with higher precision. Both the models have shown high accuracy towards their dataset that showcase their applicability to the real worls scenarios.

The YOLOv8 model has shown the exceptional accuracy in detecting static objects within its images, which contains vehicles, pedestrians, and traffic signs. As by training accurately over annoted dataset it consistently resulted with high precision and accurately localizing and classifying objects .

Figure2: Resulted Images showing a Truck, 2 Person and 1 Motorcycle.

We can see by visualizing the resulted images that YOLOv8 model accurate precision of bounding boxes localizing the object perfectly , and classification of the object is done correctly. Its reliability is can be measured on the basis of least false negative and positives.hence , efficient to perform the real world detection images.

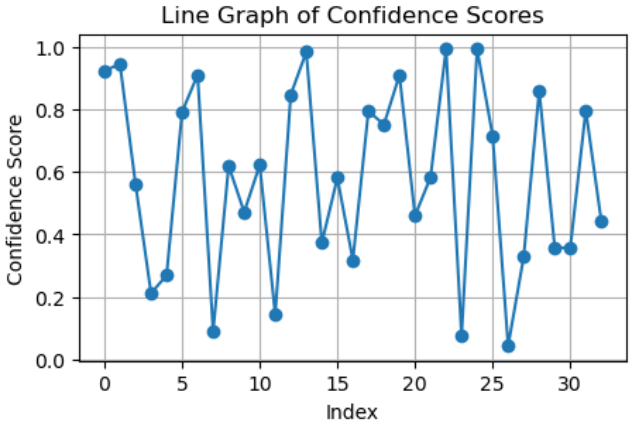

Figure 3: Confidence Score

Figure 4: Resulted Images showing a Bus, 1 Truck and 1 Motorcycle

Similarly, the other model i.e. EfficientDet is also able to Perform impressively in tracking the moving objects. In the video streams the fast moving objects are better to be detected and tracked using EfficientDet model.It is also able to detect the crowd in urban areas .We can say that EfficientDet model can perform in the challenging conditions like fast-moving object displays its ability to perform robustly. Its High accuracy shows its ability to efficiently.The model tracks rthe moving objects in frames and provide us the bounding boxes on the final state .it also localize and tells the classification of the object accurately..

Now, the results of the object detection and tracking using YOLOv8 and EfficientDet models have been highly appreciatable after looking at the model's accuracy in detection and labeling the objects.As it can be seen here both the models has shown their accuracy in different aspect of detection tracking respectively. which will significantly contribute in the increasing demand of Computer vision tech, and can be applicable in various domains in the field field of the ML.

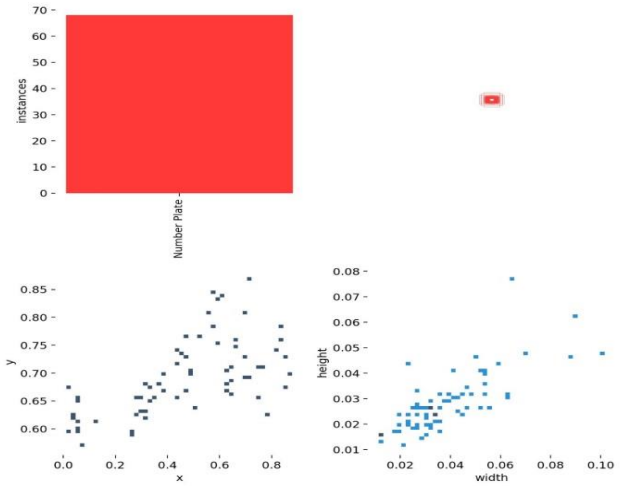

Figure 5: Graph Regarding their feature comparison.

The successful working of YOLOv8 and EfficientDet models tells their ability to identify complex objects and will help in solving the complex problems As futher feedback after the real-world deployment are expected to enhance thei performance and will make it adaptable to diverse use cases and environments.

VI. CONCLUSION

This research paper proposes a YOLOv8 model and EfficientDet model which is one Stage network algorithm using state-of-art model to detect and classify objects .

Through the implementation and experimentation, we have shown the efficiency and potential of these models in different complex scenarios having dull and low pixel images on the board .The result shown above demonstrate the high capability of both the models in detecting and tracking the objects respectively with high accuracy and efficiency, As seen in above resulted images the bounding boxes were accurately localized and classification was also done very well whereas the EfficientDet model helped to track the fast-moving objects in great manner .Apart from all these great accuracy achieved by both of these models ,we can not deny from the fact that there are still some situation where both the models lack in the detection and localization like in Densly populated scene the model lacks to detect the individual object which can still be its limitation. This problem can be called under the overlapping objects topic which still needs some more research and innovation in this increasing demanded technique for better classification of the objects .In addition to it, when our research was performing very well it is very important to recognize the errorness and false positive in the real worlds scenarios ,like varying lighting in the different images can drastically reduce its efficiency of localization which also tells the importance of the validation and continous testing over required dataset to enhance its capability.Now looking at the further research , the future of the object detection still has the chances of further improvements and advancement in the technology of computer vision. By showing the challenges faced and advancement needed in this domain will help in advancement and innovation in this field . Future researchers can also focus on the improvement of the existing previous models , developing good and efficient algorithms ans can integrate with the multi-model data to have better accuracy on the real world applications.

In conclusion, while our research has made significant strides in advancing object detection using YOLOv8 and EfficientDet models, there remains much to explore and improve upon. By embracing the complexities and challenges inherent in real-world scenarios, we are poised to unlock new insights, capabilities, and opportunities for innovation in the field of computer vision and beyond.

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)

[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] Repository and documentation by Ultralytics.

[7] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv.

[8] Yang, L., Wang, L., & Wu, S. (2018, April). Real-time object recognition algorithm based on deep convolutional neural network. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 331-335). IEEE.

[9] OpenCV: Open Source Computer Vision Library. (n.d.).