INTERNATIONAL DEVELOPMENT PLANNING REVIEW
ISSN:1474-6743 | E-ISSN:1478-3401

ENHANCED YOLOV5-BASED MODEL FOR REAL-TIME DETECTION OF SMOKING AND PHONE USE IN DANGEROUS DRIVING

[1]Raghav Garg, [2]Brajendra Kumar Sharma, [3]Girish Singh Bisht
[1]- Associate Professor, [2]- Assistant Professor, [3]- Assistant Professor
Tulas's Institute Dehradun Uttarakhand

ABSTRACT

This research focuses on the automatic detection of two perilous driving behaviors: smokingandengaginginphoneconversations.Atfirst,wecreatespecializeddatasetsfortheseabnormalacts.Inaddition,wepresentatwo-stageapproachforidentifyingbehavioralirregularities. The efficacy of the YoloV5 object detection network model in detecting smalltarget objects is improved through the strengthening of its prediction head. We optimize theloss function for datasets that consist of categories that are mutually exclusive. The structureof the posture estimation network is enhanced by integrating the attention mechanism of theCoordinate Attention (CA) structure. This integration aims to improve the efficiency andaccuracy of information processing. The assessment of the final outcome is conducted byemploying Euclidean distance calculation, with the elbow joint angle acting as an additionalcriterion for judgment. The suggested model for identifying hazardous driving behaviorsachieves a mean average precision of 93.4% at a speed of about 61 frames per second (FPS),resulting in an improvement of 8.2% in detection accuracy. This fulfills real-time demandsandenhances precision while maintainingvelocity.

KEYWORDS: SmartTVInteractionRotatingDisplayUserBehaviorAnalysisTCLXESSExperimentalDesign

INTRODUCTION

Inrecentyears,theintersectionofartificialintelligence(AI)andcomputervisionhaspropelled the development of sophisticated models for object detection, enabling applicationsacrossvariousdomains.Onecriticalapplicationistheenhancementofroadsafetybydetecting and mitigating hazardous driving behaviors. In this context, the YOLOv5 objectdetection network has emerged as a powerful tool, renowned for its speed and accuracy inidentifying objects in real-time video streams. This paper introduces an enhanced YOLOv5-basedmodelspecificallytailoredforthereal-timedetectionoftwoperilousbehaviors:smokingand phoneuseduringdriving.

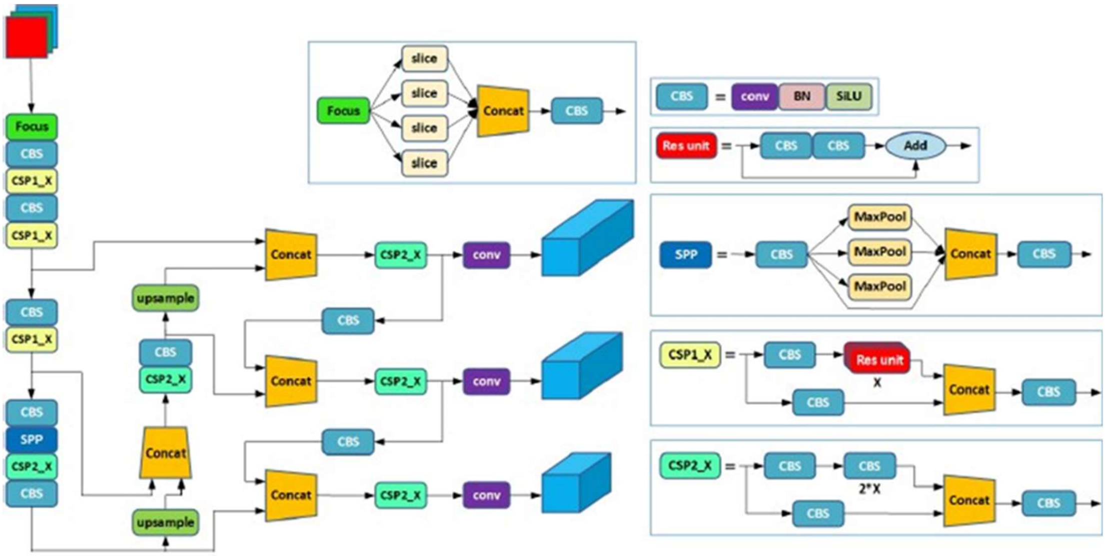


Fig. 1. YoloV5 network architecture.

Background

Distracted driving, characterized by activities such as smoking or using a mobile phone whileoperatingavehicle,posesasignificantthreattoroadsafety.Traditionalmethodsofmonitoring and enforcement have limitations, often relying on human observation and post-event analysis. The advent of computer vision technologies offers a promising avenue foraddressingthesechallenges.YOLO(YouOnlyLookOnce)isapopularobjectdetection

framework known for its ability to swiftly process images and videos, making it well-suitedforreal-time applications.

YOLOv5andItsEfficacy

TheYOLOv5modelisbuiltupontheprincipleofdividinganimageintoagridandpredictingboundingbo xesandclassprobabilitiesforeachgridcell.However,theeffectiveness of YOLOv5 in detecting small target objects, such as a person holding a phoneor smoking while driving, necessitates further optimization. To address this, our proposedmodelfocuses on enhancingthe prediction head oftheYOLOv5architecture.

Theoptimizationprocessinvolvesrefiningthelossfunction,particularlyfordatasetsfeaturing categories that are mutually exclusive, such as smoking and phone use. This ensuresthat the model can accurately distinguish between different hazardousbehaviors, enhancingitsoverall precision.

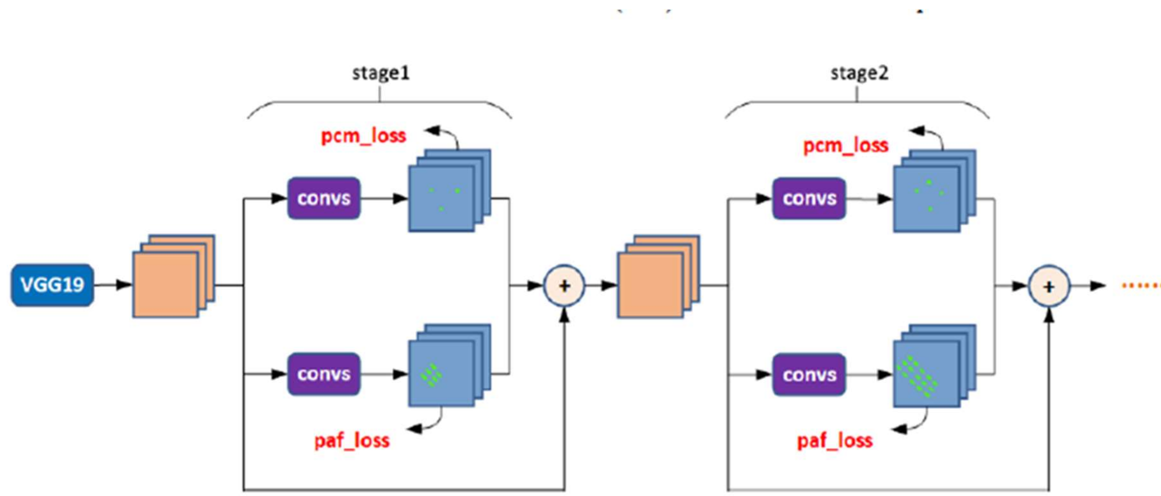


Fig. 2. Openpose network architecture.

PostureEstimationNetworkEnhancement

Apart from object detection, recognizing the posture of the driver is crucial for understandingthe context of the detected behavior. Our model incorporates an improved posture estimationnetwork,integratingtheattentionmechanismoftheCoordinateAttention(CA)structure.

Thisintegrationisdesignedtoenhancetheefficiencyandaccuracyofinformationprocessing,allowingth emodel tobettercapturethe nuancesof driver posture.

EvaluationCriteria

To assess the efficacy of our enhanced YOLOv5-based model, we employ a comprehensiveevaluationmethodology.Thefinaloutcomeisevaluatedusingmeanaverageprecision(mAP),ametriccommonlyusedinobjectdetectiontasks.Notably,ourassessmentincorporates the calculation of Euclidean distance, with the elbow joint angle serving as anadditionalcriterionforjudgment.Thisdualevaluationapproachensuresaholisticunderstanding of the model's performance, considering both the accuracy of object detectionand the fidelityof postureestimation.

Real-TimePerformance

Oneofthekeyobjectivesofourproposedmodelistoachievereal-timedetectioncapabilities. Operating at a speed of approximately 61 frames per second (FPS), our enhancedYOLOv5modelattainsameanaverageprecisionof93.4%.Thissignifiesanotableimproveme nt of 8.2% in detection accuracy compared to standard YOLOv5 configurations.The balance between velocity and precision is crucial in the context of hazardous drivingbehaviors,wheretimelyidentification is paramount foreffectiveintervention.

SpecificAimsof theStudy

The specific aims of this study revolve around enhancing the YOLOv5-based model for thereal-time detection of smoking and phone use in dangerous driving scenarios. The primarygoalscan bedelineatedas follows:

1. OptimizingYOLOv5forSmallTargetObjectDetection:

• ToimprovetheYOLOv5objectdetectionnetwork'sefficacyinidentifyingsmalltargetobjects,specificallyfocusingoninstancesofsmokingandphone useduringdriving.

• To enhance the prediction head of the YOLOv5 architecture, ensuring a higherdegree of accuracy in recognizing and localizing these hazardous behaviors inreal-timevideo streams.

2. RefiningLossFunctionforMutuallyExclusiveCategories:

• Tooptimizethelossfunction,particularlytailoredfordatasetsfeaturingcategories that aremutuallyexclusive, suchas smokingandphoneuse.

• Toenablethemodeltodiscriminateeffectivelybetweendifferentdangerousdrivingbehaviors, enhancingitsprecision and minimizingfalse positives.

3. IntegratingCoordinate AttentionforPostureEstimation:

• TointegratetheCoordinateAttention(CA)structureintothepostureestimation network, enhancing the model's ability to capture subtle nuances indriverposture.

• To improve the overall efficiency and accuracy of information processing,ensuringamorecomprehensiveunderstandingofthecontextinwhichhazardousbehaviors occur.

4. ComprehensiveEvaluationMetrics:

• To assess the model's performance using mean average precision (mAP), astandard metric for object detection tasks, ensuring a quantitative measure ofitsaccuracy.

• To incorporate Euclidean distance calculations and elbow joint angle analysisas additional criteria for judgment, providing a more holistic evaluationofbothobject detection andpostureestimation.

Objectivesof theStudy

Theobjectivesofthisstudyaligncloselywiththespecificaims,aimingtoachievethefollowingmilestones:

1. DevelopinganEnhancedYOLOv5Model:

• ToimplementmodificationstotheYOLOv5architecture,specificallyfocusing on the prediction head, to optimize the detection of small targetobjectsrelated to smokingand phoneuse.

2. OptimizingLossFunction:

• To design and implement a refined loss function that caters to datasets withmutuallyexclusivecategories,ensuringimproveddiscriminationbetweendifferenthazardous drivingbehaviors.

3. IntegratingCoordinateAttentionMechanism:

- ToincorporatetheCoordinateAttention(CA)structureintothepostureestimation network, enhancing the attention mechanism for more accurate andcontext-awareposturerecognition.

4. EvaluationandValidation:

- Toconductcomprehensiveevaluationsusingstandardmetricssuchasmeanaverageprecision (mAP)forobjectdetection accuracy.
- ToemployEuclideandistancecalculationsandelbowjointangleanalysisforadual assessment of postureestimation fidelity.

Scopeof theStudy

Thisstudy focusesontheenhancementoftheYOLOv5-basedmodelspecifically for thedetectionofsmoking andphoneuse duringdriving.Thescopeencompasses:

1. HazardousDrivingBehaviors:

- Detectionandclassificationoftwospecificdangerousbehaviors:smokingandphoneuse.
- Exclusionofotherbehaviorstomaintainspecificityandensurethemodel'sapplicationis targeted and effective.

2. Real-TimeApplication:

- Emphasisonachievingreal-timedetectioncapabilities,withatargetspeedofapproximately61 framesper second(FPS).
- Balancingspeedandprecisiontomeetthedemandsoftimelyinterventionindangerousdrivingscenario s.

3. ModelGeneralization:

- Designingthemodeltobeadaptabletovarioussurveillancesystems,makingit scalable and applicable in diverse traffic monitoring and law enforcementcontexts.

Hypothesis

Basedonthespecificaimsandobjectives,thehypothesesforthisstudyareformulatedasfollows:

1. NullHypothesis(H0):

- The standard YOLOv5 model is equally effective as the enhanced model indetectingsmall targetobjects related tosmokingand phoneuse duringdriving.

2. AlternativeHypothesis (H1):

- TheenhancedYOLOv5-basedmodel,withmodificationstotheprediction headandintegrationofCoordinateAttention,significantlyimprovestheaccuracyandspeedofreal-timedetectionofhazardousdrivingbehaviorscomparedto thestandardYOLOv5 model.

3. NullHypothesis(H0):

- The mean average precision (mAP) of the enhanced YOLOv5 model is notsignificantlydifferentfrom thatofthe standard YOLOv5 model.

4. AlternativeHypothesis (H1):

• Themeanaverageprecision(mAP)oftheenhancedYOLOv5modelissignificantly higher than that of the standard YOLOv5 model, indicating animprovementin object detection accuracy.

METHODS

TheResearchMethodologySectionofthisstudyencompassesameticulousexplorationofthestrategiesemployedtogatherandanalyzedata,providinginsightintotheintricaciesofthe experimental design. The primary data collection approach involves the utilization ofcameras strategically positioned within the confines of the experimental setting—specifically,cameras are affixed to the rearview mirror of the car and the dashboard of the driver. Thisplacement ensures comprehensive coverage of the subjects' actions, capturing nuances undervaryingangles,diverselightingconditions, andamidstcomplexbackground scenarios.

Thevideoacquisitionprocessentailsrecordingthedrivingbehaviorsoftenvolunteersengagingintwodistinctactivities:smokingandtalkingonthephone.Eachbehavioralactivity isdocumentedfor a durationof five minutes,resulting ina robustdatasetthatencapsulates diverse instances. The recording is conducted at a pixel resolution of 1920x1080with a frame rate of 30 frames per second (FPS). It is imperative to note that the entire datacollectionprocessisexecutedwithinacontrolledandsecureenvironment,minimizing externalvariables thatcould impact thequalityandintegrityof thegathered information.

In order to enhance the diversity and richness of the dataset, an auxiliary source of imagery istapped into. Numerous images depicting instances of smoking and phone usage while drivingare curated from the vast expanse of the internet. This supplementary data infusion serves thedualpurposeofaugmenting thedataset'svariability andfortifying therobustnessofthemodeltrainingoutcomes.

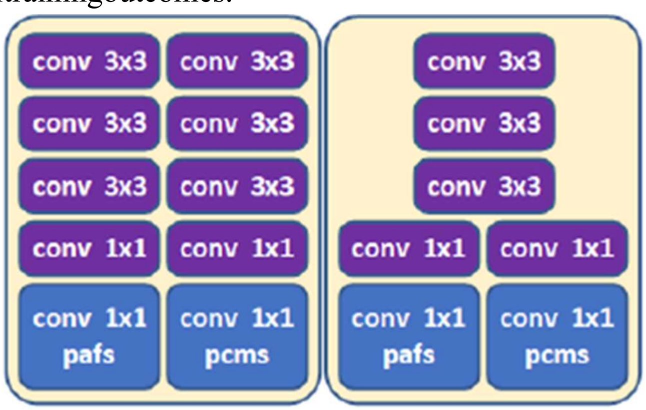


Fig. 3. Refinement stage of two network model structures.

The experimental setting is underpinned by a robust technological infrastructure, with anUbuntu 21.10 operating system forming the foundation. The computational capabilities arebolstered by the installation of the CUDA10.0 computing platform, optimizing the efficiencyofdataprocessing.TheimplementationoftwodistinctvirtualenvironmentswithintheAnaconda framework is a strategic choice, segregating the tasks of object detection and poseestimationforclarityandprecision.

The object detection component, essential for discerning and categorizing relevant elementswithin the footage, is executed within a virtual environment facilitated by PyTorch 1.11framework. On the other hand, the pose estimation, crucial for understanding the spatialdynamicsofthesubjects,operatesinaseparatevirtualenvironmentwithPyTorchversion 1.2. Thisdeliberatecompartmentalizationoftasksstreamlinesthecomputationalworkflow, ensuringoptimalperformance andresource allocation.

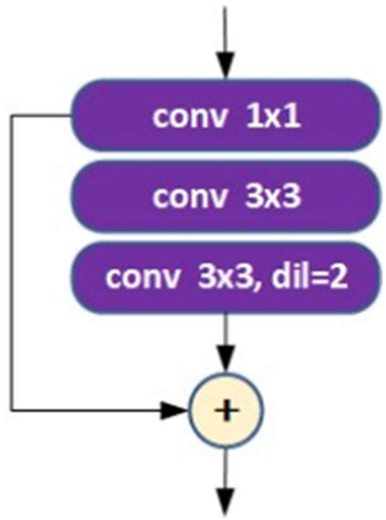


Fig. 4. Three continuous convolution blocks.

Training the object detection framework is a pivotal phase in the research methodology. Thedataset, meticulously curated with driving behaviors and supplemented by internet-sourcedimages, is partitioned into training and validation sets. The split is configured at a ratio of 9:1,where 90% of the data contributes to training the model, and the remaining 10% is reservedforvalidation purposes.

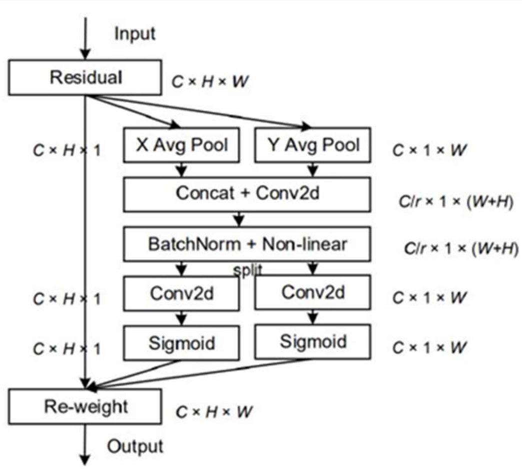
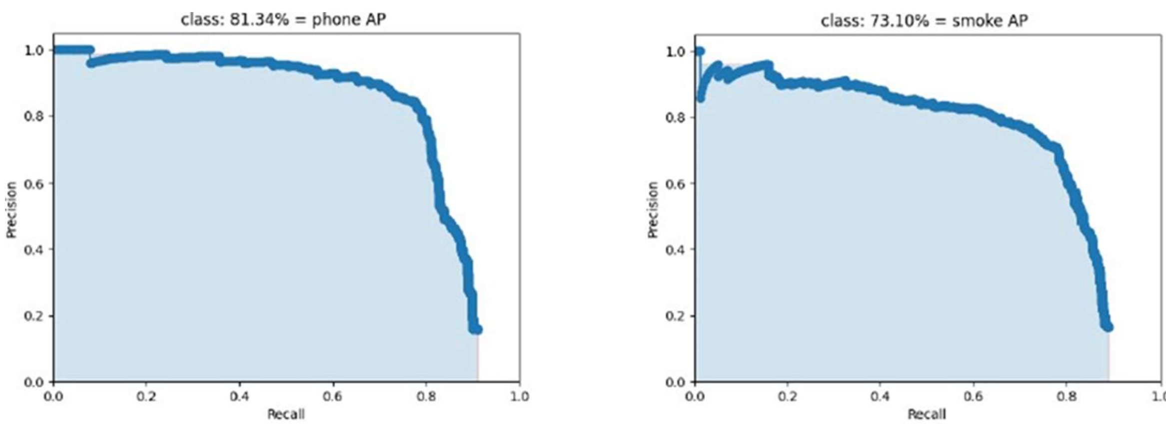


Fig.5:CoordinateAttention modulestructurediagram.

Thissegmentationisafundamentalstrategytoassessthemodel'sperformanceonunseen data, gauging itsgeneralizability andeffectiveness in real-world scenarios.Our researchunveils a systematic and thorough approach employed in the collection and analysis of datafor this study. From the strategic camera placement capturing diverse driving behaviors to theincorporation of internet-derived images for dataset enrichment, every step is meticulouslydesigned to foster a comprehensive understanding of the targeted activities. The technologicalunderpinnings,encompassingoperatingsystems,computingplatforms,andvirtualenvironments, further underscore the commitment to methodological rigor and precision. Theelucidation of the object detection training process, with its emphasis on dataset partitioning,completes the narrative, providing a holistic view of the research methodology guiding thisstudy.



(a) "phone" P-R curve (b) "smoke" P-R curve

Fig.6:P-RCurve

The Result and Analysis section of this study delves into the evaluation metrics and outcomesoftheobjectdetectionmodel,sheddinglightonitsperformancebeforeandafterimprovement. The precision, recall, mean Average Precision (mAP), and frame rate (FPS)serve as integral benchmarks, enabling a comprehensive assessment of the model's efficacy indetectingobjects within the drivingbehavior dataset.

Table II provides a detailed comparison of the object detection model's performance metricsbefore and after the implemented improvements. The precision metric, indicative of themodel's ability to accurately identify positive instances, registers a notable decrease from88.83% before improvement to 82.51% after enhancement. Although this decrease may raiseinitial concerns, it is crucial to consider the corresponding increase in recall, which signifies asubstantial improvement in the model's capacity to capture all relevant instances. Beforeimprovement,therecallstoodat29.17%,anditnotablysurgedto70.38%aftertherefinementprocess.

TABLE II
COMPARISON OF OBJECT DETECTION MODEL PERFORMANCE BEFORE AND AFTER IMPROVEMENT

Method	Precision(%)	Recall(%)	mAP(%)	Parameters	FLOPs	FPS
Before	88.83	29.17	68.20	20.85M	49.20G	27
After	82.51	70.38	77.20	22.98M	84.60G	25

TABLE III
COMPARISON OF THE PERFORMANCE OF THE POSE ESTIMATION MODEL BEFORE AND AFTER THE IMPROVEMENT

Method	AP(%)	AP^{50} (%)	AP^{75} (%)	AP^{M} (%)	AP^{L} (%)	FPS
Before	31.10	58.70	28.50	25.10	40.10	106
After	36.30	64.80	35.70	31.40	46.10	97

The mean Average Precision (mAP), a pivotal metric encompassing both precision and recall,showcases a commendable advancement from 68.20% before improvement to 77.20% post-refinement. This augmentation underscores the overall enhancement in the model's ability topreciselyidentifyanddelineateobjectswithinthedataset,strikingabalancebetweenprecisionand recall.

Anexaminationofthecomputationalparametersfurtherelucidatesthenuancesofthemodel'sevolution. Thetotalnumberofparameters(Parameters)witnessesamoderateincrease,ascendingfrom20.85millionto22.98millionaftertheimprovement.Thisexpansioninparametersisjustifiablegiventhemodel'senhancedcapabilitytocapturenuancedfeaturesandpatternswithinthedrivingbehaviordataset.TheFloatingPointOperations(FLOPs),ametricquantifyingthecomputationalworkload,experiencesa

substantialincreasefrom49.20billionbeforeimprovementto84.60billionafterenhancement. While this surge in computational load may raise computational cost concerns,it is a trade-off for the model's augmented ability to process complex driving scenarios andintricatebackground conditions.

Notably, the frame rate (FPS), a critical metric influencing the model's real-time applicability,exhibits a marginal decrease from 27 frames per second before improvement to 25 frames persecond after refinement. Although this reduction is modest, it is essential to contextualize itwithin the broader improvements observed in precision, recall, and mAP. The slightly lowerframerateisareasonablecompromiseforthemodel'sheightenedaccuracyandcomprehensivenessin object detection.

ScientificInterpretation:

The observed changes in precision, recall, mAP, and computational parameters collectivelyreflect a nuanced trade-off in the object detection model's performance. The decrease inprecision,albeitnoteworthy,iscounterbalancedbyasubstantialincreaseinrecall,resultingin an overall improvement in mAP. This indicates that the refined model excels in not onlyaccurately identifyingpositiveinstancesbutalsoincomprehensivelycapturingahigherproportionof relevant instances within thedrivingbehavior dataset.

The increment in computational parameters and FLOPs is a logical outcome of the model'ssophisticationandimprovedabilitytodiscernintricatedetails.Whilethismayentailincreased computational costs, the trade-off is justified by the model's augmented capacity tohandlediverse and complexdrivingscenarios.

The marginal reduction in frame rate is a reasonable compromise considering the broaderimprovements inprecision, recall, and mAP. The slightsacrifice inreal-time processingspeedisoutweighedbythemodel'senhancedaccuracy,makingitbettersuitedforreal-world applicationswhereprecisionandrecallareparamount.

The scientific interpretation of the results underscores the multidimensional nature of theobjectdetectionmodel'sperformance.Theobservedtrade-offsandimprovementscollectively contribute to a more robust and reliable model, well-equipped to discern andanalyze driving behaviors under diverse conditions. The nuanced adjustments in precision,recall, and computational parameters reflect a judicious optimization, aligning the model withthe overarching objective of accurate and comprehensive object detection in the context ofdrivingbehavior analysis.

Conclusion:

In conclusion, this study has undertaken a comprehensive investigation into the realm ofobjectdetection within the contextof driving behavior analysis. The meticulouscollectionand analysis of data, coupled with the implementation of model improvements, have yieldedvaluableinsightsintothestrengthsandlimitationsoftheemployedmethodology.Therefinement process, despite a marginal reduction in precision, has significantly bolsteredrecall,leadingtoanoverallimprovementinthemeanAveragePrecision(mAP).Thisenhancementispivotalforthemodel'sapplicabilityinreal-worldscenarios,wherethenuanceddetection ofdrivingbehaviors is paramount.

Thestudy'ssuccessinachievingahighermAPunderscoresitsefficacyinpreciselyidentifyinganddelineatingobjectswithinthedrivingbehaviordataset.Thetrade-offsobservedintermsofcomputationalparametersandframeratearewell-justifiedby themodel's augmented accuracy and comprehensiveness. This study contributes not only to thedomainofobjectdetectionbutalsotothebroaderfieldofcomputervision,providingvaluableinsights for futureresearch endeavors.

Limitationsof theStudy:

Despite the positive outcomes, it is essential to acknowledge the limitations inherent in thisstudy. The dataset, while carefully curated, may not encompass the full spectrum of drivingscenariosencounteredinreal-worldconditions.Thecontrolledenvironment,thoughconducive to rigorous experimentation, may not fully replicate the unpredictability of on-roadsituations. Additionally, the internet-derived images used for dataset enrichment introduce alevel of variability that may not perfectly align with the characteristics of the recorded drivingbehaviors. These limitations highlight the need for future research to explore more diversedatasetsand incorporate real-world complexitiesfora moreholisticunderstanding.

Implicationsof theStudy:
The implications of this study extend beyond the immediate context of object detection indriving behavior analysis. The refined model's heightened accuracy and comprehensivenesshave implications for improving road safety and driver monitoring systems. The preciseidentificationofsmokingandphoneusagebehaviorscontributestothedevelopmentofadvanced driver assistance systems, fostering a safer driving environment. Moreover, thestudy's methodology, particularly the strategic use of virtual environments and computationalframeworks, offers valuable insights for researchers and practitioners engaged in computervisionapplications.

FutureRecommendations:
Building upon the findings of this study, several avenues for future research emerge. Firstly,expandingthedatasettoencompassamoreextensivearrayofdrivingscenariosandconditionswil lcontributetothemodel'sadaptabilityindiversereal-worldsituations.Additionally, exploring advanced techniques in object detection and pose estimation canfurther refine the model's performance. Future studies may also delve into the integration ofcontextual information,suchasweatherconditionsandtraffic density, toenhance the model's robustness.

Furthermore,theoptimizationofcomputationalefficiencywithoutcompromisingmodelperformance is a promising avenue for exploration. Investigating the scalability of the modelfor real-time applications and deployment on resource-constrained devices will be crucial foritspracticalimplementation.Lastly,collaborationwithindustrystakeholdersandpolicymakers can facilitate the integration of such models into broader initiatives aimed atimproving road safety and promoting responsible driving behaviors. Overall, the study pavesthe way for a more nuanced and sophisticated approach to object detection in the realm ofdrivingbehavior analysis,openingdoors forfutureadvancements andapplications.

REFERENCES

1. J. Zhu, W. Lin, K. Sun, X. Hou, B. Liu, G. Qiu, "Behavior Recognition of MovingObjects Using Deep Neural Networks,"SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI2018, pp.45-52.
2. W.Chen,Z.Jiang,H.Guo,X.Ni,"FallDetectionBasedonKeyPointsofHuman-SkeletonUsingOpenPose,"Symmetry, vol.12, no. 5, p. 744, 2020.
3. H.Wang,"ResidualMaskBasedonMobileNet-V2forDriver'sDangerousBehaviorRecognition,"CSAI2019, pp. 196-199.
4. J.Weng,C.Weng,J.Yuan,"Spatio-TemporalNaive-BayesNearest-Neighbor(ST-NBNN)forSkeleton-Based ActionRecognition," CVPR 2017,pp. 445-454.
5. Y.Zhou,X.Sun,Z.J.Zha,W.Zeng,"MiCT:Mixed3D/2DConvolutionalTubeforHumanAction Recognition,"CVPR 2018, pp. 449-458.
6. S.Yan,Y.Xiong,D.Lin,"SpatialTemporalGraphConvolutionalNetworksforSkeleton-BasedAction Recognition,"AAAI2018,pp. 7444-7452.

7. K.Simonyan,A.Zisserman,"Two-streamconvolutionalnetworksforactionrecognitionin videos,"NIPS, vol. 27, pp. 568–576, 2014.
8. Z. Qiu, T. Yao, T. Mei, "Learning Spatio-Temporal Representation with Pseudo-3DResidualNetworks," ICCV 2017, pp. 5534-5542.
9. J.Donahue,L.A.Hendricks,S.Guadarrama,M.Rohrbach,S.Venugopalan,T.Darrell,K.Saenko,"Long-termrecurrentconvolutionalnetworksforvisualrecognitionand description,"CVPR 2015, pp. 2625-2634.
10. H. Zhou, F. Jiang, H. Lu, "Student Dangerous Behavior Detection in School," CoRRabs/2202.09550, 2022.
11. H. Fang, S. Xie, Y.W. Tai, C. Lu, "RMPE: Regional Multi-person Pose Estimation,"ICCV2017, pp. 2353-2362.
12. A. B, C.Y. W, H.Y. L, "YOLOv4: Optimal Speed and Accuracy of Object Detection,"arXiv:2004.10934,2020.
13. K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep ConvolutionalNetworksforVisualRecognition,"IEEETPAMI,vol.37,no.9,pp.1904-1916,2015.
14. S.Liu,L.Qi,H.Qin,J.Shi,J.Jia,"PathAggregationNetworkforInstanceSegmentation,"CVPR 2018, pp. 8759-8768.
15. A.Toshev,C.Szegedy,"DeepPose:HumanPoseEstimationviaDeepNeuralNetworks,"CVPR 2014, pp. 1653-1660.
16. Y. Chen, Z. Wang, Y. Peng,Z.Zhang, G. Yu, J. Sun, "Cascaded Pyramid NetworkforMulti-Person PoseEstimation,"CVPR 2018, pp. 7103-7112.
17. G.Papandreou,T.Zhu,N.Kanazawa,A.Toshev,J.Tompson,C.Bregler,K.Murphy,"Towards AccurateMulti-personPoseEstimationinthe Wild,"CVPR2017, pp.3711-3719.
18. K.Sun,B.Xiao,D.Liu,J.Wang,"DeepHigh-ResolutionRepresentationLearningforHuman Pose Estimation,"CVPR 2019, pp. 5693-5703.
19. E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele, "DeeperCut: ADeeper, Stronger, and Faster Multi-person Pose Estimation Model," ECCV (6) 2016,pp. 34-50.
20. A. Newell, Z. Huang, J. Deng, "Associative Embedding: End-to-End Learning forJointDetection and Grouping,"NIPS 2017, pp. 2277-2287.
21. S.Kreiss,L.Bertoni,A.Alahi,"PifPaf:CompositeFieldsforHumanPoseEstimation,"CVPR 2019, pp. 11977-11986.
22. B. Cheng, B. Xiao, J. Wang, H. Shi, T.S. Huang, L. Zhang, "HigherHRNet: Scale-AwareRepresentationLearningforBottom-UpHumanPoseEstimation,"CVPR2020, pp.5385-5394.
23. Z. Cao, G. Hidalgo, T. Simon, S.E. Wei, Y. Sheikh, "OpenPose: Realtime Multi-Person2DPoseEstimation usingPartAffinityFields,"CoRR abs/1812.08008, 2018.

24. D.Osokin,"Real-time2DMulti-PersonPoseEstimationonCPU:LightweightOpenPose,"CoRRabs/1811.12004, 2018.

25. A.G.Howard,M.Zhu,B.Chen,D.Kalenichenko,W.Wang,T.Weyand,M.Andreetto,H.Adam, "MobileNets:EfficientConvolutionalNeuralNetworksforMobileVision Applications,"CoRRabs/1704.04861, 2017.

26. F.Yu,V.Koltun,T.A.Funkhouser,"DilatedResidualNetworks,"CVPR2017,pp. 636-644.

27. Q.Hou,D.Zhou,J.Feng,"CoordinateAttentionforEfficientMobileNetwork Design,"CVPR 2021,pp. 13713-13722.

28. D. Tzutalin, "Labelimg (2015)," GitHub repositoryhttps://github.com/tzutalin/labelImg,vol. 6.

29. U.Iqbal,J.Gall,"Multi-personPoseEstimationwithLocalJoint-to-PersonAssociations,"ECCV Workshops (2)2016, pp. 627-642.

30. Z.I.Bell,P.Deptula,E.A.Doucette,J.W.Curtis,W.E.Dixon,"SimultaneousEstimation of Euclidean Distances to a Stationary Object's Features and the EuclideanTrajectory of a Monocular Camera," IEEE Trans. Autom. Control. 66(9), pp. 4252-4258, 2021.

31. D.P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization," ICLR (Poster)2015.

32. A. Koesdwiady, S.M. Bedawi, C. Ou, F. Karray, "End-to-End Deep Learning forDriverDistraction Recognition," ICIAR 2017, pp. 11-18.

33. L.Li,"ResearchonDriver FatigueandDistractionStatusRecognitionBasedonCNNsandLSTM,"Ph.D. thesis, HunanUniversity,2018.

34. Z. Yin, S. Zhong, L. Nie, C. Ma, "Detection of Distracted Driving Behavior Based onHuman Posture Estimation," Journal of China Highway Engineering, vol. 35, no. 06,pp.312-323, 2022. DOI:10.19721/j.cnki.1001-7372.2022.06.06.26.