International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 3, May-June 2024
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Deep Learning for Automotive Sign Detection

B.Karthiga[1],M.Sivakumar[2],S.Sivasubramanian[3],S.Venkatesan[4],V.Baranidharan[5]
Department of Electronics and Communication Engineering,
DhanalakshmiSrinivasan Engineering College(Autonomous), Perambalur, India.
Email:[1]karthiga.jaya15@gmail.com,[2]esivakumarm@gmail.com,
[3]ssivasubramanian@gmail.com,[4]svenkatesanv7@gmail.com,[5]jacksbani303@gmail.com

***Abstract*—Trafficsignandlightdetectionsarecorecomponentsof Advanced Driver Assistance Systems (ADAS) and self-drivingvehicles. To this end, the automotive industry is widely exploitingcomputer vision (CV) and deep learning (DL) techniques. Thispaper presents a lightweight traffic sign and light detector byharnessing a single-stage, single-shot multi-object detector (SSD).For accelerating the inference speed of the detector, its originalbackbone, VGG16 is replaced by MobileNet V2 that expertlymanages detection speed and network size. In autonomous driv-ing,quickerdetectionperformancewithrespecttothedistanceof an object is of particular interest, for a comfortable braking.However,fartherdistancemakestheobjectstobedetectedappearsmaller.Unfortunately,theoriginalSSDstrugglestodetectsmall objects.Thus,thisworkfurtheroptimizesthenumber of feature map layers of the SSD for the detection ofsmallobjectsalongwithabettertrade-offbetweendetectionprecision and inference time. Experimental analysis confirms theeffectiveness of the proposed model, which achieves 2 times (ormore) faster detection time than the baseline SSD models and acompetitiveprecisionof76.7%.**

***Index Terms*—Computer vision, object detection, ADAS, deeplearning**

I. INTRODUCTION

A reliable real-time detection of traffic sign and light oncomputationally limited platforms is an important concern forautonomous driving. Therefore, only the models with fewerparametersandlowcomputationalcomplexityareneeded.In thisline,variousdeeplearningframeworks,two-stageand single-stage object detectors have been proposed by theresearch community. For example, the Darknet [2] is a simplearchitecture, having fast inference speed. But it specificallysupports only NVIDIA CUDA for acceleration. Hence, thetwo-stage models, like the faster region-based convolutionalneuralnetworks(R-CNN)[3]andregion-basedfullycon-volutional networks (R-FCN) [4] have better detection rate;however,theyrequireimmensecomputationalpowerthatmakethemunacceptableforreal-timeapplications.Ontheotherhand, the single-stage models bridge region proposal, clas-sification and regression tasks, as a single multi-task learning.For instance, the you only look once (YOLO) [5] and single-shotmulti-boxdetector(SSD)[1].Thesemodelsimprovethedetectionspeedandthereforecanbeimplementedonembedded platforms. However, their efficiency is lower intermsofaccuracywhencomparedwithtwo-stagedetectors.

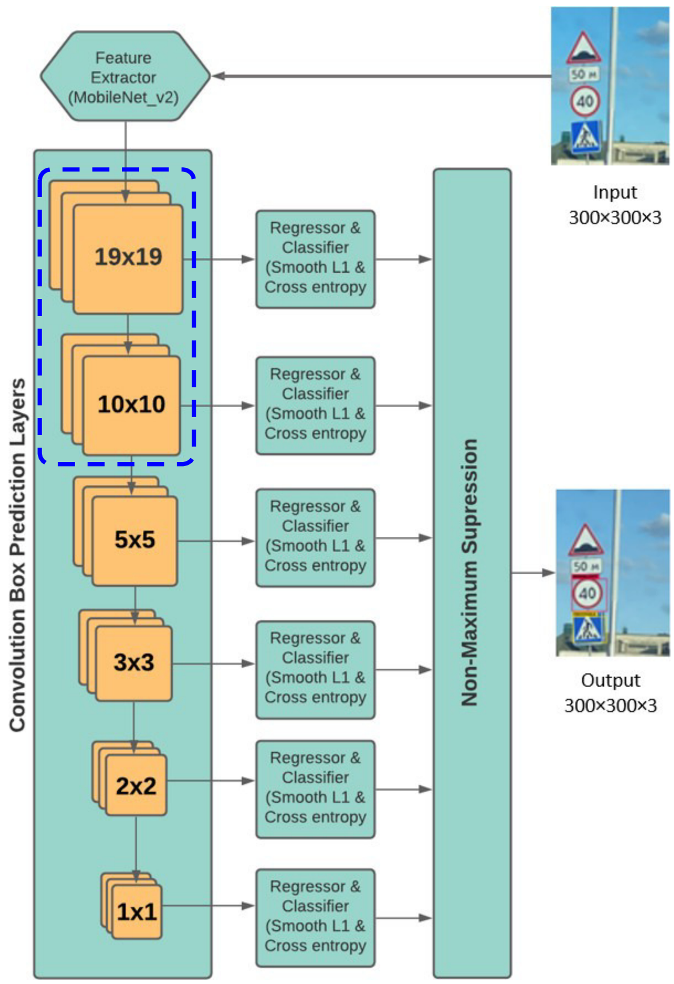

Fig. 1.The proposed SSD-based traffic sign and light detection framework.The blue broken lines indicate the only two feature map layers exploited bythe proposed model out of the 6-box prediction layers in the original SSDmodelin[1].TherationaleofthismodificationisinvestigatedinSectionIII.

There is another major problem with SSD that it is not good atdealing with small objects. Regardless of their small size, thetraffic signs and lights provide key contextual information forintelligent transportation and safety. This work address theseissues of SSD. The proposed detection framework integratesthemulti-scalefeaturemapsobtainedbyabackbonenetwork to enhance the detection performance. The original SSD usessix feature map layers of different scales; however, it strugglesinsmallobjectdetectionwhenallsixfeaturemaplayersare used. Through exhaustive experimental investigation wefound that the first two feature map layers (19×19 and 10×10.cf. Fig. 1) exclusively provides not only better results in termsofsmallobjectdetectionbutalsosignificantreductionindetection time. To further improve the inference speed, thebase network of SSD that is originally the VGG16 [6] is alsoreplacedbyMobileNetv2[7]architecture.

The rest of the paper is organized as follows. Section IIbriefly describes the state-of-the-art models. Section III elabo-rates the proposed SSD-based approach and the improvementsmade in this work. Section IV presents a thorough experimen-tal analysis. Section V closes the paper with conclusion andfuturedirection.

II. RELATEDWORK

Most of the existing solutions focus on a single objectiveeither the detection of the traffic sign or traffic light; not onboth. In retrospect, each of the state-of-the-art for traffic signandtrafficlightdetectionisseparatedbytwoapproaches:conventionalmodelsanddeeplearning(DL)models.

A. ConventionalModels

Thesemethodssolelydependsonthefeatureextractionalgorithms.Letitbeaclassificationordetectionproblem,theymandatorily require low-level features, like color intensities,edge details, and shape features. For example, Kim *et al.* [8]and Diaz-Cabrera *et al.* [9] device a traffic light detectionsystemusingcolorinformation,viz.RGB,hue-saturation-intensity (HSI), and hue-saturation-value (HSV). On the otherhand, the researchers in [10], handle the traffic light detectionas a shape extraction of the traffic light. Similarly, Swathi *etal.* [11] and Supreesh*et al.* [12] also used color clues, as themain features to locate the traffic signs on the road scenes.Hence, Nguyen *et al.* [13] and Yang *et al.* [14] take advantageof shape descriptors, like Hough transform to extract classspecificfeatures,suchascircles,andrectangles.Latertheyuse these features to train a machine learning model to locatethe traffic signs on a given image. These models are featurespecific, and although they achieve a higher precision, theylackrobustnessanddonotgeneralizewellasawholesystem.

B. DeepLearningModels

The advanced object detection, segmentation, and classi-fication algorithms for intelligent transportation applicationsexploits deep learning [15]–[17]. Thus, researchers have fo-cussed on building deep neural networks (DNNs), for trafficlightandsigndetection.Forinstance,Weber*etal.*[18]implementmodel,coinedasDeepTLR.Thenetworkreturnsa pixel-segmented image and then they apply a bounding boxregressor for detection of traffic lights. Similarly, Behrendt *etal.*[19]introduceasystemfordetection,tracking,andclassifi-cationusingadeepconvolutionalneuralnetwork(DCNN).Byutilizingthegeneralobjectdetector,theSSD,Muller*etal.*[20] introduceamodelonlyfortrafficlightdetection.Followingthework of [19], Yudin *et al.* [21] propose another traffic lightdetector based on fully convolutional network (FCN). Theyuse the FCN to get a heat-map highlighting plausible areas oftraffic lights, then employ a high-speed clustering algorithm toobtain traffic light bounding boxes. Besides being a transferlearning approach, it has a very low precision of detection ascomparedtoSSD-basedsolutions,likein[20].

Ontheotherhand,fortrafficsigndetection,Zhang*etal.*[22]usea modifiedversionofYOLOv2[23]objectdetector.Theymanipulatethesizeoffilters,andthenumberof layers to obtain a balance between detection accuracy andspeed. Zhu *et al.* [24] design a custom-built CNN architectureto target more on the smaller size objects. Such target-specificimplementation also faces lack of generalization performance.Thesesolutionscanachievegoodrobustnesscomparedtothe conventional counterparts, as they self learn the featurecorrespondence between the raw inputs and targets. However,they face the criticism for being hungry for data and computepower.

III. PROPOSEDTRAFFICLIGHTANDSIGNDETECTIONFRAMEWORK

A. BasicConcept

Forobjectdetection,aCNN-basedfeatureextractor(cf.Fig. 1) is extended to a larger network by eradicating topclassificationlayersofthebasenetworkandaddingsomesuccessive layers. The successive layers are connected to twomain heads: (1) a regressor to predict bounding boxes, (2) aclassifier to classify each of the detected boxes. Then, a non-maximum suppression (NMS) algorithm is applied to discardinsignificant regions, finalizing the most probable boundingboxes.

Backbone selection: The SSD in [1] was originally builtupontheImageNetpretrainedVGG16visualclassificationnetwork [6], whereby the VGG16 was exploited as a high-level feature extractor. Although VGG16 has good object rep-resentation capability, it is quite a large network architecturewith 24.1 M parameters (cf. Table III). To address this, thiswork strategically replaces the heavy computing VGG16 withMobileNetv2 [7] (cf. Fig. 1). MobileNet versions of CNNsare lighter architectures due to the usage of depth-wise sepa-rable convolution operations. For example, the MobileNet v1andv2havenearlya1/30ofthecomputationalcostandmodelsize ascomparedtoVGG16.

Multi-scale feature maps: The standard SSD model usesVGG16andadditionalsixfeaturemaplayersasshowninFig. 1. They explore features from input images at multiplefield of view for detecting objects of various sizes. TableIshows a comparison of the multi-scale object handling withvarious backbone networks that are considered in this studyalong with the proposed model. In Table I, we can learn that,the SSD with MobileNet V2 (MB v2) uses feature maps withsizesof19×19,10×10,5×5,3×3,2×2and1×1,whichisdifferentfromVGG16-basedSSDmodel.Thefirstfeaturemaplayerwiththesizeof38×38intheVGG16isashallowand

TABLEI
THENUMBEROFFEATUREMAPSUSEDWITH VARIOUSBACKBONES INTHISWORKFORABLATIONSTUDY.

Feature maplayer	SSDwith VGG16	SSDwith MobileNetV2	Proposed model withMobileNetV2
Layer1	38×38×512	19×19×96	19×19×96
Layer2	19×19×1024	10×10×1280	10×10×1280
Layer3	10×10×512	5×5×512	-
Layer4	5×5×256	3×3×256	-
Layer5	3×3×256	2×2×256	-
Layer6	1×1×256	1×1×256	-

notsoeffectiveinextractingkeyattributesofinputimages.Therefore,itismodifiedtohave19×19featuremap.

B. ProposedFastTrafficLight&SignDetectionModel

InSSDwithMBv2,itisempiricallyfoundthatthesixfeaturemaplayersfrom19×19to1×1calculatedpriorboxes

withascaleof0.1,0.2,0.375,0.55,0.725and0.9[1].Therefore, larger feature maps have prior boxes with smallerscales,sotheyareidealfordetectingsmallerobjects.Withthis institution, we modify the standard SSD. As traffic-singsand traffic-lights occupy relatively a smaller fraction of theentire road scene, we exploit the first two prediction layers:19×19, and 10×10 (cf. Table I and Fig. 1), and discard theremaining ineffective layers. This strategic modification savesa lot of memory and computations, resulting in a faster objectdetection.

C.ObjectiveFunctionandEvaluationMetrics

To train the proposed model, we use a weighted sum of theclassification confidence loss (L_{conf}) and the localization loss(L_{loc}),asgiveninEq.(1).

$$L = \frac{1}{N}(L_{conf}+\alpha\cdot L_{loc}), \quad (1)$$

where$N$isthenumberofmatchedpriorboxes,andαisaconstantsetto 1.Thelocalizationlossisthesmooth$L1$loss between the predicted and ground truth bounding boxes.Hence,theconfidencelossiscalculatedusingaSoftmaxactivationwithcategoricalcross-entropy.

The mean average precision (mAP), which is the primarymetricusedforcomparingtheperformanceofobjectdetectors.The average precision (AP) is computed as the average ofmaximumprecisionvaluesatachosen11recallvalues.Hence,theAPforclass$c$isdefinedasinEq.(2).

$$AP_c = \frac{1}{11}\sum_{r}^{R} max(P(r)), \quad (2)$$

where $P(r)$ is the precision for one of the 11 recalls (cf. Eq. 4),$r$and$r \in$ 0.0,··· ,1.0.Hence,themAPforobjectdetectionis the average of the APs calculated over all the classes asshown in Eq. (3), where C is the total number of classes andAP_cisAPforclass$c$withIoU(cf.Eq.(5))of0.5.

$$mAP=\frac{1}{C}\sum_{c}^{C} AP_c. \quad (3)$$

$$Recall= \frac{TruePositive}{TruePositive+FalseNegative}. \quad (4)$$

$$IntersectionOverUnion(IoU)=\frac{Apred\cap Agt}{Apred\cup Agt}, \quad (5)$$

where $Apred$and Agtstand for the area of the predicted boxesand ground truth boxes, respectively. An IoU threshold of 0.5isusedtoclassifywhetherthepredictionisatruepositiveorafalse positive.

IV. EXPERIMENTALANALYSIS

A.Dataset

This study uses the publicly available road sign detectiondataset[25].Asfaraswareconcerned,currently,therearenopeer-reviewedresultspublishedonthisnewlyemerging dataset. It consists of 877 images of four classes: speed limit,stop,crosswalk,andtrafficlight.Thedatasetissplitintotrainingandtestsetsbytakingaratioof8:2.Thus,thetrainingandtestsetshave,respectively,701sampleswith2103objects,and175samples with525objects.

B. ComputationalPlatformandTrainingSetup

TheablationstudywascarriedoutontheGoogleColabPro computational platform with the following specifications.A2.20GHzIntel(R)Xeon(R)CPUwitha12GBmemory,anda 1.59 GHz NVIDIA T4 GPU with 16 GB memory. The entireprogramwaswritteninPyTorch1.9.0+cu102withPython 3.7.10.Themodelsweretrainedusingstochasticgradientdescent (SGD) optimizer with an initial learning rate of 0.001,momentumof0.9,weightdecayof0.0005,learningratedecaypolicydropbyafactorof10,andbatchsizeissetto16.

C. QuantitativeAnalysis

To study the impact of proposed framework, four indepen- dent investigations, Model A - D, were carried out consideringdifferentnumberoffeaturemaplayerswithdifferentscalesas listed in Table II. Where, Model A uses all the six featuremap layers, while Model B was created with only the first19×19 feature map exclusively omitting other layers. Hence,the Model C was built with only the fist two feature maplayers,19×19and10×10.ThelastModelDwasbuiltusingfist three feature map layers: 19×19, 10×10, and 5×5. All themodels were fine tuned for 10 epochs. This sanity test showsthat the Model C produces the best results and the mAP startsdeterioratingwhenthenumberoflayersisincreasedtothree. Tofurthervalidate,thesegregatedobjects'areasareanalyzed basedonthesizeoftheircorrespondingboundingboxeswrtthefullimagetocalculatethepercentageofcoverage.FromFig.2,wecanseethat95.4%(41.9+53.46)ofobjectshavingmaximum20%ofimagedimensioncoverage,i.e.,majorityoftheobjectscomeundertherangeofsmall/mediumsizecomparedtotheinputimagedimensions.ThatisthereasonforincreasedmAPoftheModelCascomparedtoModelA.FurtherreferringtoTableII,onecanseethat,intermsof thenumberoftrainableparameters,ourproposedframework

TABLEII
PERFORMANCEOFTHESSDWITHDIFFERENTNUMBEROFFEATUREMAPLAYER SOFTHEMOBILENETV2(WITHOUTDATAAUGMENTATION)

Model	mAP@ IoU0.5	#of Pa-rameters	# of predic-tionsperclass
ModelA:FullSSDwith MBv2	0.307	7,222,518	2,268
ModelB:OnelayerSSDwit hMBv2	0.391	3,536,524	1,444
ModelC:Two-layerSSDwithMBv2	0.433	4,158,146	2,044
ModelD:Three-layerSSDwithMBv2	0.370	6,932,088	2,194

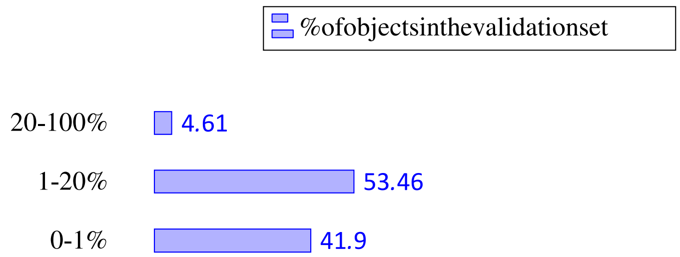

Fig.2.Percentagecoverageoftheobjects'groundtruthboundingboxwrtthe whole image dimension. The ordinates shows objects' coverage. Based onthe % of area cover the bounding boxes they are grouped into 0-1% - small,1-20%-medium,and20-100%-largeobjects.

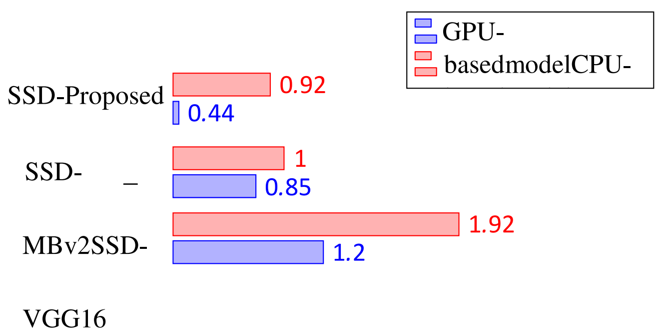

Fig.3.Detectiontimeanalysisoftheproposedtwo-layerSSDwithMobileNetV2.Thebarsshowthetiminginsecond.

TABLEIII
COMPARATIVE ANALYSISOF THE PROPOSED MODEL WITH BASELINE SSD MODELS (WITH DATA AUGMENTATION)

Model	mAP@I oU0.5	DetectionTime(s)		#ofPara--meters
		GPU	CPU	
SSDwithVGG16	76.7	1.203	1.917	24.1M
SSDwithMBv2	68.8	0.848	0.997	7.2M
ProposedSSDwithtwo -layerMBv2	73.8	0.443	0.924	4.1M

with two-layer SSD (Model C) has less number of parametersas compared to Model A. Less number of parameters means alighter architecture with less computational complexity. Also,the Model C involves 224 lesser number of predictions perclassascomparedtotheModelAwithallthesixlayersleading to a smaller number of computations. All these prop-erties of the proposed solution make it more suitable for real-time detection and deployment of the model on an embeddedsystem.

Using the experimental findings discussed earlier (cf. Ta-bleII)asaproofofconcept,wefinalizethetwo-layer SSD with MBv2 and conduct further ablation study. In theseexperiments,wealsoapplyfourdataaugmentationtechniques:photometric distortions, expand image (zoom out), randomlycrop image, and horizontal flip to get better generalizationperformance. The models are trained with early stopping (thebestmAPisfoundat58thepoch).Theexperimentalresultsare tabulated in Table III, and compared in Fig. 3, and Fig. 4.The average detection times on CPU and GPU are comparedfor all the three models in Fig. 3. The results show noticeableimprovements even in the case of CPU. These results provethat the proposed framework can be deployed on an embeddedplatformforreal-timetrafficsignandlightdetection.

Overall, theproposed two-layer SSDwithMobileNet V2isfoundtobefasterthanthebaselineSSDwithVGG16. Asexpected,theSSDwithVGG16hasabettermAPduetothedeeperfeaturelearningarchitecture(cf.TableIII).However, the proposed model renders a comparable detectionperformance and achieves a 5% improvement in mAP whencompared to the full layer SSD with MBv2. As noticed inFig.3,intermsofdetectiontime,thebaselineSSDwith VGG16is63%and52%slowerwhencomparedtotheproposed two-layer SSD model, respectively on GPU-basedand CPU-based implementations. In a nutshell, the proposedmodel, nearly 2 times faster than the baseline model with aminorcompromiseondetectionprecision<3%.

D. *QualitativeAnalysis*

Fig. 4 shows few visual results for comparing the top-2models: the proposed two-layer SSD with MB v2 and thebaseline SSD with VGG16. We can clearly notice that theproposed model's performance is much better as compared tothebaselinemodel,especiallyinimageIDs:road807,road748,road716androad213intermsofsmallobjectdetection.Itis

also noticeable in image ID: road821 that the proposed modelcandetecttwoextratrafficlightsthatthebaselinemodelisnot abletodetect.

V. CONCLUSION

This works presents an efficient exploitation of the SSDmodelforafasttrafficsignandlightdetection.Intheproposedframework,thestandardSSDarchitecture'sbackbone,VGG16is replaced with the lighter MobileNet v2 and only the top-twofeaturemaplayersoftheSSDareused.Suchmodificationsareproved to be essential not only for small object detection butalso quicker detection. The proposed model achieves 2 times(or more) lesser detection time than the baseline SSD modelsand a comparable detection performance of 76.7% mAP. Theproposed solution can be further improved by training on abigger dataset. Apart from the intended purpose, it can bebeneficial for the applications, where the detection of smallobjectsplaysanimportantrole.

REFERENCES

[1] W.Liu,D.Anguelov,D.Erhan,C.Szegedy,S.Reed,C.-Y.Fu,and A.C.Berg,"Ssd:Singleshotmultiboxdetector,"in*EuropeanConf.oncomput. vis.*Springer,2016,pp.21–37.

DetectionResultsfromtheStandardSSDwithVGG16

DetectionResultsfromtheProposedModel(SSDwithTwo-layerMobileNetV2)

Fig.4.FewSamplesofQualitativeResults.ImageIDsfromCol.1to5:road807,road748,road716,road213,androad821.

[2] Y. Koo, C. You, and S. Kim, “Opencl-darknet: An openclimplementation for object detection,” in *2018 IEEE Intl. Conf. on Big Data andSmartComputing(BigComp)*.IEEE,2018,pp.631–634.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: towards real-timeobject detection with region proposal networks,” *IEEE trans. on patternanalysisandmachineintelligence*,vol.39,no.6,pp.1137–1149,2016.

[4] J. Dai, Y. Li, K. He, and J. Sun, “R-fcn: Object detection via region-based fully convolutional networks,” in *Advances in neural informationprocessingsys.*,2016,pp.379–387.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only lookonce:Unified,real-timeobjectdetection,”in*Proc.oftheIEEEConf.oncomput.vis.andpatternrecog.*,2016,pp.779–788.

[6] K. Simonyan and A. Zisserman, “Very deep convolutional networks forlarge-scaleimagerecog.”*arXivpreprintarXiv:1409.1556*,2014.

[7] M.Sandler,A.Howard,M.Zhu,A.Zhmoginov,andL.-C.Chen,“Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proc. of theIEEEConf.oncomput.vis.andpatternrecog.*,2018,pp.4510–4520.

[8] H.-K. Kim, Y.-N. Shin, S.-g. Kuk, J. H. Park, and H.-Y. Jung, “Night-timetrafficlightdetectionbasedonsvmwithgeometricmomentfeatures,” *Intl. Journal of comput. and Information Engineering*, vol. 7,no.4,pp.472–475,2013.

[9] M. Diaz-Cabrera, P. Cerri, and J. Sanchez-Medina, “Suspended trafficlights detection and distance estimation using color features,” in *201215th Intl. IEEE Conf. on intell. Transportation sys.*IEEE, 2012, pp.1315–1320.

[10] M. Omachi and S. Omachi, “Traffic light detection with color and edgeinformation,” in *2009 2nd IEEE Intl. Conf. on comput. Science andInformationTechnology*.IEEE,2009,pp.284–287.

[11] M. Swathi and K. Suresh, “Automatic traffic sign detection and recog.:A review,” in *2017 Intl. Conf. on Algorithms, Methodology, Models andApplicationsinEmergingTechnologies*.IEEE,2017,pp.1–6.

[12] H. Supreeth and C. M. Patil, “An approach towards efficient detectionand recog. of traffic signs in videos using neural networks,” in *2016 Intl.Conf. on Wireless Communications, Signal Processing and Networking(WiSPNET)*.IEEE,2016,pp.456–459.

[13] B. T. Nguyen, S. J. Ryong, and K. J. Kyu, “Fast traffic sign detectionunder challenging conditions,” in *2014 Intl. Conf. on Audio, LanguageandImageProcessing*.IEEE,2014,pp.749–752.

[14] Y. Yang, H. Luo, H. Xu, and F. Wu, “Towards real-time traffic signdetection and classification,” *IEEE trans. on intell. transportation sys.*,vol.17,no.7,pp.2022–2031,2015.

[15] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, “A 3d cnn-lstm-based image-to-image foreground segmentation,” *IEEE Transactions onIntelligentTransportationSystems*,vol.21,no.3,pp.959–971,2019.

[16] T. Akilan and Q. J. Wu, “sendec: An improved image to image cnn forforeground localization,” *IEEE Transactions on Intelligent Transporta-tionSystems*,vol.21,no.10,pp.4435–4443,2019.

[17] T.Akilan,Q.J.Wu,A.Safaei,andW.Jiang,“Alatefusionapproachforharnessingmulti-cnnmodelhigh-levelfeatures,”in*2017IEEEInternationalConferenceonSystems,Man,and Cybernetics(SMC)*.IEEE,2017,pp.566–571.

[18] M.Weber,P.Wolf,andJ.M.Zöllner,“Deeptlr:Asingledeepconvolu- tional network for detection and classification of traffic lights,” in *2016IEEEintell.vehiclessymposium(IV)*.IEEE,2016,pp.342–348.

[19] L. B. Karsten, N. Libor, and B. Rami, “A deep learning approach totraffic lights: Detection,” in *2017 IEEE Intl. Conf. on Robotics andAutomation*,2017,pp.1370–1377.

[20] J.MüllerandK.Dietmayer,“Detectingtrafficlightsbysingleshotdetection,” in *2018 21st Intl. Conf. on intell. Transportation sys. (ITSC)*.IEEE,2018,pp.266–273.

[21] D. Yudin and D. Slavioglo, “Usage of fully convolutional network withclustering for traffic light detection,” in *2018 7th Mediterranean Conf.onEmbeddedComputing(MECO)*.IEEE,2018,pp.1–6.

[22] J. Zhang, M. Huang, X. Jin, and X. Li, “A real-time chinese traffic signdetectionalgorithmbasedonmodifiedyolov2,”*Algorithms*,vol.10,no.4, p.127,2017.

[23] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” in *Proc.of the IEEE Conf. on comput. vis. and pattern recog.*, 2017, pp. 7263–7271.

[24] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, “Traffic-signdetection and classification in the wild,” in *Proc. of the IEEE Conf. oncomput.vis.andpatternrecog.*,2016,pp.2110–2118.

[25] M.Andrew,“Roadsigndetectiondataset,”https://www.kaggle.com/andrewmvd/road-sign-detection,May2020.

