iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70152

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Autonomous Navigation Using Deep Learning

Aman Kumar[1], Danesh Sharma[2], Gurpreet Singh[3], Kuldeep Kumar[4], Asst. Prof. Jagveer Singh[5]
Department of Computer Science Meerut Institute of Engineering and Technology Meerut, India

***Abstract:** Withusesinrobotics,industrialautomation,autonomousvehicles,andsurveillance,objectdetectionis a basic computer vision problem. WithinthecontextoftheCOCOdataset,thisworkcomparestheperformanceof severalstate-of-the-artobjectrecognition models,includingMaskRCNN (Detectron2),YOLOv8s,YOLOv8l,and YOLOv11s.Some of the significant parameters such as mean Average Precision (mAP), precision, recall, and inference speed are utilized to compare models.*

The results indicate that while Mask R-CNN is accurate, its computation makes it less suitable for real-time use. YOLOmodels,particularlyYOLOv8s,arehoweveracompromisebetweenaccuracyandspeedandthusareideal for real-time detection processes.YOLOv8l is however computationally more demanding but somewhat offers higheraccuracy. Duetoitsspeedand accuracy,YOLOv8sisthemostsuitablemodeltoapplyinreal-time,asstated in the review. In selecting the most suitable object detection models for various applications, researchers and developers can learn a lot from this study.

***Keywords:** YOLO,DETECTRON,R-CNN, Object Detection.*

I. INTRODUCTION

Object detection is one of the pillars of computer vision that allows machines to detect and locate objects withinanimageorvideo.Itisusedinawidevarietyofapplicationsinrobotics,industrialautomation,medical imaging, autonomous cars, and security.Several object detection models based on deep learning have been proposed in recent years to make it faster and more accurate. While deciding on the most suitable model for real-time object detection, we are considering Mask R-CNN (Detectron2), YOLOv8s, YOLOv8l, and YOLOv11s based on their performances.

Prior object detection methods utilized region-based methodologies, like Faster R-CNN, which were computationallyintensive but very precise. YOLO algorithmshave, nonetheless,transformedobjectdetection by remarkably optimizing speed with no compromise in accuracy.While Detectron2, which is an implementationofMaskR-CNN,isveryefficientinsegmentingobjectsprecisely,itshighcomputationalcost renders it inappropriate for real-timeUse.

Based on the COCO dataset, herein, various models are comparedon thebasis of testing of remarkableperformance measures like precision, recall, inference time, and mean Average Precision (mAP).Determination of the most balanced model regarding speed and accuracy for various applications is the focus.This study will guide developers and researchers in choosing the most suitable model for object detection applications.

II. LITERATURE REVIEW

Deepmodelarchitecture,particularlythatofYOLO-type models, Faster R-CNN, and SSD, has revolutionized object detection. Some other approaches with a focus on segmentation and tracking included the unsupervised Siam Mask model, which ranked top in self-driving object segmentation and tracking experiments [1].

Real-time processes have heavily depended on Detectron2, which is a mature object detection model based on Faster R-CNN. Detectron2 employs region proposal networks to achieve higher accuracy and detects traffic entities in real-time through Faster R-CNN efficiently, according to research [2].

Tinier-YOLO offers a very good solution for real-time processing by reducing computational complexity, and optimizations for constrained environments have also been realized in YOLO models [3]. With negligible real-time processing rate compromise, YOLO v3-Tiny optimizations significantly improved detection accuracy [4].

YOLO modelfamilyhavebeen focuseduponheavilyby objectdetectionresearches.YOLOwasinitiallypresentedin earlier work as a unified real-time object detection method able to detect a high number of variant objects [5][6].The model's speed was enhanced while keeping accuracy, incorporating hierarchical detection with YOLO9000 [7]. YOLOv3 continued to advance in order to enhance feature extraction as well as its ability to detect small objects [8].

There has been a great amount of research also that has beendoneonobjectdetectionmethodsregion-based.RCNN establishedthegroundworkforthenextgenerationofmodels with dense feature hierarchiesallowing object segmentation and identification [9].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

With the addition of region proposal networks, Faster R-CNN added a great amount of speed up withoutanydecrementinaccuracy[10].InSSD,yetanother widely employed class, was illustrated a one-shot detection architecture which efficiently balanced between speed and accuracy [11].

CNNs and residual learning models are also the key to further promoting object detection.The success of deep learninginimageclassificationproblemswasestablished with theadventofconvolutionalneuralnetworks(CNNs)in ImageNet classification [12].Residual learning techniques enabled networks that were much deeper without gradient vanishing, further promoting feature extraction [13]. To enhancedenseobjectdetectionmodelsandenhanceaccuracy inthecaseofclassimbalance,focallossalgorithmshavealso beendeveloped[14].Moreover,theinclusionofOpenCVin computer vision software has made it easier to implement object detection frameworks [15].

III. METHODOLOGY

A. *About the Dataset*

During this research, the Common Objects in Context (COCO) dataset, being the most popular dataset used for objectdetection,imagesegmentation,andimagecaptioning,isutilized.COCOistheworldlargestobjectdetectiondatasetwithmoretha n200,000labeledimagesand80objectclasses.Denseannotationsofthedatasetasboundingboxes,instancemasks,andkeypointsenable successfuldeeplearningmodel training and assessment. For convenience, we particularly use theCOCOvalidationsetmadeavailableonKaggletotryout part of our object detection models used in the project. Because objects in the dataset vary in size (small, medium, large),wearethusin apositionto comparetheperformance ofvariousmodelsasafunctionofobjectsizevariability.Itis ideal for testing the generalization capacity of existing YOLO and Mask R-CNN models because it is challenging, with occluded objects and challenging backgrounds. We want to carry out an unbiased and balanced assessment of various detection models in COCO.

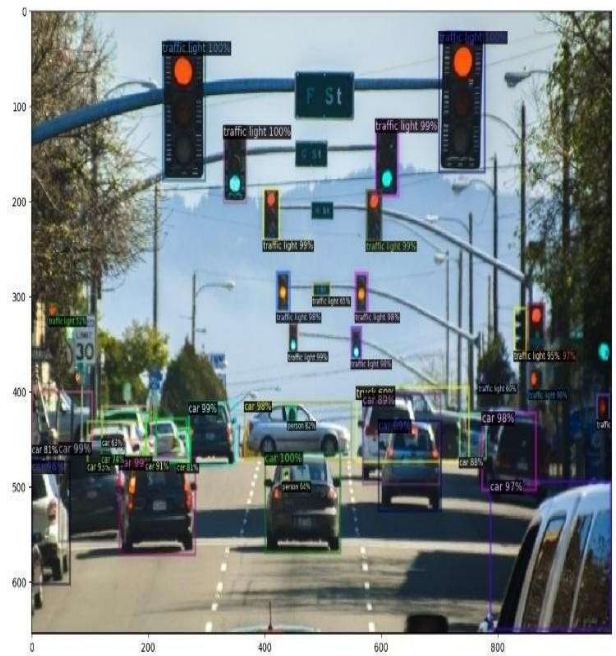


Figure1.ObjectsDetectedimagefromdataset

B. *OverviewofObjectDetectionFrameWorks1.Detectron2*

Facebook AI Research (FAIR) came up with the cuttingedge object detection architecture known as Detectron2. Its extensible and modular building blocks for the object detection and segmentation tasks are based on PyTorch.Detectron2supportshigh-performancedetectionof objectswithawiderangeofmodelsincludingFasterR-CNN, Mask R-CNN, and RetinaNet. It also features Region Proposal Networks (RPN) and Feature Pyramid Networks (FPN)toprovideimprovedaccuracy.Itissimplertoemploy onrealworldproblemsusingthepre-trainedmodelsthathave beentrainedacrosstheCOCOdataset.DuetoitsstrongGPU acceleration optimization, Detectron2 offers training and inferenceefficiency.Peoplecantrainmodelsforcertaintasks aswellbecausetoitscapacitytohandlebespokedatasets.

COCO metrics for performance evaluation are some of the tools within the framework. Detectron2 is widely used by videoanalysis,medicalimaging,andautonomousdriving.It is accessible to researchers and developers since it has a straightforwardAPI.

Yolo

The cutting-edge object detection model YOLO (You Only Look Once) is renowned for its real-time speed and accuracy.By formulating object detection as a one-stage regression task,YOLO both predicts class probabilities and bounding boxesfromtheinputimageatonce,asopposed to traditional region-based detectors.This does not entail large processing overhead, thereby making it extremely fast for real-timeapplicationslikerobots, autonomous vehicles, and surveillance. In order to provide real-time and effective localization, YOLO utilizes a deep convolutional network that splits an image into a grid and allocates detection work to all the cells in the grid. Since the model has undergone several revisions from YOLOv1 to YOLOv8, each of them has improved precision, velocity, and potency. To provide improved feature extraction as well as generalization, new releases include transformer-based advancements and leading-edge topologies like CSPDarknet. Additionally, YOLO also accommodates multi-scale detection, which enables it to detect differently sized objects.

Itwillbecapableofgeneralizingformanysituationssinceit was trained on large datasets such as COCO.

YOLOv8-nanoandYOLOv8-smallare reducedsize models thathavebeenoptimizedfordeploymentonprocessing-poor devices. YOLO continues to be a leading object detection framework used in research and also applied in commercial deploymentsbecauseofitsabilitytostrikeabalancebetween accuracy and speed.

C. DifferenceBetweenDetectron2andYolo

While both Detectron2 and YOLO are robust object detection models, there are a few differences in application, structure, and strategy between the two models as well. While the Faster R-CNN and Mask R-CNN models were designed to do a few things, Facebook AI developed Detectron2 to detect objects accurately and segment instances. It is expensivecomputationallybuthighlyprecisebecauseituses a two-stage detection model by first generating regions of interestand refining themsubsequently.YOLO is one-stage detection, and this is highly efficient and quick real-time runner as it predicts bounding boxes and class probability simultaneously.As YOLO is efficiency-oriented, it is utilized whereinferenceneedstobeexecutedasquicklyaspossible, i.e., self-driving vehicles and surveillance cameras. Slower but more efficient at executing tasks such as panoptic segmentation and occluded objects is Detectron2. While Detectron2 is more computationally intensive, YOLO models such as YOLOv8 are light and can be run on edge devices. WhileDetectron2isidealtodeeper levelsin ordertoenable deeper learning and investigation, YOLO is ideal for a beginner.Itissimpletocomprehendandisnotcomplicated. Thedecisiontheniswhetheronewouldpreferdetectionand segmentation with high precision (Detectron2)or inreal-time (YOLO).

IV. ARCHITECTURE OF MODELS

1) YOLOv8l

The head-neck-backbone architecture of YOLOv8l (Large) is optimized for accuracy over efficiency. The backbone utilizes CSPDarknet, gradient flow-friendly, with reduced redundancy, and feature extraction improved through cross-stage partial connections. PANet (Path Aggregation Network)isemployedwithintheneckformultiscalefeature fusion withstrong objectdetection irrespectiveofsizes.The detecting head is anchor-free, and this improves the localization accuracy and makes bounding box regression easier. Decoupled head architecture is employed for higher accuracy, wherein it decouples the task of localization from thetaskofcategorization.Depthwiseseparableconvolutions are utilized in the model to avoid compromising accuracy with an increase in computational burden. YOLOv8l achieved better results compared to the lower models in project evaluation with precision of 0.810, recall of 0.690, mAP50 of 0.775, and mAP50-95 of 0.610. A few of the reasonsfortheerrorsarefalsepositivesonthickbackgrounds and lack of capability to identify objects that are little in occluded scenes. YOLOv8l is perfect for high-performance real-time detection because it neither loses speed nor accuracy despite these constraints.

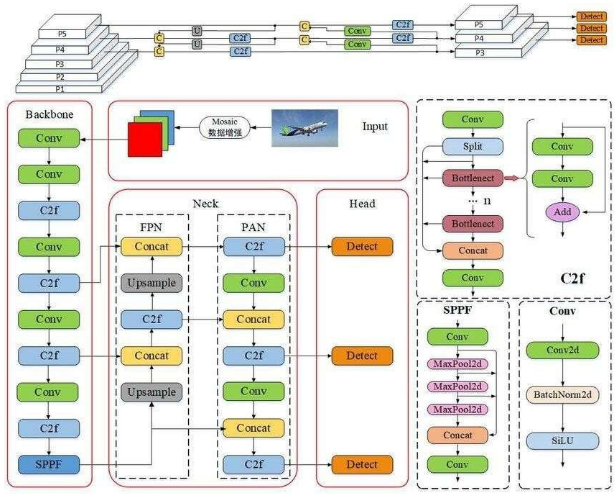


Figure2YOLOv8Models

2) YOLO11s

YOLO11s,apriciervariantoftheYOLOseries,supportsrealtimeobjectdetectionwithincreasedspeedandaccuracy.Forextractingfeaturesatreducedcomputationalexpenses,it follows a backbone-neck-head structure with a CSPDarknetbased backbone.Small and large object detectionisenhancedthroughadditionalfine-grainedmulti- scale fusion of features by an enhanced Path Aggregation Network (PANet). Since the anchor-free structure of the detection head contains no special anchor boxes, the prediction flexibility is enhanced.YOLO11s applies depthwise separable convolutions in an attempt to enhance the computing efficiency without compromising the detection accuracy.Even more sophisticated than any other YOLO model,YOLO11shasbeendesignedtodetectobjectsinreal- time with increased speed and accuracy.For easier computational feature extraction, it employs a backbone- neck-head architecture with a CSPDarknet backbone.The neck encourages small and large object detection through multiscale feature fusion optimization via an improved Path AggregationNetwork(PANet).Duetotheanchor-freenature ofthedetectionhead,someoftheanchorboxesarenot necessary, and prediction is more flexible. Depthwise separableconvolutionsareemployedbyYOLO11stoobtain computation efficiency without loss of detection accuracy.

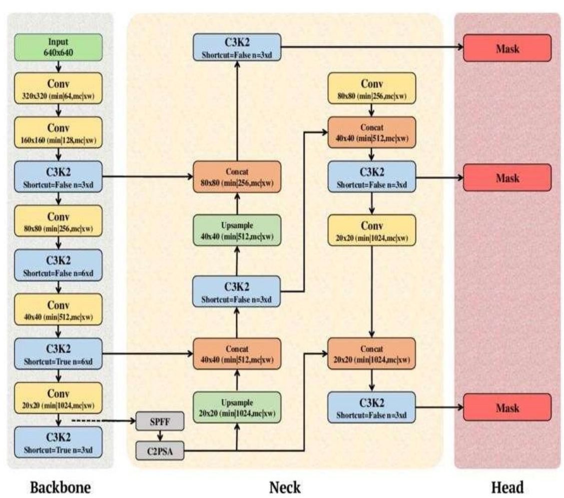


Figure3YOLOv11sModels

3) DETECTRON2(R-CNN)

Employinga FeaturePyramid Network (FPN)multi-scale feature extraction, Detectron2 model Mask R-CNN R_50_FPN_3x is a two-stage object segmentation and detection model with a ResNet-50 backbone.

A Region ProposalNetwork(RPN)generatescandidateobjectregions duringthefirststage.Duringthesecondstage,theproposals are enhanced, objects class-annotated, and bounding boxes predicted in addition to instance segmentation by an independent mask head. By fusing low-level and high-level features, FPN encourages feature learning and enhances detectionprecisiononvariousobjectsizes.Alongertraining duration is denoted by the "3x" in the model name name, which enhances precision and convergence. This model performed well at most IoU thresholds with Average Precision(AP)of0.375onCOCO2017validationset,AP50 of0.546,andAP75of 0.419.WithanAverageRecall(AR)of 0.445, themeasuresofrecallarereflectingtheerrorrateand include some missed detections, i.e., on small objects (AR small=0.232).Themodeliscomputationallyexpensiveand needsaGPUtorunin real-time inferencebuthasverygood accuracy.

TheYOLOv8s architecture is a lightweight and efficient object detection model optimized for real-time applications. It utilizes a CSPDarknet backbone for feature extraction, a PANet(PathAggregationNetwork)forfeaturefusion,anda YOLO detection head for final predictions. The model incorporates decoupled heads to improve accuracy in classificationandlocalizationtasks.Itprocessesimagesina single pass, making it significantly faster than two-stage detectors like Faster R-CNN. In our project, YOLOv8s achieved a precision of 78.42%, a recall of 67.05%, an mAP50 of 76.01%, and an mAP50-95 of 58.79% on the COCO validation dataset. These results indicate that YOLOv8s provides a good balance between speed and accuracy, outperforming older YOLO versions while being computationally efficient. However, its performance is slightlylowerthanlargermodelslikeYOLOv8lbutremains more suitable for deployment on edge devices. The model also demonstrated improved inference speed compared to Detectron2, making it ideal for real-time detection applications.Itslightweightnatureandefficient architecture

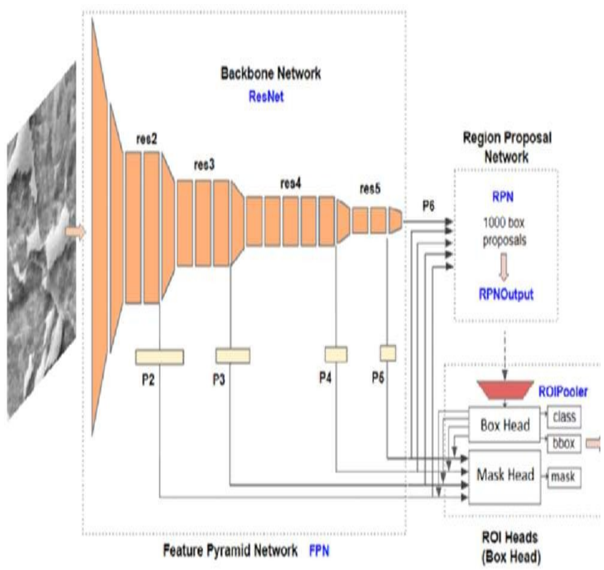


Figure4Detectron2Models

4) YOLO8s

make it a strong choice for tasks requiring both speed and accuracy in object detection.

V. RESULTS OBTAINED

WeInthepresentcase,wearetryingtocomparehowwell orefficientlysomeobjectdetectionmodelsareontheCOCO dataset,forexample,MaskR-CNN(Detectron2),YOLOv8s, YOLOv8l, and YOLOv11s.

AveragePrecision(AP),AverageRecall(AR),precision, and inference speed are a few of the used metrics.

WithAPof0.375,AP50of0.546,andAP75of0.419,

Mask R-CNN (Detectron2) was extremely accurate in detection with relatively higher computational expense. Recall(AR=0.445)reflectsthelimitationofdetectingsmall objects efficiently.

With the decreased inference time of 4.95ms,YOLOv8s achievedanmAP50of0.760andanmAP75-95of0.587and outperformed Mask R-CNN in speed and accuracy but marginally.

Higher recall and accuracy were achieved byYOLOv8l, which was bigger in size but with greater computational requirements.

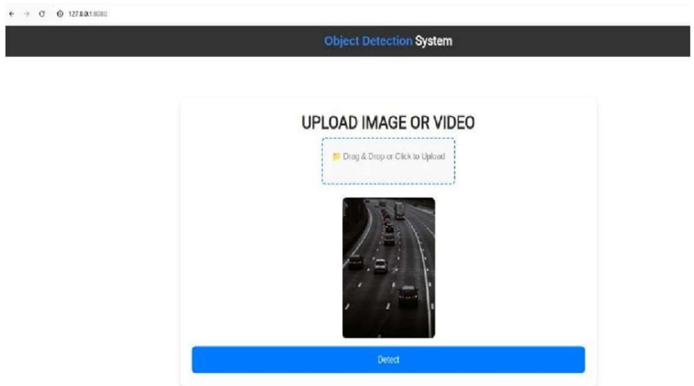


Figure4Rawimage

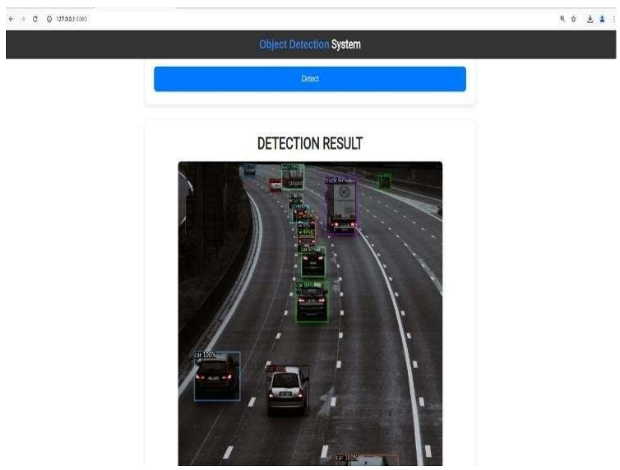


Figure5ObjectDetected Image

YOLOv11sachievedthesameperformanceofYOLOv8s at mAP50-95 of 0.578 but with the ability to perform better under some detection conditions.

YOLO models offer a better trade-off between real-time speed and accuracy and are therefore suitable for real-time tasks such as autonomous driving and video tracking, but MaskR-CNNisgoodatchallengingsegmentation.The relative importance of speed, accuracy, and computational expenserequiredbyanoperationdeterminesthemodeltobe used.

Model	AP (IoU 0.50:0.95)	AP50	AP75	Precision	Recall	Inference Time (ms)	Best Use Case
Mask R-CNN (Detectron2)	0.375	0.546	0.419	-	0.445	High (~50ms)	High-accuracy tasks, segmentati
YOLOv8s	0.588	0.76	-	0.784	0.67	4.95	Real-time application edge device
YOLOv8l	0.6	0.77	-	Higher	Higher	~7-10	Balanced accuracy & speed
YOLOv11s	0.578	0.751	-	0.722	0.693	~6.86	Fast detection with moderate accuracy

Table1ComparisontableofObjectDetectionModels

VI. CONCLUSION

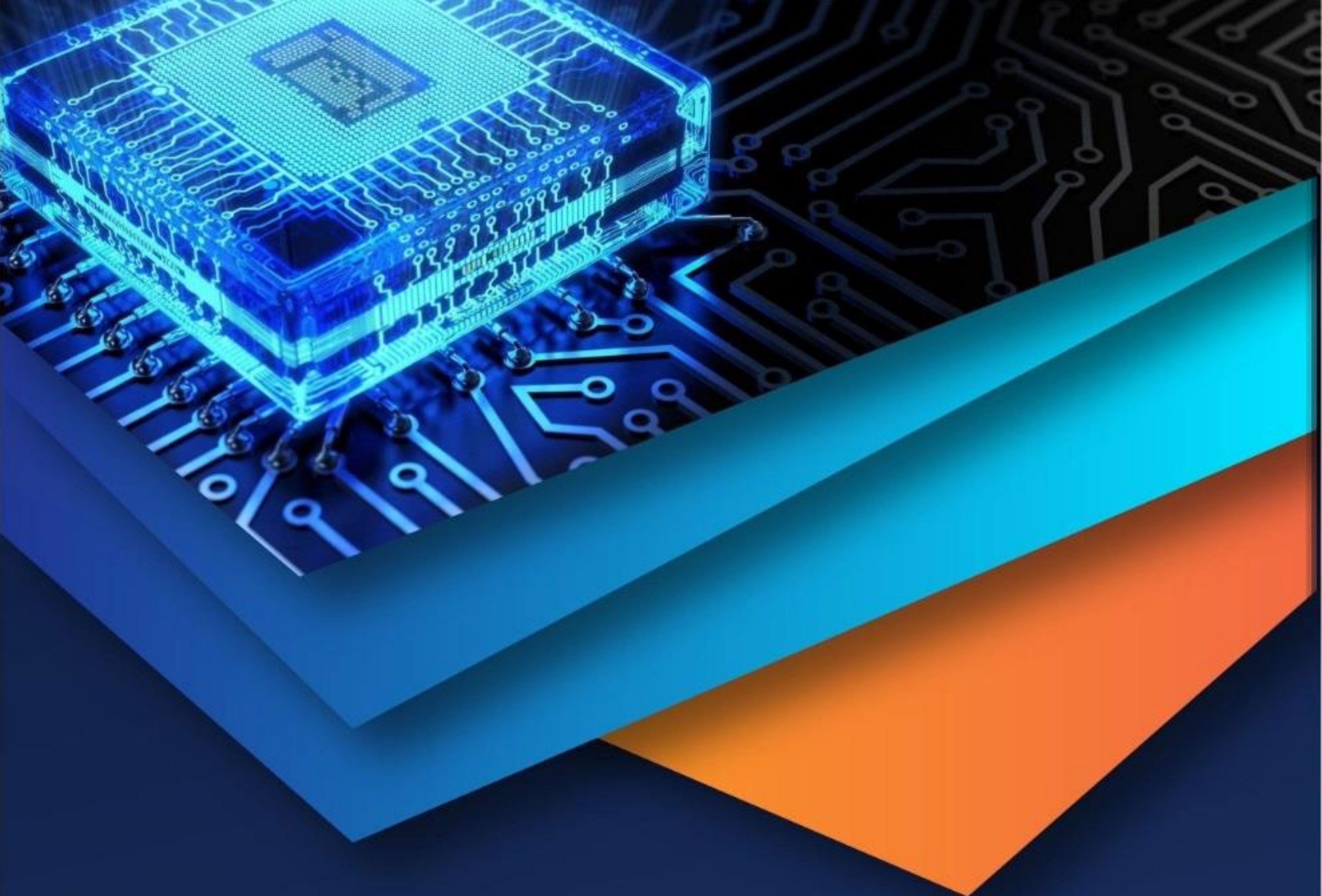
Herein,wehaveexperimentedandcomparedvariousobject detectionmodelssuchasMaskR-CNN(Detectron2), YOLOv8s,YOLOv8l,andYOLOv11son accuracy measures (AP, mAP), precision, recall, and inference time with the COCOdataset.OurresultsindicatethatalthoughMask RCNNisappropriateforinstancesegmentationandhasvery highaccuracy (AP=0.375,AP50=0.546), itconsumesalot of resources and thus is not very appropriate for real-time use. Conversely,YOLOmodelssuchasYOLOv8sandYOLOv8l performedbetteratanmAP50of0.760and0.770, respectively, butwith much lower inference times (4.95ms for YOLOv8s), but YOLOv8l was better than YOLOv8s with increased computations.

WhileYOLOv11modelshaveanmAP50-95valueof0.578, they were not significantly better compared to YOLOv8 models.

The topreal-timeobjectdetection model isYOLOv8s when inferencespeed,accuracy,andcomputationalefficiency trade-offs are considered.It can be used for real-time tracking, surveillance, and autonomous use due to its accuracy-speed ratio.Model optimization and hybrid methods to improve detection efficiency can be explored in future research.

REFERENCES

[1] S. Noor, M. Waqas, M. I. Saleem, and H. N. Minhas, “AutomaticObject Tracking and Segmentation Using Unsupervised SiamMask,”IEEE Access, vol. 9, pp. 106550-106559, 2021.
[2] A.Obi-Obuoha,V.S.Rizama,I.Okafor,H.E.Ovwenkekpere,K.Obe,and J. Ekundayo, “Real-timetraffic objectdetection using detectron 2with faster R-CNN,” World Journal of Advanced Research andReviews, vol. 24, no. 02, pp. 2173-2189, 2024.
[3] W. Fang, L. Wang, and P. Ren, “Tinier-YOLO: A Real-Time ObjectDetection Method for Constrained Environments,” IEEEAccess, vol.8, pp. 1935-1944, 2020.
[4] P.Adarsh,P.Rathi,andM.Kumar,“YOLOv3-Tiny:ObjectDetectionand Recognition using one stage improved model,” in 2020 6thInternational Conference onAdvanced Computing & CommunicationSystems (ICACCS), 2020, pp. 687-691.
[5] C. Liu,Y.Tao, J. Liang, K.Li, andY. Chen,“ObjectDetection Basedon YOLO Network,” in 2018 IEEE 4th Information Technology andMechatronicsEngineeringConference(ITOEC2018),2018,pp.5054.
[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only lookonce: Unified, real-time object detection,” arXiv:1506.02640, 2015.
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “YOLO9000:Better, Faster, Stronger,” arXiv:1612.08242v1, 2016.
[8] J.RedmonandA.Farhadi,“YOLOv3:AnIncrementalImprovement,”arXiv:1804.02767,2018.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich featurehierarchies for object detection and semantic segmentation,” inProceedings of the IEEE conference on computer vision and patternrecognition, 2014, pp. 580-587.
[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towardsrealtime object detection with region proposal networks,” IEEETransactionsonPatternAnalysisandMachineIntelligence,vol.39,no.6, pp. 1137-1149, 2017.
[11] W.Liu,D.Anguelov,D.Erhan,C.Szegedy,S.Reed,C.Y.Fu,andA.
[12] C.Berg,“SSD:Singleshotmultiboxdetector,”inEuropeanconferenceon computer vision, 2016, pp. 21-37.
[13] A. Krizhevsky, I. Sutskever, and G.E.Hinton,“ImageNet classificationwith deep convolutional neural networks,” Advances in neuralinformationprocessing systems,vol.25,2012.[13]K. He,X. Zhang,
[14] S. Ren, and J. Sun, “Deep residuallearning for imagerecognition,”inProceedings of the IEEE conference on computer vision and patternrecognition,2016, pp.770-778.[cite:1011, 1012, 13,14,15,16, 45].
[15] T.-Y.Lin,P. Goyal,R. Girshick,K.He,and P. Dollár, “Focalloss fordense object detection,” in Proceedings of the IEEE internationalconference on computer vision, 2017, pp. 2980-2988.
[16] G.BradskiandA.Kaehler,“LearningOpenCV:Computervisionwiththe OpenCV library,” O'Reilly Media, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY