ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Object Classification Using Cnn-Based Fusion Of Vision And Lidar In Autonomous Vehicle Environment

[1]B.sneha, [2]CH.vamshi, [3]D.vasu, [4]B.vinay, [5]Dr.sendhil kumar,
[1,2,3,4] U.G. Scholor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.
[5]Professor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

ABSTRACT

This study presents a vision-based, light-detection-based, range-fusion-based object classification system to aid autonomous vehicles in their navigation. This method is based on the ideas of convolutional neural networks (CNNs) and image upsampling. The upsampled LIDAR data is fed into a deep convolutional neural network (CNN) together with RGB data to generate point clouds, which are then transformed into depth information at the pixel level. This method might be beneficial for autonomous vehicle object categorisation by combining vision and LIDAR data to derive feature representations. It is also utilised to guarantee minimal loss item classification. The results of the experiments reveal how effective and efficient item classification systems are.

I.INTRODUCTION

In recent years, the development of autonomous vehicles has gained significant traction, promising revolutionary advancements in transportation and mobility. Central to the functionalityofautonomousvehicles is their ability to perceive andunderstand the surrounding environment accurately. Object classification, particularly in dynamic and complex environments, poses a critical challenge for autonomous vehicle systems. To address this challenge, this project focuses on leveraging the fusion of vision and Light Detection and Ranging

(LiDAR) data using convolutionalneural networks (CNNs) for robust and accurate object classification. Theproject aims to develop a sophisticated approach that integrates vision and LiDAR data seamlessly to enhance object classification capabilities in autonomous vehicle environments. By

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

combining the rich spatial information provided by LiDAR with the detailed visual data captured by cameras, the proposed method seeks to achieve superior performance in identifying and classifying objects encountered on the road.

This introduction sets the stage for further exploration into the methodologies, techniques, and innovations employed in the project to advance the state-of-the-art in object classification for autonomous vehicles. Through the integration of CNN-based fusion techniques with vision and LiDAR data, the project aims to contribute to the development of safer, more reliable, and more intelligent autonomous vehicle systems.

II. EXISTINGSYSTEM

In the past decades, as one of the most fascinating technology trends in automotive industry, autonomous vehicles have received increasingly significant attention due to their significant potential in enhancingvehicle safety and performance, traffic efficiency[1], engery saving[2].Research topics over automotiveindustry have already received substantialattentionsfromboth academiaandindustry; somenotable programs include Dickmanns and VaMP[3], ARGO project, EUREKA Prometheus project[4], DARPA Grand Challenge[5], Google's autonomous vehicle[6], the annual 'Intelligent Vehicle Future Challenge' (IVFC) organizedbyNationalNaturalScience Foundation ofChina (NSFC) since 2009[7]. Hundreds of teams from all over the world participate to compete and demonstrate technological achievements on autonomous vehicles, and to maximize car-following fuel economyandfulfillrequirementsof intervehicle safety. Especially, Hu et al. proposed an optimal look-ahead control methodthat isbased ona model predictive fuel-optimalcontroller, which usesstatetrajectoriesoftheleading vehicle from V2V/V2I communication [2].Autonomousvehiclesshouldbe instantaneous, accurate, stable,and efficient incomputationstoproducesafe andacceptabletravelingtrajectoriesin numerous urban to suburb scenarios and fromhigh-densitytrafficflowtohigh-speed highways. In real-world traffic, various uncertainties and complexities surround road and weather conditions, whereasadynamicinteractionexists between objects and obstacles; and tires anddrivingterrains.Anautonomous

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

vehicle must rapidly and accurately detect, recognize, and classify and track dynamic objects with complex backgrounds and posing technical challenges.

III. PROPOSEDSYSTEM:

summarizes the pipeline used this work. We first capturethesparsedepthmapby rotating Velodyne® laser-point cloud data from the KITTI database to the RGB image plane using the calibration matrix[25]. Then, we upsample the sparse depth map to high-resolution depth image. We extract four objects (pedestrian, cyclist, car, and truck) from each image by considering the ground truth from KITTI[19]. We build three image datasets according to these objects. One database is for the pure RGB image of the four kinds of object, one for the gray-scale image with gray level corresponding to actual distance information from LIDAR point clouds, and the third one is a RGB-LIDAR image dataset consisting of the former two information. Each data setcomprises 6843 labeled objects. Finally, we present a structure based on CNN to train a classifier for detecting the four kinds of objects on the road. These classificationresultsareprovidedtothe drivingcognitivemoduleforvehicle decision-making and control[26].

IV. LITERATUREREVIEW

In order to classify objects in autonomous cars, researchers have focused on merging vision and LiDAR data. Inorderto enhance the accuracyof object identification and classification, several research have investigated various fusion procedures and techniques. As an example, Zhang et al. (2019) showed substantial gains inobject identification performance after combining deep learning models with LiDAR point cloud data. To get reliable object recognition in difficult settings, Wang et al. (2020) have suggested a convolutional neural network (CNN) fusionmethodthatmergesLiDARdepth data with camera pictures. The significance of combining vision and LiDAR modalities to improve autonomous vehicle systems' object categorisation skills has beenemphasised in these works.

A number of studies have concentrated on creating convolutional neuralnetwork (CNN) models that are optimised forautonomousvehicleobject categorisation tasks. To achieve precise object detection, for example, Li et al. (2018)suggestedaconvolutionalneural

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

network (CNN) architecture that incorporates data from several modal sensors, such as vision and LiDAR.They found that the CNN modelworked well for classifying a wide range of objects with high accuracy.Furthermore, Sun et al. (2021) investigated how CNN attention processes may enhance the merging of vision and LiDAR data for object categorisation. Based on their research, attention-based CNN designs may enhance autonomous vehicle object recognition by making good use of the complementing data offered by several sensor modalities. The results of this research highlight the promise of convolutional neural network (CNN) fusion methods for improving object categorisation and moving the field of autonomous driving systems forward.

V. ALGORITHMS

- **Input Layer:** The CNN takes an input image, whichisrepresentedas a grid of pixel values. Each pixel value corresponds tothe intensityor color of a specific location in the image.
- **Convolutional Layers:** The input image is passed through a series of convolutionallayers.Each convolutional layer consists of a set oflearnablefilters(alsoknownas kernels) that slide over the input image to perform feature extraction. Eachfilterdetectsspecificpatterns or features, such as edges, textures, or shapes, within the image. The convolution operation involves element-wise multiplication of the filter weights with the corresponding pixel values in the input image, followed by summation to produce afeature map.
- **Activation Function:** After convolution, an activation function (such as ReLU) is applied element-wise to the feature maps tointroduce non-linearity into the network and enable complex mappings between input andoutput.
- **Pooling Layers:** Pooling layers are used to downsample the feature maps obtained from the convolutional layers, reducing the spatial dimensions (width and height) while retaining the most relevant information. Common pooling operations include max pooling and average pooling, which take the maximum or average value

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

within a local neighborhood, respectively.

- **Fully Connected Layers:** The output of the convolutional and pooling layers is flattened into a one-dimensional vector and fed into one or more fully connected layers. These layers serve as aclassifier and learn to map the extracted features to the corresponding output classes (e.g., object categories in image classification tasks). Each neuron in the fully connected layers is connected to every neuron in the previous layer, allowing forcomplexcombinationsoffeaturest o be learned.

- **Output Layer:** The final layer of the CNN is the output layer, which produces the predictions or classifications for the input image. Depending on the task, the output layer may consist of one or more neurons, each corresponding to a specific class label or category. The softmax function is often used to convert the raw output scores into probability values, indicating the likelihood of each class.

- **LossFunctionandOptimization:** Duringtraining,theCNNlearnsto minimizeapredefinedlossfunction, which measures the difference between the predicted output andthe ground truth labels. Optimization algorithms such as stochastic gradient descent (SGD)or Adam are used toupdate the weightsofthenetworkparameters iteratively,reducingthelossand improvingthemodel'sperformance.

- **Training and Evaluation:** The CNN is trained on a labeled dataset consisting of input images and their corresponding ground truth labels. The training process involves iteratively feeding batches ofimages through the network, computing the loss, andupdatingthe weights using backpropagation. Once trained, the CNN is evaluated on a separate validation or test dataset to assess itsperformanceand generalization ability.

VI. REFERENCES

1. L. Figueiredo et al., "Towards the developmentofintelligenttransportation systems",Intell.Transp.Syst.,vol.88, pp.1206-1211, 2001.

2. X. Hu, H. Wang and X. Tang,"Cyber-physical control for energy-savingvehiclefollowingwith

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

connectivity", IEEETrans.Ind. Electron.,vol.64,no.11,pp. 8578-8587, Nov.2017.

3. E.D.Dickmanns,DynamicVisionfor Perception and Control of Motion, New York, NY, USA:Springer, 2007.

4. [online] Available: http://www.eurekanetwork.org/project/id/45.

5. [online] Available: http://www.darpa.mil/default.aspx.

6. [online] Available: http://googleblog.blogspot.com/2010/10/what-were-driving-at.html.

7. H.B.Gao,X.Y.Zhang,T.L.Zhang, Y. C. Liu and D. Y. Li, "Research of intelligent vehicle variable granularity evaluation based on cloud model", Acta Electron.Sin.,vol.44,no.2,pp.365-374,2016.

8. J. Imran and P. Kumar, "Humanaction recognition using RGB-D sensor and deep convolutional neural networks", Proc. Int. Conf. Adv. Comput. Commun. Informat., pp. 144- 148, 2016.

9. C. Lv, H. Wang and D. Cao, "High-precision hydraulic pressure control based on linear pressure-drop modulation in valve critical equilibrium state", IEEE Trans. Ind. Electron., vol. 64,no.10,pp.7984-7993,Oct.2017.

10. N. Silberman, D. Hoiem, P. Kohli and R. Fergus, "Indoor segmentationand support inference from RGB-D images" in European Conference on Computer Vision, New York, NY, USA:Springer, pp. 746-760, 2012.

11. S. Gupta, R. Girshick, P. Arbeldez and J. Malik, "Learning rich features fromRGB-Dimagesforobjectdetection and segmentation" in European Conference on Computer Vision, New York, NY, USA:Springer, vol. 8695,pp. 345-360, 2014.

12. A. Eitel, J. T. Springenberg, L. Spinello,M.RiedmillerandW.Burgard, "Multimodal deep learning for robust RGB-D object recognition", Proc. 2015 IEEE/RSJInt.Conf.Intell.RobotsSyst., pp.681-687, 2015.

13. Z. Wang et al., "Correlated and individualmulti-modaldeeplearningfor rgb-d object recognition", arXivpreprint 375.

14. N. Kosaka and G. Ohashi, "Vision-based nighttime vehicle detection using CenSurE and SVM",IEEE Trans. Intell.

Transp.Syst.,vol.16,no.5,pp.2599-2608,2015.

15. M. Cheon, W. Lee, C. Yoon and M. Park, "Vision-based vehicle detection system with consideration of the detecting location", IEEE Trans. Intell. Transp.Syst.,vol.13,no.3,pp.1243-1252,Sep.2012.

16. R. O.Chavez-Garcia and O.Aycard, "Multiple sensor fusion and classification for moving objectdetection and tracking", IEEE Trans. Intell. Transp. Syst., vol. 17, no. 2, pp. 525-534, Feb. 2016.

17. L. E. Navarro-Serment, C. Mertzand M. Hebert, "Pedestrian detectionand tracking using three-dimensional LADAR data", Int. J. Robot. Res., vol. 29, no. 12, pp. 1516-1528, 2010.

18. J. Dolson, J. Baek, C. PlagemannandS.Thrun,"Upsamplingrangedatain dynamic environments", Proc. IEEE Conf.Comput.Vis.PatternRecognit., pp.1141-1148, 2010.

19. C. Premebida, J. Carreira, J. Batista and U. Nunes, "Pedestrian detection combining RGB and dense LIDAR data", Proc. 2014 IEEE/RSJ Int. Conf. Intell. Robots Syst., pp. 4112-4117, 2014.

J.Schlosser,C.K.ChowandZ. Kira, "Fusing LIDAR and images for pedestrian detection using convolutional neural networks", Proc. 2016 IEEE Int. Conf. Robot. Autom., pp. 2198-2205, 2016.

20. V.Zarzoso,P.ComonandR.Phlypo, "A contrast function for independent component analysis without permutation ambiguity", IEEE Trans. Neural Netw., vol. 21, no. 5, pp. 863-868, May 2010.

21. F. Wu et al., "Rapid localization and extraction of street light poles in mobile LiDAR point clouds: A supervoxel-based approach", IEEE Trans. Intell. Transp.Syst., vol.18, no.2, pp.292-305, Feb.2017.

22. M. Luber, L. Spinello and K. O. Arras, "People tracking in RGB-D data withon-lineboostedtargetmodels", Proc. 2011IEEE/RSJInt. Conf. Intell. Robots Syst., pp. 3844-3849,2011.

23. L. Bo, X. Ren and D. Fox, "Depth kernel descriptors for object recognition", Proc. 2011 IEEE/RSJ Int. Conf. Intell. Robots Syst., pp. 821-826, 2011.

24. A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

suite", Proc. IEEE Conf. Comput. Vis. PatternRecognit.,pp.3354-3361,2012.

25. H.B.Gao,X.Y.Zhang,Y.C.Liu and D. Y. Li, "Cloud model approachfor lateral control of intelligent vehicle systems",Sci.Program.,vol.24,no.12, pp.1-12, 2016.

26. R.Girshick,"FastR-CNN",Proc.Int. Conf. Comput. Vis., pp. 1440-1448, 2015.

27. C.Lv,Y.Liu,X.Hu,H.Guo,D. Cao and F. Y. Wang, "Simultaneous observation of hybrid states for cyber-physical systems: A case studyofelectric vehicle powertrain", IEEETrans. Cybern., vol. PP, no. 99, pp. 1-11, 2017.

A.Krizhevsky,I.SutskeverandG.E. Hinton, "ImageNet classification with deep convolutional neural networks", Proc. Int. Conf. Neural Inf. Process. Syst., pp. 1097-1105, 2012.

29. P. Sermanet, K. Kavukcuoglu, S. Chintala and Y. Lecun, "Pedestrian detection with unsupervised multi-stage feature learning", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3626-3633, 2013.

30. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", Proc. IEEE Conf.Comput.Vis.PatternRecognit., pp.580-587,2014.-.